

# Metadaten-Grundlagen

## Arten von Datenformaten

Jakob Voß

KIM Workshop, Mannheim, 2023-05-08

# Was ist eigentlich ein Datenformat?

- ▶ Formale Regeln (Schemas)
  - ▶ Menge von *möglichen* Dokumenten (meist unendlich)
  - ▶ Basieren auf einer Strukturierungssprache
  - ▶ Syntax
- ▶ Weitere implizite oder semi-formale Regeln
  - ▶ Was ist erlaubt/nicht erlaubt?
  - ▶ Teils in Software versteckt
- ▶ Semantik
  - ▶ Das was Menschen erwarten
- ▶ Praxis
  - ▶ Wie es tatsächlich angewandt wird

# Arten von Datenformaten

# Einordnung von Datenformaten

- ▶ Nach Einsatzzweck
- ▶ Nach Herkunft
- ▶ Nach Eigenschaften
- ▶ Nach Beziehungen untereinander
- ▶ ...

# Datenformate nach Einsatzzweck

*Fragen: Wie lassen sich die Datenformate nach Einsatzzweck gruppieren? Wofür werden Datenformate angewandt?*

*Gruppen bilden in Gruppen!*

# Bibliographische Datenformate (Metadatenformate)

Beschreiben Publikationen oder andere Dokumente

- ▶ MARC
- ▶ PICA
- ▶ ONIX
- ▶ BibTeX
- ▶ ...

Beschreiben Entitäten (Personen, Orte, Organisationen, Themen...)

- ▶ MARC
- ▶ PICA
- ▶ SKOS
- ▶ JSKOS
- ▶ ...

# Dokumentformate

## Enthalten Dokumente

- ▶ HTML
- ▶ LaTeX
- ▶ TEI
- ▶ ...



## Strukturieren Daten in kleinere Einheiten

- ▶ JSON
- ▶ XML
- ▶ CSV
- ▶ *Excel?*
- ▶ *PICA?*
- ▶ ...

Dienen zur Abfrage/Referenzierung von Teilen aus Datensätzen

- ▶ XPath
- ▶ JSON Pointer
- ▶ jq
- ▶ SQL
- ▶ MARCspec
- ▶ ...

*Beziehen sich immer auf eine Strukturierungssprache!*

Legen (und beschreiben) eine Teilmenge einer Datenstrukturierungssprache fest: formale Definition eines Datenformats.

- ▶ JSON Schema
- ▶ XML Schema
- ▶ Avram
- ▶ ...

*Beziehen sich immer auf eine Strukturierungssprache!*

# Containerformate

- ▶ Gepacktes Verzeichnis (ZIP, TAR...)
- ▶ Die meisten Dokumentformate
- ▶ ...

# Wie hängen Datenformate zusammen?

- ▶ Kodierungen
  - ▶ PICA/JSON, PICA+, PICA/XML...
  - ▶ MARC/JSON, MARC/XML, MARC Maker, ISO MARC...
- ▶ Formatbeschreibungen
- ▶ Schemas
- ▶ ...