

# KIM-Workshop 2019

## Session „Anforderungen an moderne Metadatenhaltung“, Zusammenfassung

Jana Hentschke, April 2019

### Zusammengetragene Anforderungen

Schritt #2. Gruppierung nach Themenblöcken. Stichpunkte sortiert nach Häufigkeit der Nennungen.

#### Herkunft

- Datenquelle (Institution)
  - als Kriterium für die Vertrauenswürdigkeit der Daten ("Trust")
  - um Datenfehler melden zu können
  - um in nachnutzenden End-Nutzer-Systemen Herkunftsnachweis anzeigen zu können
  - um bei der Weiterverarbeitung selektieren zu können
- Entstehung der Daten
  - Prozesskennzeichnungen (administrativ)  
z.B. maschinell vs. intellektuell
    - um bei der Verarbeitung selektieren zu können
  - Bezugsweg der Daten
    - für administrative Selektionen

#### Entstehungsgrundlage

- Regeln  
z.B. Regelwerke, Algorithmen in Vorschlagssystemen
  - um Grad der Normierung ableiten zu können
  - um Verwendungskontext ableiten zu können
  - um Grad der eindeutigen, maschinenlesbaren Identifizierung von Ressourcen ableiten zu können (z.B. ISSN/p-ISSN/e-ISSN/ISBN)
  - in dem Zusammenhang: *öffentlich verfügbare, auffindbare und lesbare Regelwerksdokumentation sollten gewährleistet sein*
- Beziehung zwischen verschiedenen Repräsentationen von Metadaten ausweisen  
z.B. "Abgeleitet von"
  - um Probleme nachvollziehen zu können (administrativ)
  - um inhaltliche Redundanzen absehen zu können
- Ausgangsmetadatenformat

- um auf Interoperabilität und Strukturen rückschließen zu können
- Sprache der Metadaten(inhalte)
  - um diese in datennutzenden Anwendungen berücksichtigen zu können

## Nachvollziehbarkeit von Veränderung

- Versionierung (mit referenzierbaren Versionen)
  - um Reproduzierbarkeit wissenschaftlicher Forschungsarbeit zu gewährleisten (betrifft auch andere Arten von Daten als Metadaten)
  - um Probleme nachvollziehen zu können (administrativ)
- Transparente Dokumentation der Datengrundlage und ggf. der Update-Mechanismen
  - um Redundanzen erkennen zu können
  - um Datenaktualität einschätzen zu können
- Konfiguration von Tools, die die Metadaten erzeugt haben  
*verwandt mit Prozessbezeichnungen (administrativ), s.o.*
- Akteur, der Änderungen durchgeführt hat  
z.B. Einrichtung, Softwareversion
  - um Rückschlüsse auf Qualität ziehen zu können
  - um Probleme nachvollziehen zu können

## Qualität

Ergibt sich häufig indirekt aus den Punkten in den anderen Themenblöcken, z.B.

- Vertrauenswürdigkeit der Datenquelle
  - Entstehung der Daten (z.B. maschinell vs. intellektuell), s.o.
  - Regelwerksgrundlage (z.B. Grad an Normierung vs. Freitext, Verlinkungen), s.o.
- Anhaltspunkte zur Vollständigkeit von Metadatenzusammenstellungen  
z.B. bezogen auf einen Datensatz
  - um Datenübernahmeentscheidungen treffen zu können
- Konfidenzwerte
  - um Zuverlässigkeit beurteilen zu können

## Lizenzinformation, Rechte

- Lizenzen / Nutzungsbedingungen der Metadaten
  - um Metadatennachnutzern die Selektion zu ermöglichen

# Matrix

Analyse der von den Session-TeilnehmerInnen kooperativ befüllten Matrix aus Schritt #3

<http://data.bib.uni-mannheim.de/kimws19/matrix> (Kann noch weiter befüllt werden)

Leider blieb zu wenig Zeit am Session-Ende, um die Matrix systematisch zu befüllen.

Die vorhandenen Einträge deuten jedoch an, dass an den unterschiedlichen Stellen (Datenerfassungssystemen, Datenschnittstellen, Metadatenstandards, ...) Lücken in Bezug auf die zusammengetragenen Anforderungen bestehen. Teilweise sind diese auch darauf zurückzuführen, dass nicht an allen Stellen alle Anforderungen relevant sind. Zum Beispiel besteht im rein intellektuell befüllten Datenerfassungssystem vielleicht gar nicht die praktische Notwendigkeit, die für alle Daten geltende Erfassungsgrundlage explizit zu verzeichnen. Ebenso wie ein Metadatenaustauschformat nicht die Stelle ist, die die Funktion "Versionierung der Daten" bereitstellen könnte (wohl aber kann dort natürlich eine Datenversion ausgewiesen werden!).

Eventuell weisen die häufigen vagen Aussagen ("teilweise ...") darauf hin, dass in der Praxis (notgedrungen) Lösungen gefunden werden, um bestehende Anforderungen umzusetzen und sich Möglichkeiten entwickeln.

Die Gestalt der Matrix zeigt ein mögliches Vorgehen auf, um - systematisch angegangen - eine vollständigere Übersicht der aktuellen Situation zu erfassen.

