

Vom Datenkatalog zum Wissensgraph – Forschungsdaten im konzeptuellen Modell von FRBR

Ingo Frank

Leibniz-Institut für Ost- und Südosteuropaforschung,
Regensburg, Deutschland
frank@ios-regensburg.de

Abstract

Die Beschreibung von Forschungsdaten im Bereich des Forschungsdatenmanagements ist oft ungenau, unvollständig oder nicht konsistent bzw. leidet unter einer nicht konsequent durchgeführten Metadatenkuration. Der Beitrag stellt einen bibliotheks- und informationswissenschaftlich motivierten Ansatz vor, wie Metadaten über Forschungsdaten mithilfe des konzeptuellen Modells von FRBR (Functional Requirements for Bibliographic Records) verbessert werden können. Das konkrete Ziel dabei ist die Konstruktion eines Wissensgraphen, der FRBRisierte Metadaten aus einem Datenkatalog mit Metadaten aus einem Bibliothekskatalog sowie mit Forschungsinformation integriert. Die Methode baut auf einem Datenkatalog mit einem auf DCAT (Data Catalog Vocabulary) und Disco (DDI-RDF Discovery Vocabulary) basierenden Anwendungsprofil als Metadaten-schema auf. Die Metadaten im Datenkatalog werden mit SHACL (Shapes Constraint Language) validiert und dienen somit als Grundlage für die FRBRisierung zum Aufbau des Wissensgraphen mit FaBiO (FRBR-aligned Bibliographic Ontology) als FRBR-basiertem Datenmodell. Die FRBRisierten und integrierten Metadaten im Wissensgraphen unterstützen schließlich aufgrund der besseren Metadatenqualität und der Vorgehensweise zur Verlinkung von Entitäten aus Datenkatalog, Bibliothekskatalog und Forschungsinformationssystem insbesondere die Versionierung und Provenienzinformation von Forschungsdaten und nicht zuletzt auch die Datenzitation. Der FRBRisierungsansatz trägt dadurch zur Verbesserung des Information Retrieval in Discovery-Systemen für Forschungsdaten bei.

Keywords: Forschungsdatenmanagement; FRBR; Informationsintegration; Wissensgraphen; Metadatenqualität; Digital Humanities; Forschungsdateninfrastruktur

1 Einleitung

Daten bzw. Forschungsdaten werden in der Forschungsdatenmanagement-Community oft unzureichend definiert (s. Hjørland, 2018). Problematisch ist auch die verbreitete Auffassung, dass Daten die Grundlage für Information und Wissen bilden, wovon im Modell der Wissenspyramide ausgegangen wird (s. Frické, 2019). Dazu kommt die häufig vorkommende Verwechslung von Daten und (Daten-)Dokument (s. Schöpfel et al., 2020) – was sich alles negativ auf die Qualität der Metadaten zur Beschreibung von Forschungsdaten auswirken kann. Wir wollen mit unserem kurzen Arbeitsbericht (work in progress) zeigen, welchen bibliotheks- und informationswissenschaftlichen Beitrag man im konzeptuellen Rahmenwerk von FRBR (Functional Requirements for Bibliographic Records) zur Verbesserung der Metadatenqualität und des Information Retrieval im Forschungsdatenmanagement leisten kann.

Datenkataloge, die DCAT (Data Catalog Vocabulary) verwenden, stellen Metadaten über Datensätze und deren Distribution in verschiedenen Formaten oder über verschiedene Dienste bereit. Trotz Best Practices zur Forschungsdatenkuratation gelingt diese Trennung nicht immer präzise. Bibliothekskataloge stellen Metadaten über Publikationen in der Regel auf der Ebene von FRBRs *Manifestation*-Begriff bereit. Datensätze in DCAT-Datenkatalogen sind allerdings mehr oder weniger auf *FRBR-Expression*-Ebene zu verorten, was ein weiteres Problem für die Integration von Information aus Datenkatalogen und Bibliothekskatalogen darstellt. Mit der FRBRisierung der Metadaten wollen wir diese Probleme angehen und die implizit vorhandenen Unterscheidungen von Entitäten und deren Zusammenhänge explizit machen und exakt erfassen, um die Qualität der in einem Wissensgraphen integrierten Metadaten zu verbessern.

Im Rahmen des institutionellen Forschungsdatenmanagements und des Aufbaus eines institutionellen Repositoriums, das sowohl Digitalisate, Publikationen, aber eben auch Forschungsdaten einschließt, besteht unsere konkrete Anforderung darin, den Datenkatalog mit dem Bibliothekskatalog und auch mit Forschungsinformation zu integrieren. Im Prinzip geht es darum, die für das institutionelle Repositorium relevante Information in einem Wissensgraphen zu integrieren.¹ Unsere Vorgehensweise zur Integration von

¹ Hintergrund bzw. Projektkontext ist die Integration des institutionellen Forschungsdatenrepositoriums LaMBDA mit einem im Aufbau befindlichen Informationsintegra-

Metadaten über Forschungsdaten aus den verschiedenen Quellen in einem ‚Institutional Knowledge Graph‘ wird im nächsten Abschnitt dargestellt.

2 Vorgehensweise

FRBR is not a data model. FRBR is not a metadata scheme.

FRBR is not a system design structure.

It is a conceptual model of the bibliographic universe.

– Tillett (2005)

Ausgehend von einem mit SHACL (Shapes Constraint Language) validierten Datenkatalog wird dieser FRBRisiert und schließlich mit zusätzlicher Information aus dem Bibliothekskatalog und dem Forschungsinformationssystem angereichert und in einem Wissensgraphen integriert. FRBR verwenden wir dabei als konzeptuelles Modell des bibliografischen Bereichs. Als konkretes Datenmodell zur Modellierung der FRBRisierten Metadaten verwenden wir die FRBR-aligned Bibliographic Ontology (FaBiO) (Peroni/Shotton, 2012).

2.1 DCAT- und Disco-basiertes Anwendungsprofil als Metadatenschema für erweiterte Datenkataloge

Zur Beschreibung von Forschungsdaten im institutionellen Forschungsdaten-repositorium haben wir ein im Wesentlichen auf DCAT basierendes Anwendungsprofil entworfen. Wesentliche Grundlage des Anwendungsprofils sind die Klassen `dc:Dataset` und `dc:Distribution` und deren für uns relevante Eigenschaften, die aber keine Information über Erhebungsmethoden etc. vorsehen. Heery und Patel (2000) definieren Anwendungsprofile als Metadatenschemas, “which consist of data elements drawn from one or more namespaces, combined together by implementors, and optimised for a particular local application”. Damit wir demnach methodologische Information über die Erhebung bzw. Erstellung der Forschungsdaten erfassen können, haben wir das Metadatenschema um Eigenschaften aus Disco (DDI-RDF Discovery Vocabulary) (Bosch et al., 2013) erweitert. Disco legt den Schwer-

punkt auf die Unterstützung der für das Information Retrieval in Discovery-Systemen relevanten Elemente aus dem DDI-Metadatenstandard.

Um beim Aufbau unseres Metadatenschemas keine ‚Frankenstein-Ontologie‘ (vgl. Corcho et al., 2015) zu entwerfen, bei der – womöglich lediglich von der Bezeichnung her passende – Klassen und Eigenschaften aus bereits vorhandenen Ontologien zusammengesucht werden, verwenden wir SHACL zur Vermeidung von Inkonsistenzen und zur Validierung der Metadaten. Der folgende Auszug aus den SHACL-Constraints zeigt als Beispiel die Festlegung der gemäß Forschungsdatenmanagement-Policy im Forschungsdaten-Repositorium möglichen Zugriffsrechte auf Forschungsdaten in einem *property shape*.

```
sh:property [
  sh:name "Availability of Data" ;
  sh:description "Access rights or restrictions to data
    distributions" ;
  sh:path dct:accessRights ;
  sh:class skos:Concept ;
  sh:nodeKind sh:IRI ;
  sh:in (
    <http://purl.org/coar/access_right/c_abf2>
    <http://purl.org/coar/access_right/c_flcf>
    <http://purl.org/coar/access_right/c_16ec>
    <http://purl.org/coar/access_right/c_14cb>
  ) ;
  sh:maxCount 1 ;
] ;
```

Mit dieser *property shape* wird eine bestimmte Auswahl an Zugriffsrechten aus dem *Access Rights*-Vokabular der COAR Controlled Vocabularies festgelegt. Auf ähnliche Weise können mit SHACL auch Einschränkungen formuliert werden, die festlegen, welche Wissensorganisationssysteme zum Einsatz kommen können. So haben wir etwa festgelegt, dass die Erhebungs- bzw. Erstellungsmethoden anhand der Eigenschaft `disco:collection-Mode` nicht nur mit dem auf die Sozial- und Wirtschaftswissenschaften zugeschnittenen *Mode of Collection*-Vokabular der DDI Controlled Vocabularies (Jaaskelainen et al., 2010), sondern auch mit Begriffen aus der TaDiRAH (Taxonomy of Digital Research Activities in the Humanities)² beschrieben werden können, um den (digitalen) Geisteswissenschaften entgegenzukommen.

2 TaDiRAH im Skosmos-Browser des ACDH-CH: <https://vocabs.dariah.eu/tadirah2/en/>

Wir gelangen so zu einer interdisziplinären Wissensorganisation – im Prinzip einer Facettenklassifikation – zur Erfassung und Beschreibung der Untersuchungsgegenstände, Erhebungs-/Erstellungsmethoden sowie außerdem dem theoretischen und disziplinären Hintergrund. Diese methodologische Information geht über die üblichen bibliografischen Angaben hinaus, kann aber anhand entsprechender SHACL-Festlegungen auch auf die – im nächsten Schritt – FRBRisierten Metadaten, d.h. `fabio:DataFile` auf FRBR-Expression-Ebene, eingebunden werden.

2.2 FRBRisierung des Datenkatalogs mit FaBiO

Zur FRBRisierung der DCAT-Entitäten `dcat:Dataset` und `dcat:Distribution` wenden wir FaBiO an.³ FaBiO ist gerade bei der Integration in einen Wissensgraphen deutlich einfacher zu handhaben (s. Abb. 1) als die Ontologie FRBRoo (Functional Requirements for Bibliographic Records object-oriented)⁴, die auf ereignisbasierter Modellierung mit CIDOC CRM (Doerr, 2003) basiert.

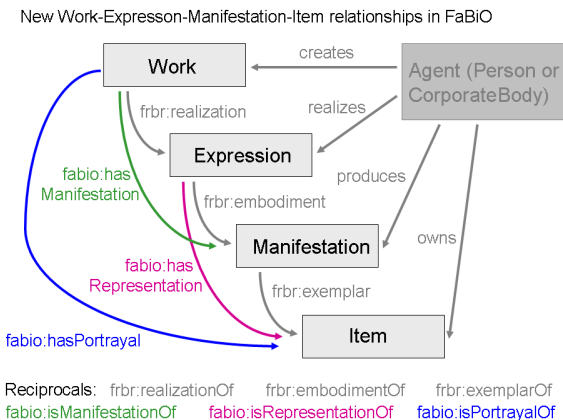


Abb. 1 Die FRBR-Konzepte *Work*, *Expression*, *Manifestation* und *Item* (WEMI-Entitäten) in FaBiO (Diagramm aus der Online-Dokumentation unter <http://purl.org/spar/fabio/2019-02-19>)

- 3 Das Metadatenschema von DataCite hat diese Unterscheidung übrigens nicht, was die Indexierung von DCAT-basierten Datenkatalogen in DataCite auf die Datensatz-Ebene (bzw. FRBR-Expression-Ebene) einschränkt. Dieses Problem soll hier aber nicht weiter behandelt werden.
- 4 Siehe OWL-Implementierung von FRBRoo unter <http://erlangen-crm.org/efrbroo>.

Genau genommen sollte eine `dcat:Distribution` die Daten eines Datensatzes repräsentieren. Bei mehreren Datendistributionen kann sich somit das Datenformat unterscheiden, aber nicht der Inhalt, d. h. die Daten selbst.⁵ Distributionen werden über die Eigenschaft `dcat:distribution` mit dem Datensatz, den sie in verschiedenen Formaten repräsentieren, verbunden. Verbindungen zu anderen Publikationen oder auch zur Datendokumentation sollten über die Eigenschaft `dcat:qualifiedRelation` jeweils als `dcat:Relationship` explizit gemacht werden.

Eine `dcat:Distribution` kann daher als `fabio:DigitalManifestation` (Unterklasse von `fabio:Manifestation`) aufgefasst werden. Wie bereits im vorigen Abschnitt angedeutet, betrachten wir ein `dcat:Dataset` mitsamt den zugehörigen bibliografischen, aber auch methodologischen Metadaten als `fabio:DataFile`. Der Klassenname ‚DataFile‘ ist etwas irreführend: Ein `fabio:DataFile` ist jedoch eine Unterklasse von `fabio:Expression` und als “realisation of a `fabio:Dataset` (a `frbr:Work`) containing a defined collection of data with specific content and possibly with a specific version number, that can be embodied as a `fabio:Digital Manifestation` (a `frbr:Manifestation` with a specific format)”⁶ definiert.

Ein typischer Anwendungsfall für den Ansatz ist die Versionierung von Datensätzen. Wie in Abbildung 2 zu sehen ist, ergibt sich durch den Einsatz von FRBR-Klassen – z. B. durch die Einbindung von `dcat:Dataset` und `dcat:Distribution` als Unterklassen von `fabio:DataFile` und `fabio:DigitalManifestation` – der Vorteil, dass der Zusammenhang von Versionen eines Datensatzes (als Werk) explizit gemacht werden kann. Das ist insbesondere für das Information Retrieval in einem Discovery-System relevant, um dem Benutzer in den Suchergebnissen oder der Navigation den Zugriff auf alle Versionen eines Datensatz zu ermöglichen. Bei der FRBRisierung des Datenkatalogs können die *Work*-Entitäten (`fabio:Dataset`) anhand der typisierten Relationen zwischen den versionierten Datensätze (`dcat:Dataset`) abgeleitet und generiert werden.

5 Vgl. dazu die DCAT-Dokumentation unter <https://www.w3.org/TR/vocab-dcat-2/#Class:Distribution>.

6 Siehe FaBiO-Dokumentation unter <https://sparantologies.github.io/fabio/current/fabio.html#d4e2752>.

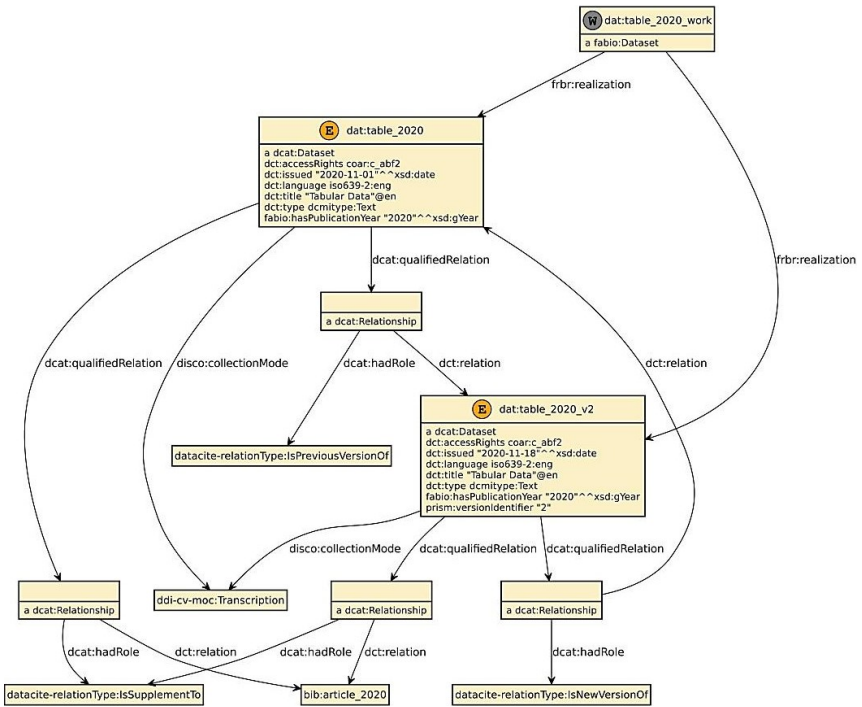


Abb. 2 DCAT hat keine Klasse zur Repräsentation eines Werkes im Sinne von FRBR. Deshalb kann in DCAT ein Datensatz nicht als Entität auf FRBR-Work-Ebene identifiziert werden, der z. B. in zwei Expressions realisiert ist, wie etwa zwei Versionen des Datensatzes. (Die Metadaten im Diagramm sind aus Platzgründen und zur übersichtlicheren Visualisierung nicht vollständig dargestellt. Es fehlen z. B. eigentlich notwendige Angaben zu Autoren, Sacherschließung, DOIs, etc.)

Einen vergleichbaren Ansatz zur FRBRisierung von Datenkatalogen hat bereits McCusker et al. (2012) vorgeschlagen. Die Autoren wollten damit die Provenienzinformation für integrierte offene Verwaltungsdaten verbessern.⁷ Sie haben dazu die FRBR-Konzepte angewendet, um zwischen der Transformation des Inhalts und der Transformation des Formats der Daten unter-

7 Hourlé (2008) verwendet das konzeptuelle Model von FRBR, um die Verwechslung zwischen Daten und digitaler Repräsentation der Daten bei wissenschaftlichen Daten – also Forschungsdaten – aufzuklären. Klump et al. (2020) schlagen einen aktuellen Ansatz vor, um mit FRBR ebenfalls die Metadaten über die Versionierung von Forschungsdaten zu verbessern.

scheiden zu können (s. Abb. 3). McCusker et al. (ebd., S. 57) beschreiben und begründen ihren FRBRisierungsansatz wie folgt:

FRBR's four levels of abstraction also naturally apply to electronic information sources. Copies of files (Items) are exemplars of the same Manifestation (byte sequence). Similarly, an Excel file created from a CSV will have a different Manifestation, but maintain the same Expression because they both store the same data. Finally, if the data is modified, the original and resulting spreadsheets have different Expressions (visual content) of the same abstract Work.

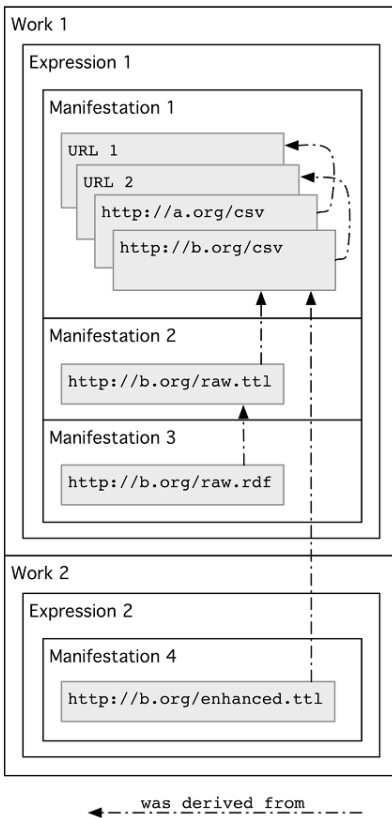


Abb. 3 Die WEMI-Entitäten von FRBR ermöglichen z.B. die Repräsentation eines Datensatzes (Work 1) und seiner (digitalen) Manifestationen sowie eines weiteren Datensatzes (Work 2) mit neuem *Inhalt* (Expression 2 von Work 2), der durch die erweiterte Manifestation 4 verkörpert wird. Diese liegt im Turtle-Format vor und wurde von der CSV-Repräsentation des ersten Datensatzes abgeleitet, was im Beispiel durch die *was derived from*-Relation ausgedrückt wird (Diagramm aus McCusker et al. 2012).

2.3 Integration von Datenkatalog, Bibliothekskatalog und Forschungsinformation in einem Wissensgraph

Nach der Definition von Ontotext (2020) repräsentiert ein Wissensgraph “a collection of interlinked descriptions of entities – real-world objects and events, or abstract concepts (e.g., documents)”. Mit Wallis (2013) kann man daher unsere Vorgehensweise auch sehr treffend mit “From Cataloging to Catalinking” bezeichnen. Unser Ansatz zur Konstruktion des Wissensgraphen ist somit die Verlinkung der relevanten Entitäten, die mit verschiedenen Ontologien repräsentiert werden:

- DCAT u. Disco: Datensätze und deren Datendistributionen im Datenkatalog
- BIBO u. FRBR: Publikationen im Bibliothekskatalog
- FRAPO u. FOAF: Forschungsinformation über Projekte, Drittmittelförderung, Forschungoutput, etc.
- FaBiO: FRBRisierte Entitäten im Wissensgraph.

Forschungsaktivitäten im Forschungsdatenlebenszyklus, wie z.B. die für Digital-Humanities-Projekte typischen Digitalisierungsprozesse, können auf einfache Weise mit der PROvenance Interchange Ontology (PROV-O) (Lebo et al., 2013) als `prov:Activity` dokumentiert werden. Die Forschungsinformation kann im Kontext des Forschungsdatenmanagements noch granularer aufbereitet werden, indem etwa die Scholarly Ontology (SO) (Pertsas/Constantopoulos 2017) zur ereignisbasierten Modellierung zum Einsatz kommt. Der Life Cycle of Historical Information (Meroño-Peñuela et al., 2014) aus dem Bereich der Historical Information Science (s. ebd.) ist auf typische Digital-History-Projekte zugeschnitten und eignet sich dabei als fachspezifischer Forschungsdatenlebenszyklus zusammen mit der TaDiRAH zur Klassifikation der Forschungsaktivitäten und -methoden besser für das Forschungsdatenmanagement entsprechender Projekte als der allgemeine Forschungsdatenlebenszyklus.

Abschließend betrachtet kann aufgrund der FRBRisierten Metadaten zu den Datensätzen die Datenzitation wesentlich genauer erfolgen, indem auf die im Wissensgraph integrierte Information aus Datenkatalog, Bibliothekskatalog und Forschungsinformationssystem zurückgegriffen werden kann (s. Abb. 4).

Abb. 4 Beispiel zu FRBRisierten und integrierten Metadaten aus Datenkatalog (Präfix `dat:`), Bibliothekskatalog (Präfix `bib:`) und Forschungsinformationssystem (Präfix `ris:`) mit verbesserter Provenienzinformation als Grundlage für genauere Datenzitation (Im Beispiel wird davon ausgegangen, dass die Aufbereitung des GREG-Datensatzes als Linked Data noch kein neues Werk ausmacht, sondern lediglich eine neue FRBR-Expression bildet. Die Frage nach der Schöpfungshöhe, die ein neues Werk ausmacht, ist ggf. individuell zu diskutieren.) →

3 Zusammenfassung und Ausblick

Wir haben gezeigt, wie mithilfe des konzeptuellen Modells von FRBR und FaBiO als FRBR-basiertem Datenmodell ein ‚Institutional Knowledge Graph‘ aus Metadaten aus Datenkatalog, Bibliothekskatalog und Forschungsinformationssystem konstruiert werden kann. FRBR dient dabei beim Übergang von Metadaten zu Wissensgraphen nicht nur als konzeptuelles Rahmenwerk, sondern quasi auch als Werkzeug zur Begriffsklärung – um auf den Beginn der Einleitung zurückzukommen: Was sind ‚Forschungsdaten‘? In welchen ‚Dokumenten‘ werden Forschungsdaten beschrieben und repräsentiert?

Der Mehrwert der bibliotheks- und informationswissenschaftlich informierten FRBRisierung und Integration von Metadaten in einem Wissensgraphen besteht zusammengefasst in

- der Möglichkeit zur Erfassung von Versionen (d. h. *FRBR-Expressions*) eines Datensatzes (als *Werk*) für das Information Retrieval,
- der erweiterten Information über den Entstehungskontext und die Quellen von Forschungsdaten anhand der Modellierung verschiedener – z. B. in verschiedenen Digitalisierungsprozessen erstellter – digitaler *Manifestationen* eines Datensatzes und
- nicht zuletzt der Verbesserung der Datenzitation aufgrund von detaillierterer Information über Datensätze und deren verschiedene *Datendistributionen* bzw. *-manifestationen*.

Als Implikation für die Entwicklung von Forschungsdatenmanagementsystemen ist allerdings festzuhalten, dass erst auf Semantic-Web-Basis eine direkte Verwendung der Metadatenstandards DCAT, Disco, FaBiO, etc. und die konsequente Nutzung von einheitlichen SKOS-Vokabularen zur interdisziplinären Wissensorganisation möglich ist.⁸ Dadurch wird die automatische Evaluierung der Konsistenz und Vollständigkeit der Metadaten mit SHACL ermöglicht.⁹ Ebenso kann die aus verschiedenen Quellen zusammengetragene

8 Eine Datenmanagementplattform, die direkt Semantic-Web- und Linked-Data-Technologie unterstützt, wird derzeit von Kirstein et al. (2020) entwickelt.

9 Die Wichtigkeit der Erfüllung der klassischen Qualitätskriterien für Metadaten – Korrektheit, Vollständigkeit und Konsistenz (s. Park 2009) – zeigt sich insbesondere auch in der gegenwärtigen Pandemie in der Schwierigkeit, aufgrund von fehlenden, inkonsistenten oder gar falschen Metadaten, gezielt nach relevanten klinischen Stu-

ne Information im Wissensgraphen mit SHACL evaluiert werden. Wissensgraphen eröffnen schließlich auch Möglichkeiten zur Verbesserung der Usability. Sehr naheliegend ist dazu die Verknüpfung von Forschungsdatensätzen und in Beziehung stehender Publikationen mittels eines interaktiven Visualisierungswerkzeugs zur Forschungsdatenkuration.¹⁰

Literatur

- Bosch, T.; Cyganiak, R.; Gregory, A.; Wackerow, J. (2013): DDI-RDF Discovery Vocabulary: A Metadata Vocabulary for Documenting Research and Survey Data. In: *Proceedings of the WWW2013 Workshop on Linked Data on the Web. CEUR Workshop Proceedings*. <http://ceur-ws.org/Vol-996/papers/ldow2013-paper-12.pdf>
- Corcho, O.; Poveda-Villalón, M.; Gómez-Pérez, A. (2015): Ontology Engineering in the Era of Linked Data. *Bulletin of the Association for Information Science and Technology* 41 (4), 13–17. doi: [10.1002/bult.2015.1720410407](https://doi.org/10.1002/bult.2015.1720410407)
- Doerr, M. (2003): The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine* 24 (3), 75–92.
- Frické, M. (2019): The knowledge pyramid: the DIKW hierarchy. *Knowledge Organization* 49 (1), 33–36.
- Heery, R.; Patel, M. (2000): Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne* (25). <http://www.ariadne.ac.uk/issue/25/appprofiles/>
- Hjørland, B. (2018): Data (with big data and database semantics). *Knowledge Organization* 45 (8), 685–708.
- Hourclé, J. A. (2008): FRBR applied to scientific data. *Proceedings of the American Society for Information Science and Technology* 45 (1), 1–4. doi: [10.1002/meet.2008.14504503102](https://doi.org/10.1002/meet.2008.14504503102)

dien oder Daten (z. B. für die Durchführung von Metastudien) zu suchen (vgl. Schriml et al., 2020). In Anlehnung an den Artikel „Medical Usability: How to Kill Patients Through Bad Design“ von Nielsen (2005) kann man das Problem mangelhafter Metadatenqualität hier treffend mit ‘How to Kill Patients Through Bad Metadata’ beschreiben.

¹⁰ Eine entsprechende Implementierung scheint es bisher ansatzweise nur für das Datenmanagementsystem CKAN zu geben: Milić et al. (2015) schlagen die Architektur Linked Relations (LIRE) zur visuellen Erstellung von Verknüpfungen von Datensätzen vor.

- Jaaskelainen, T.; Moschner, M.; Wackerow, J. (2010): Controlled Vocabularies for DDI 3: Enhancing Machine-Actionability. *IASSIST Quarterly* 33 (1–2). <https://iassistquarterly.com/index.php/iassist/article/view/649>
- Kirstein, F.; Stefanidis, K.; Dittwald, B.; Dutkowski, S.; Urbanek, S.; Hauswirth, M. (2020): Piveau: A Large-Scale Open Data Management Platform Based on Semantic Web Technologies. In: A. Harth et al. (Hrsg.): *The Semantic Web – 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings* (S. 648–664). Cham: Springer International Publishing. doi: [10.1007/978-3-030-49461-2_38](https://doi.org/10.1007/978-3-030-49461-2_38)
- Klump, J.; Wyborn, L.; Downs, R.; Asmi, A.; Wu, M.; Ryder, G.; Martin, J. (2020): Principles and best practices in data versioning for all data sets big and small. Revised Report of the Research Data Alliance Data Versioning Working Group. doi: [10.15497/RDA00042](https://doi.org/10.15497/RDA00042)
- Lebo, T.; Sahoo, S.; McGuinness, D.; Belhajjame, K.; Cheney, J.; Corsar, D. ... Zhao, J. (2013): PROV-O: The PROV Ontology. W3C Recommendation. World Wide Web Consortium. <http://www.w3.org/TR/2013/REC-prov-o-20130430/>
- McCusker, J.; Lebo, T.; Chang, C.; McGuinness, D. L.; McGuinness, D. (2012): Parallel Identities for Managing Open Government Data. *IEEE Intelligent Systems* 27 (3), 55–62. doi: [10.1109/MIS.2012.5](https://doi.org/10.1109/MIS.2012.5)
- Meroño-Peñuela, A.; Ashkpour, A.; van Erp, M.; Mandemakers, K.; Breure, L.; Scharnhorst, A. ... van Harmelen, F. (2014): Semantic Technologies for Historical Research: A Survey. *Semantic Web* 6, 539–564.
- Milić, P.; Veljković, N.; Stoimenov, L. (2015): Linked Relations Architecture for Production and Consumption of Linksets in Open Government Data. In: *Open and Big Data Management and Innovation. 14th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2015* (S. 212–222). Cham: Springer International Publishing. doi: [10.1007/978-3-319-25013-7_17](https://doi.org/10.1007/978-3-319-25013-7_17)
- Nielsen, J. (2005): Medical Usability: How to Kill Patients Through Bad Design. <https://www.nngroup.com/articles/medical-usability/>
- Ontotext (2020): What is a Knowledge Graph? Ontotext Knowledge Hub. Ontotext Fundamentals. <https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/>
- Park, J.-R. (2009): Metadata Quality in Digital Repositories: A Survey of the Current State of the Art. *Cataloging and Classification Quarterly* 74 (3–4), 213–228. doi: [10.1080/01639370902737240](https://doi.org/10.1080/01639370902737240)
- Peroni, S.; Shotton, D. (2012): FaBIO and CiTO: Ontologies for Describing Bibliographic Resources and Citations. *Journal of Web Semantics First Look* 17, 33–43. doi: [10.1016/j.websem.2012.08.001](https://doi.org/10.1016/j.websem.2012.08.001)

- Pertsas, V.; Constantopoulos, P. (2017): Scholarly Ontology: modelling scholarly practices. *International Journal on Digital Libraries* 18, 173–190. doi: [10.1007/s00799-016-0169-3](https://doi.org/10.1007/s00799-016-0169-3)
- Schöpfel, J.; Farace, D.; Prost, H.; Zane, A.; Hjørland, B. (2020): Data documents. In: B. Hjørland, C. Gnoli (Hrsg.): *Encyclopedia of Knowledge Organization*. http://www.isko.org/cyclo/data_documents
- Schriml, L. M.; Davies, M. C. N.; Eloë-Fadrosh, E. A.; Finn, R. D.; Hugenholtz, P.; Hunter, C. I. ... Walls, R. (2020): COVID-19 pandemic reveals the peril of ignoring metadata standards. *Scientific Data* 7. doi: [10.1038/s41597-020-0524-5](https://doi.org/10.1038/s41597-020-0524-5)
- Tillett, B. (2005): FRBR and Cataloging for the Future. *Cataloging and Classification Quarterly* 39 (3), 197–205.
- Wallis, R. (2013): Linked Data for Libraries: Great Progress, but What Is the Benefit? In: *SWIB13 – Semantic Web in Libraries Conference*.

In: T. Schmidt, C. Wolff (Eds.): Information between Data and Knowledge. Information Science and its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021), Regensburg, Germany, 8th–10th March 2021. Glückstadt: Verlag Werner Hülsbusch, pp. 333–346. DOI: doi.org/10.5283/epub.44952.