Digitales Sammlungsmanagement

(Grundlagen, Kriterien, Show- und Usecases)

Peter Bubestinger-Steindl
(peter @ ArkThis.com)

Juli 2023

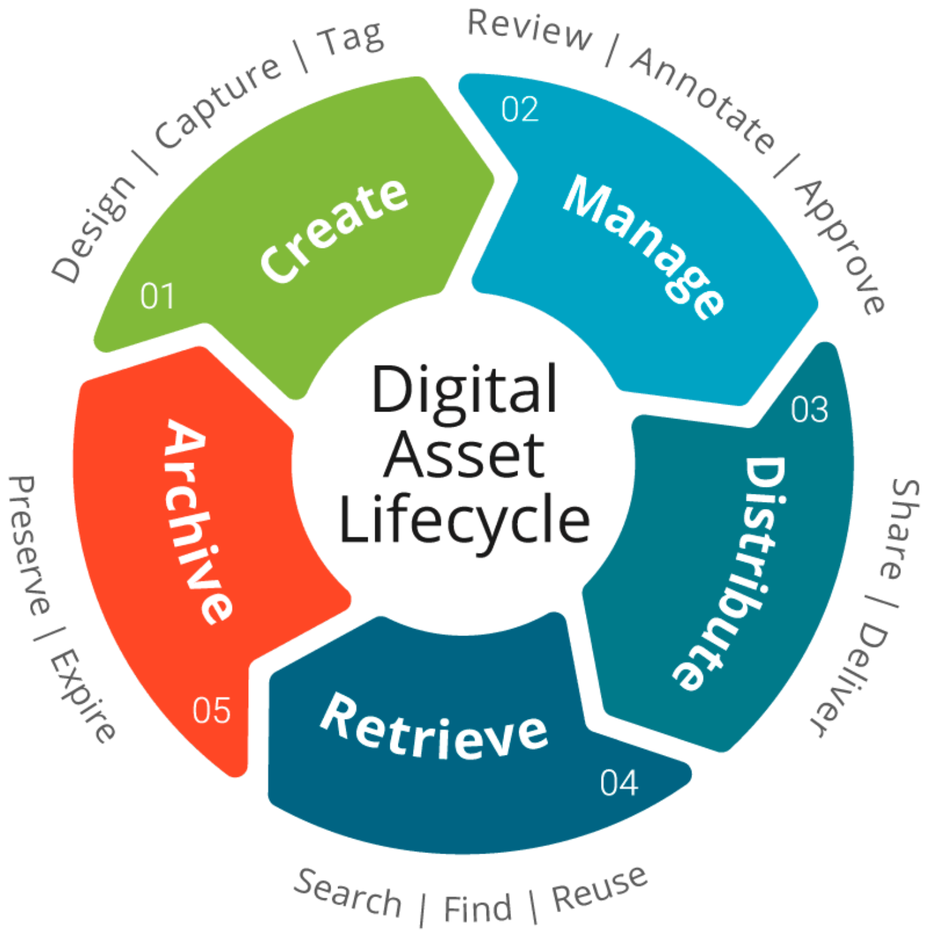
Speaker notes

No notes on this slide.

Speaker notes

No notes on this slide.

“Digitale Assets” verwalten



Speaker notes

No notes on this slide.

Sammlungsverwaltungssysteme? Digital?

Speaker notes

No notes on this slide.

Puh. Wo fangen wir an...?

Es ist eine Datenbank mit einer UI und wahrscheinlich Referenzen auf Files. Gedacht um Dinge suchen und organisieren zu können.

- **CMS:** Collection Management System
- **MAM:** Media Asset Management
- **DAM:** Digital Asset Management

Speaker notes

CMS usually includes a MAM, and a MAM is a media-type-aware version of a DAM.

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: The abbreviation “CMS” usually means “**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.

Eine Verbindlichkeit.

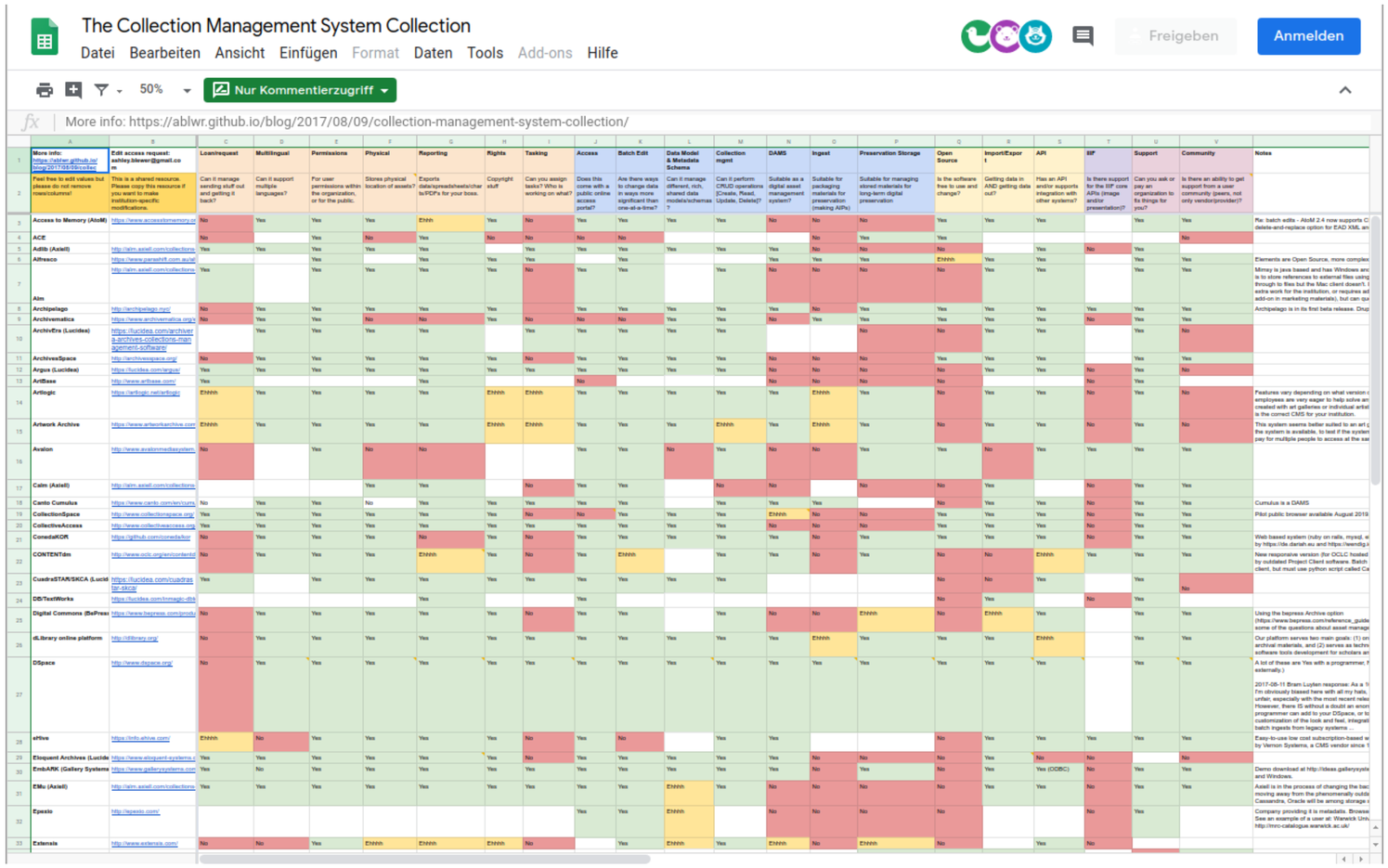
Sehr ernsthaft langfristig. *Drum prüfe, entscheide und konfiguriere weise…* 

Speaker notes

No notes on this slide.

laaange Liste von (70+) Systemen... 😮

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



Speaker notes

No notes on this slide.

Gekürztes Update (50+)

kb.kulturerbe-digital.at/books/empfehlungen-software/page/vergleich-sammlungsverwaltungssysteme

Speaker notes

No notes on this slide.

A kurze Liste

von beliebt-bekannten Open Source Systemen:

- Omeka
- AtoM (Access To Memory)
- ResourceSpace

Speaker notes

No notes on this slide.

Verschiedene DAMs sind verschieden.

- Für welche Anwendungsfälle war es gedacht?
- Ordnentliches Medien-Typ-Handling? (image, docs, AV?)
- Verschiedene Datenschemata parallel?
- Lokale App vs Browser-basiert?
- Lizenzmodelle?
- Setup/maintenance Komplexität?

Speaker notes

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 too much…
- Speak with others who have the same use-cases (or system).

DAM Kriterien - Beispiele

Speaker notes

- **Loan/request:** Can it manage sending stuff out and getting it back?
  - **Multilingual:** Can it support multiple languages?
  - **Permissions:** For user permissions within the organization, or for the public.
  - **Physical:** Stores physical location of assets?
  - **Reporting:** Exports data/spreadsheets/charts/PDFs for your boss.
  - **Rights:** Can it handle copyright stuff?
  - **Tasking:** Can you assign tasks? Who is working on what?
  - **Web-Access:** Does this come with a public online access portal?
  - **Batch Edit:** Are there ways to change data in ways more significant than one-at-a-time?
  - **Data Model & Metadata Schema:** Can it manage different, rich, shared data models/schemas?
  - **Collection Management:** Can it perform CRUD operations [Create, Read, Update, Delete]?
  - **DAMs:** Suitable as a digital asset management system?
  - **Ingest:** Suitable for packaging materials for preservation (making AIPs)?
  - **Preservation Storage:** Suitable for managing stored materials for long-term digital preservation
  - **Open Source:** Is the software free to use and change?
  - **Import/Export:** Getting data in AND getting data out?
  - **API:** Has an API and/or supports integration with other systems?
  - **IIIF:** “Is there support for the IIIF core APIs (image and/or presentation)? https://iiif.io/technical-details/”
  - **Support:** Can you ask or pay an organization to fix things for you?
  - **Community:** Is there an ability to get support from a user community (peers, not only vendor/provider)?

Basic Features?

- Mehrsprachigkeit
- Zugriffsrechte
- Physische Lagerverwaltung
- Reporting
- (Urheberinnen-)Rechte
- Web-Access

Speaker notes

No notes on this slide.

Advanced Handling?

- Stapel-Bearbeitungen
- Datenmodell & Metadaten-Schema
- Archiv-Langzeitspeicher
- Workflow/Task Management
- Leihgaben-Verwaltung

Speaker notes

No notes on this slide.

Zukunftssicher & Interoperabel?

- Open Source
- Import / Export
- APIs
- IIIF

Speaker notes

No notes on this slide.

Who you gonna call?

- Bezahlter Support
- “Community”

Speaker notes

No notes on this slide.

Sprache ist wichtig.

For tech: English, please? 😇

- Alle Feldbezeichnungen (title / ~~Titel~~)
- Alle Tech-Begriffe
- Dokumentation
- **aber:** UI und Feld-Labels übersetzbar

Speaker notes

No notes on this slide.

Datum/Uhrzeit Formate

Ein Klassiker. 😜

Speaker notes

No notes on this slide.

Datum/Uhrzeit Formate

Dates	Times
January 3rd, 1981	13:37
3. Jänner 1981	13:37:00
1/3/1981	0h13m37s
3.1.1981	123.4
1977-11-23-02	

Zeitzonen? Andere Kalendersysteme…?

Speaker notes

No notes on this slide.

ISO 8601. 😀

- **Datum:**
  1981-01-03
- **Datum & Zeit (+TZ):**
  1981-01-03T00:12:34$^{+02:00}$
- **Woche (+Wochentag):**
  2022-W46-5
- **Dauer:**
  PnYnMnDTnHnMnS

Siehe: (Wikipedia) ISO 8601

Speaker notes

Whatever you do, please:

- Support iso8601 for date/time **input** (if possible).
- Support exporting/importing iso8601 date/time/duration strings.

You may however, want to be able to at least display date/time also in different formats (configurable).

Why? The “psycho-visual perception” (does it look/feel good?) of people reading ISO8601 syntax is mixed…

- Standalone dates: ok. Sometimes cleaner to read than written date. Sometimes not.
- Date+Time: rather hard to grasp visually. (Add a timezone and you’ve lost most operator staff)
- Duration: Forget it. Hard to read as a person.

Examples:

- 1981-01-03 (nice) vs January 3rd, 1981 (warmer)
- 1981-01-03T20:15+01:00 (uck.)

Here’s a link to documents from the US LoC, about rules and extensions for date/time cataloging:

- LoC: WD8601-2_2016 Representation of dates and times (Part 1: Basic Rules)
- LoC: WD8601-2_2016 Representation of dates and times (Part 2: Extensions)

Dauer / Periode

- Unterstützung für “von/bis” in Datumsfeldern?
- Mit verschiedener, möglicherweise gemischter “Genauigkeit”?
  - “1920-1923”
  - “1920-01 - 1923”

Speaker notes

No notes on this slide.

Was ist mit “Fuzzy Dates” …?

- 1910-01?
- circa 1800?
- Bronce age
- WWII
- etc.

Oh, eine gemeinsame Syntax wäre gut.

Speaker notes

Seems trivial, but it makes sense to support fuzzy date/time.

To my knowledge, there’s no standard for fuzzy-dates (ISO8601 doesn’t cover that).

The US Library of Congress (LoC) has issued extension rules for datetime issues, that include suggestions for approximate and uncertain date/time values, as well as general rules for handling this for cataloging:

- LoC: WD8601-2_2016 Representation of dates and times (Part 1: Basic Rules)
- LoC: WD8601-2_2016 Representation of dates and times (Part 2: Extensions)

Teilweise Ungefähr Unsicher?

- “**?**”: Unsicher
- “**~**”: Ungefähr (circa)
- “**%**”: Beides: Unsicher & ungefähr
- “**X**”: Unspecified

1981-01?, 20XX, 197X-11?-XX
(*https://www.loc.gov/standards/datetime/iso-tc154-wg5_n0039_iso_wd_8601-2_2016-02-16.pdf*)

Speaker notes

No notes on this slide.

Der “Identifier” (ID)

Unbedingt notwendig. Auch bekannt als:

- Object ID, Item ID, …
- Archive signature
- UID: Unique ID
- UUID: Universally Unique ID
- PID: Persistent ID
- …

Speaker notes

Questions: * For what reason(s) would you want/need an identifier? * Why is it desirable that it’s unique? * Why is it desirable that it’s persistent? * What’s the scope that ID shall be used for? This has great influence on choosing its properties, like: persistence, uniqueness, recognizability, etc.

Alle haben einen! 😉

Beispiele

- V-00815
- W/S #00034
- 38AF2EC1A13494B9DF6FD6E75960307
- FBW002984
- 8-717774-330128
- adBDwKf_aSE
- Q83697636
- …

Speaker notes

All of the above examples are patterns used for real somewhere.

PIDs - Persistent Identifiers

Beispiel: NHMW-ZOO-FS50750

> *“[PIDs] provide [means for offering] a long-lasting click-able link to a digital object.”*
>
> (Source: https://tanc-ahrc.github.io/HeritagePIDs/

Speaker notes

No notes on this slide.

PIDs - Infos

- “PIDs 101 - Einführung in die Welt der persistenten Identifiers (by Britta Dreyer)”
- “How to implement human readable PIDs?”
- “Towards a National Collection - HeritagePIDs”
- “Heritage PID Resources”

Speaker notes

No notes on this slide.

ID: Anforderungen

- Eindeutig? Persistent?
- Stabile Syntax?
- Wiedererkennbar?
- Scope?
- Dateinamen-safe?
- Druckbar? (QR- oder Barcode-Label)

Speaker notes

No notes on this slide.

Kann’s “DAM-IT” umgehen?

- Auto-generate?
- Hochzählen?
- Validieren?
- Mehr als 1 Syntax?

Speaker notes

If you’re dealing with different entity- and object-types, that require different identifier-syntaxes: Check if your DAM supports different ID-syntax for these cases in parallel, without conflicts.

Übliche Begriffe?

> *“Des hamma immer scho so g’nannt, also isses klar. Uns.”*

- Person, Firma, Schauspielerin oder: **Agent**?
- Aufnahme, Film, Video oder: **MovingImage**?
- Aktion, Shooting oder: **Event**?
- Author oder: **Creator**?
- etc.

Können Begriffe in der UI umbenannt werden?

Speaker notes

I’m very sure, most use-cases are not unique. As said before: If someone else is collecting and cataloguing and managing similar items or has similar use-cases, please get in contact with each other.

But: As great as sharing and using common lists/vocabularies is, it brings the “problem” that you need to agree on common terms. :)

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” ;)

Be ready for trade-offs. Be ready to use different, but now common, terms.

Kontrolliertes Vokabular

> *“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.”*

Quelle: “Controlled Vocabulary” (Wikipedia)”

Speaker notes

No notes on this slide.

Wozu Listen / Taxonomien?

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 = Regisseur = Regie

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)

Taxonomien handhaben

- Klare Auswahl/Definition von Begriffen?
- Standardbegriffe?
  (e.g. ISO Listen für Sprachen, Länder, Regionen, Fachbegriffe, etc.)
- (Tastatur)kürzel um Begriffe auszuwählen?
- Verschachtelte Hierarchien möglich?
- Begriffe verwalten / ändern / hinzufügen?
- ID vs Label?

Speaker notes

No notes on this slide.

ID *und* Label 💕

Kann ihr DAM folgendes ordentlich speichern und handhaben?

Identifier(für Maschinen)	Label(Für Menschen>)	Übersetzungen?
`sq`, `sqi`, `alb`, …	Albanian	Albanisch(de), Albanais(fr)
`de`, `deu`, `ger`, …	German	Deutsch(de), Allemand(fr)
`und`	Undetermined	…?

Beispiele: ISO 639 (Sprachen), ISO 3166 (Länder),…

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.

Gemeinsame Taxonomien

- Gleiche Begriffe/Listen können in verschiedenen Systemen/Domänen verwendet werden.
- Falls möglich, gemeinsame (shared) Listen verwenden.
- Eigene Listen teilen/veröffentlichen.
- Im Austausch/Kontakt mit Anderen sein.
- Bereit sein Kompromisse einzugehen.

Speaker notes

No notes on this slide.

Dynamische Taxonomien

Wie geht ihr DAM damit um:

- Neue Begriffe (spontan) einfügen?
- Begriffe ändern, die bereits zugewiesen wurden? 😱
- Import / Export von Listen?

Speaker notes

It’s quite unrealistic to assume any term-list will not change over time. Depening on how a DAM system is designed/built under the hood, it may (or may not) handle changes to term-lists well.

Try to assign terms to records, then change/move/re-label/etc those already-assigned terms.

Some DAMs store terms by references (usually some form of identifier), others store the actual value at the time the term was assigned to a field.

Both have their pros and cons, yet with data being more and more required to become “flexible” and interoperable, it is more likely that being able to handle term-assignments dynamically will be the preference for future-proof systems.

Gemeinsame Identifier

Falls Sie nicht die einzige sind, die auf … verweist:

Wäre es dann nicht sinnvoll einen gemeinsamen, öffentlich zugänglichen Identifier zu haben?

Speaker notes

With data being exchanged and aggregated, it's becoming more interesting (and practically important) to include identifiers from different sources, to be able to relate database entries with each other.

A typical scenario is:

- in-house identifier
- additional external identifier(s), like IDs for Agents.
- external identifier from data source.
  Example: When merging multiple collections into one.

Increasingly popular:

- Store Wikidata identifier to relate to common entries.

Mehrfache IDs handhaben?

zB: Als Attribut zu deklarieren um “welche” ID es sich handelt:

```
<id schema="dnb"> ... </id>
<id schema="efg"> ... </id>
<id schema="wikidata"> Q84199 </id>
```

Speaker notes

This is a very simple “attribute” added to an identifier field. This example was taken from EN15907 standard for cataloguing cinematographic works.

Wikidata

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

Beispiele: Earth (Q2) vs Earth (Q83697636)

Speaker notes

No notes on this slide.

Datenstruktur ist relevant.

> *“klassischer” Katalog = flach, Tabellen/Zellen-Denken.*

- Besser: **Entitäten** and **Attribute** identifizieren.
- Dann: **Beziehungen** (Relationships) identifizieren
  (und deren “`@Attribute`”)

- Entity Relationship Model (ERM)
- Think in “Semantic Triples”:

```
Subject - Predicate - Object
```

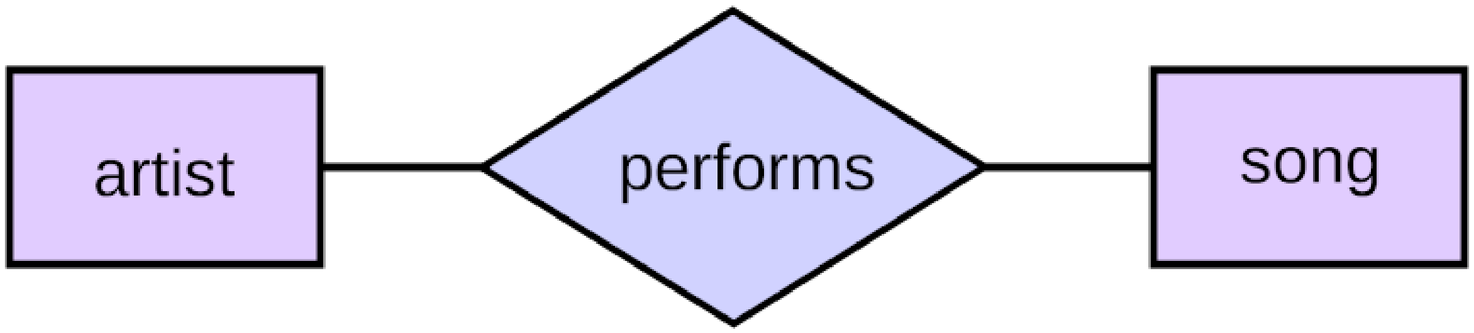
Beispiel: elgrito.witness.org

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

Relationships



- **Forward**: Agent1 `ist Elternteil` von Agent2
- **Reverse**: Agent2 `ist Kind` von 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.

Relationships beschreiben

- Für welche Zeitperiode war die Relation gültig?
  (Agent “ist verheiratet” mit Agent)

- Noch etwas, dass Sie gerne anmerken würden?

Speaker notes

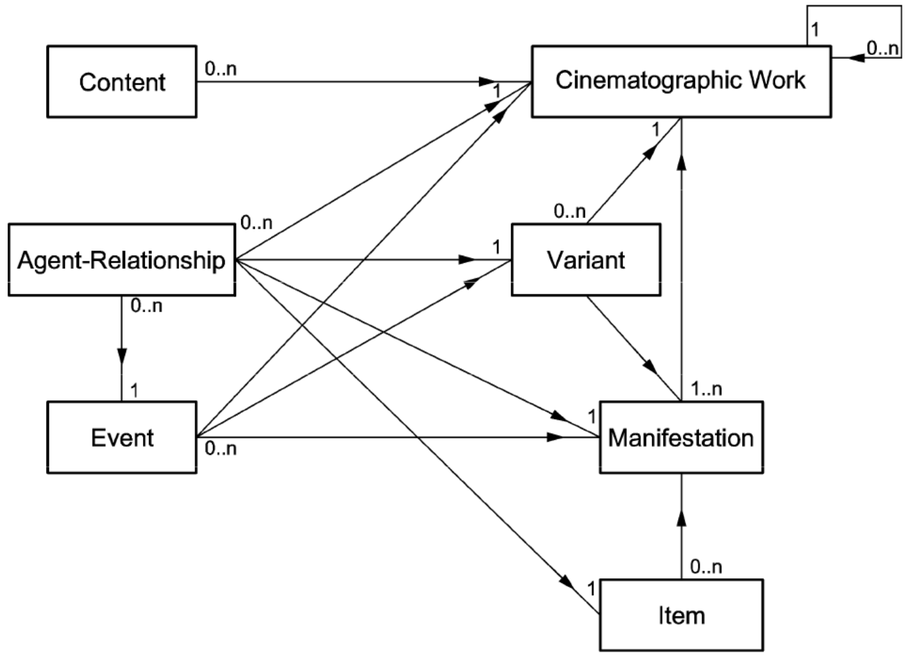
Relationships: Abfragen?

(Wie gut) kann ihr System Relationship-Details handhaben/suchen?

- In welche Werken war Agent unter dem Alias … mitwirkend als …?
- Liste alle … die in Beziehung zu … stehen, sortiert nach …?

Speaker notes

Mehr als nur 1 Standard/Schema?



Super! Was machen wir jetzt mit “film-related” Material? 😐 (Poster, Dokumente, etc)

Speaker notes

No notes on this slide.

Import / Export

Speaker notes

No notes on this slide.

Import / Export: Überlegungen

- (Nested) Listen & Vokabulare?
- Link zwischen Katalogdaten (DB) & Files? (Storage)
- Programmier(sprachenkenntnisse) benötigt?
- Value/Term/Schema Mapping Optionen?
- Metadata-Layout (Schema) valide?

Speaker notes

Referenz zwischen Katalog und Files

- Wie wird die Verknüpfung zwischen Katalogeinträgen und Dateien hergestellt (+vice versa)?
- Wie (un)abhängig vom Katalog können Dateien benutzt/geändert werden?

Speaker notes

No notes on this slide.

Dateinamen & Ordnerstruktur

- Viele DAMs vergeben diese automatisch.
- Auf nicht-menschlich lesbare Art (UUIDs).
- Welches Level an Kontrolle benötigen Sie hier?
- Gibt es externe Workflows die auf die Dateien zugreifen?

Speaker notes

No notes on this slide.

Files intakt, aber DAM kaputt?

Speaker notes

It’s good practice to somehow embed vitally necessary metadata with/inside your files that may help to identify and (re-)assign them if they got “lost” in the wild.

For example, having its “title” and some form of persistent identifier (PID) with your files, makes your data setup more robust against critical database or other failures.

See “Heritage PID Resources” for more information on how to create and use PIDs.

Zugriff & Kontrolle

- (Web) APIs: Application Programming Interfaces
- REST API:
- OAI-PMH: Protocol for Metadata Harvesting
- …

Speaker notes

No notes on this slide.

APIs: Wozu?

Zugriff & Austausch Ihrer Daten…

- … mit anderen Institutionen.
- … mit externen Tools.
- … für Backup/Migration.
- … um was völlig Anderes damit zu machen.
- etc.

Speaker notes

No notes on this slide.

APIs: Wozu noch?

Fernsteuern!

- Gemeinsames Backend + verschiedene UIs.
- Workflows/Aktionen extern automatisieren.
- etc.
- **Und:** Einfach (mehr) zukunftssicher. 😎 🤓

Speaker notes

No notes on this slide.

“Does it scale?”

Bitte mit erfahrenen Tech-Leuten besprechen.

- Geschwindigkeit? (search, retrieve, data, files, etc)
- Cache & Tuning Optionen?
- Mehrere Instanzen synchronisieren?
- Laufender Import/Export? (in-house, extern)
- Lizenzbedingungen und Auswirkungen?
- Mehrere Datenschemata parallel?
- Mehrsprachigkeit?
- Migration auf ein “besser skalierendes” System (falls notwendig)?

Speaker notes

No notes on this slide.

Ham's Plugins?

- Erfahrungsgemäßg ist ein "All-in-One" monolithisches System starr und aufwendig zu erhalten.
- Bevorzugt: Schlanker Kern + zusätzliche Features als Plugins/Module.
- Notwendig: Plugin-Interfaces offen & dokumentiert.

Speaker notes

No notes on this slide.

Ham's Support?

- Wen kann man (an)rufen im Falle von …?
- Möglichkeit(en) "Ihren Mechaniker" zu wechseln?
- Zu welchen Bedingungen?
- Supportverträge?
- Welche Lizenzmodelle?
- Ihr Personal schulen?
- Ordentliche Dokumentation?

Speaker notes

No notes on this slide.

Hat sich gut bewährt:

- Zugriff auf eine Demoinstallation (lokal oder online).
- Versuchen Sie Listen/Vocs zu importieren / exportieren.
- Testen Sie Import/Export/UI-Feel mit *Ihren* Daten (Layout). (*Bevor* Sie bindende Entscheidungen treffen)
- Katalog- & Technikleute zusammenbringen.
- Hilfe von Expertinnen (des Systems/Anwendungsfälle Ihrer Wahl) einholen.
- **So einfach wie möglich, so kompliziert wie notwendig.**

Speaker notes

No notes on this slide.

Und ja: Bitte Folgendes einplanen...

- Backups?
- Failover System?
- Supportverträge?
- Migrationsanforderungen-und-Hürden bekannt?

Speaker notes

No notes on this slide.

- Ende -

Speaker notes

No notes on this slide.

Fragen?

Kommentare?

info@kulturerbe-digital.at

Peter Bubestinger-Steindl

peter @ ArkThis.com

Speaker notes

No notes on this slide.

Oh, btw: “Says who?”

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

Speaker notes

No notes on this slide.

Oh, there’s a typo…

- Should we correct it?
- Is it even a typo, or was it proper “back then”?
- Can we keep multiple “versions”?
- (Where) can we document that? (commit msg)

Speaker notes

No notes on this slide.