

Metadaten-Grundlagen

Datenformate

Jakob Voß

KIM Workshop, Mannheim, 2023-05-08

Was gibt es für Datenformat?

Welche (Meta)datenformate kennt ihr?

1. Frage: Welche Datenformate sind das oder könnten das sein?

- ▶ Wieder in Kleingruppen
- ▶ Mehrfachantworten sind möglich
- ▶ Bitte leserliche Karteikarten
- ▶ Umhängen ist erlaubt
- ▶ Bitte nicht einfach im Internet nachschlagen!

2. Frage: Welche relevanten Datenformate fehlen?

Beispiele von Datenformaten

Tab-Separated Values

- ▶ Tabellarisches Format
- ▶ Zeilen und Spalten
- ▶ Zeilen getrennt durch Zeilenumbruch
- ▶ Spalten getrennt durch Tabulator (Bytecode 09)

Vor- und Nachteile?

JavaScript Object Notation (JSON)

- ▶ Inspiriert von JavaScript
- ▶ Inzwischen wichtigstes allgemeines Strukturierungsformat
- ▶ Definiert unter <http://json.org>
- ▶ Elemente: Objekt, Array, String, Zahl, Boolean, Null
- ▶ Grundlage zahlreicher Datenformate und Kodierungen
 - ▶ JSKOS
 - ▶ BibJSON
 - ▶ PICA/JSON
 - ▶ ...

Einige JSON Werkzeuge

Webanwendungen

- ▶ <https://jsoneditoronline.org/>
- ▶ <https://jqplay.org/>

Lokal

- ▶ Texteditor mit Syntax-Highlighting
- ▶ jq (Kommandozeile)

Datenformate rund um JSON

- ▶ Programmiersprache: jq
- ▶ Abfragesprachen: JSONPath, SQL/JSON path, JSON Pointer...
- ▶ Schemasprachen: JSON Schema...
- ▶ Zahlreichen Kodierungen *in* JSON: JSON-LD, PICA/JSON...

Datenformate nach Einsatzzweck

3. 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...)
- ▶ METS/MODS?
- ▶ ... # Vertiefende Beispiele
- ▶ PICA
- ▶ MARC
- ▶ JSON
- ▶ PICA/JSON

Arten von Datenformaten (I)

Datenstrukturierungssprachen

- ▶ Internes Datenformat der Bibliothekssysteme von OCLC PICA
- ▶ Feld-Unterfeld-Struktur (oder: Ebenen)
- ▶ Pica3/PICA+
- ▶ Verschiedene konkrete Anwendungsprofile
 - ▶ K10Plus Katalogisierungsformat
 - ▶ GND-Internformat
 - ▶ ...
- ▶ Beispiele: siehe OPAC oder WinIBW

PICA-Datenmodell

- ▶ field, tag, occurrence
- ▶ subfield, code, value

und/oder

- ▶ level 0
- ▶ level 1
- ▶ level 2

Machine-Readable Cataloging (MARC)

- ▶ Bibliothekarischer Datenaustausch seit den 1960ern
- ▶ Noch immer wichtigstes Format für Bibliotheksdaten
- ▶ *MARC Must Die* (Tennant, 2007)

<http://www.loc.gov/marc/>

Wie hängen Datenformate zusammen?

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

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
- ▶ ...