

Expertenteam RDA-Anwendungsprofil  
für die verbale Inhaltserschließung

# **Qualitätskriterien und Qualitätssicherung in der inhaltlichen Erschließung – Thesepapier des *Expertenteams* *RDA-Anwendungsprofil für die verbale Inhaltserschließung (ET RAVI)***

## **Vorbemerkung**

In der Folge eines Arbeitsauftrags des Standardisierungsausschusses wurde im Herbst 2017 das Expertenteam *RDA-Anwendungsprofil für die verbale Inhaltserschließung* (ET RAVI) eingerichtet. Dieses ist der Fachgruppe Erschließung zugeordnet und beschäftigt sich seither mit der Weiterentwicklung der verbalen Inhaltserschließung.

Eines der zu bearbeitenden Themen war die „Definition allgemeiner verfahrensunabhängiger Qualitätskriterien und ggf. Verfahren zur Qualitätssicherung“.<sup>1</sup> Auf einem eintägigen Workshop diskutierte die Gruppe die beiden Leitfragen *Welche Dimensionen von Qualität gibt es?* und *Welche Anwendungsszenarien können unterschieden werden?* Die Überlegungen wurden in mehreren Telefonkonferenzen vertieft. Einige Mitglieder des Expertenteams bearbeiteten weitere Details und erstellten ein Thesepapier, das dem Standardisierungsausschuss vorgelegt wurde. Die Ergebnisse wurden außerdem bei verschiedenen Gelegenheiten einem breiteren Fachpublikum vorgestellt.<sup>2</sup>

Das Ende 2019 dem Standardisierungsausschuss vorgelegte Thesepapier wird hier in einer im Januar 2021 leicht redigierten Form vorgelegt. Folgende Personen waren an der Erarbeitung beteiligt (die Verfasser:innen des Thesepapiers sind mit einem Sternchen gekennzeichnet): Regine Beckmann\* (Staatsbibliothek zu Berlin – Preußischer Kulturbesitz), Cyrus Beck (Zentralbibliothek

---

1 Vgl. Protokoll der 34. Sitzung des Standardisierungsausschusses vom 10. Dezember 2018, TOP 8c, S. 13, <https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Standardisierung/Protokolle/Standardisierungsausschuss/pSta20181211Abg.pdf> (18.1.2021).

2 Fortbildungsveranstaltung des VDB-Hessen *Im Fokus: Inhaltserschließung* am 2. September 2019 in Fulda; Sitzung der österreichischen Kommission für Sacherschließung am 9. September 2019 in Graz; Tagung *Netzwerk maschinelle Erschließung* am 10./11. Oktober 2019 in der DNB in Frankfurt a. M. im Rahmen des Knowledge Cafés *Inhaltserschließung im digitalen Zeitalter für Wissenschaft von heute und morgen*.

Zürich), Urs Frei (Schweizerische Nationalbibliothek Bern), Michael Franke-Maier\* (Universitätsbibliothek der Freien Universität Berlin), Werner Holbach (Bayerische Staatsbibliothek), Armin Kühn (Bibliotheksservice-Zentrum), Bettina Kunz (Niedersächsische Staats- und Universitätsbibliothek Göttingen), Julijana Nadj-Guttandin\* (Deutsche Nationalbibliothek Frankfurt am Main), Sarah Pielmeier (Universitäts- und Landesbibliothek Münster), Esther Scheven (Deutsche Nationalbibliothek Frankfurt am Main), Christoph Steiner\* (Österreichische Nationalbibliothek), Heidrun Wiesenmüller (Hochschule der Medien Stuttgart), Barbara Wolf-Dahm\* (Universitätsbibliothek Augsburg).

## **Allgemeine verfahrensunabhängige Kriterien und Maßnahmen zur Qualitätssicherung**

### **1 Ausgangslage**

In den vergangenen Jahren haben sich die Rahmenbedingungen für die bibliothekarische Kernaufgabe der Inhaltserschließung deutlich verändert und erweitert. Die zunehmende Menge von Materialien und Publikationsformen und die Heterogenität der Daten bergen zahlreiche immanente Möglichkeiten, aber auch Notwendigkeiten für die Entwicklung und den Einsatz neuer Erschließungsverfahren. Im selben Maße sind die Anforderungen an die inhaltliche Aufbereitung von Ressourcen und die Anwendungsbereiche gewachsen. Regelwerke und Normdateien öffnen sich für Anwendungsgruppen und Nutzungsszenarien jenseits der Bibliotheken. Die traditionelle intellektuelle Sacherschließung nach den Regeln für die Schlagwortkatalogisierung (RSWK) oder nach einem Klassifikationssystem ist nur noch ein Teil in einem Nebeneinander verschiedener Verfahren und Erschließungssysteme. In erheblichem Umfang sind Bestände nicht inhaltlich erschlossen und entziehen sich der thematischen Auffindbarkeit in der Menge der erschlossenen Bestände. Dementsprechend ist die aktuelle Fachdiskussion über zukunftsfähige Verfahren zur Inhaltserschließung geprägt von unterschiedlichen Sichtweisen auf die Bedeutung, den Nutzen und die Relevanz von Regelwerken und kontrollierten Vokabularen. In diesem vielschichtigen Kontext spielt der Qualitätsbegriff eine zentrale Rolle. Damit unterschiedliche Verfahren zukünftig nicht nur nebeneinander existieren, sondern unter optimaler Ausnutzung ihrer Stärken zusammenwirken können, bedarf es der Konsolidierung der Qualitätskriterien, die der inhaltlichen Erschließung zugrunde liegen. Ziel ist die Ablösung bisher hilfswise herange-

zogener Bewertungskategorien wie z. B. „gut“, „schlecht“, „nützlich“, „falsch“, „präzise“, „grob“ durch verlässliche und operationalisierbare Definitionen.

Verfahrensunabhängige Qualität besteht nicht in der Angleichung der Inhalte und Erschließungsregeln unterschiedlicher Wissensorganisationssysteme, sondern in der Verbindlichkeit von Standards für die Gestaltung von Wissensorganisationssystemen. Damit wird Interoperabilität heterogener Systeme und in heterogenen Umgebungen ermöglicht.

Die Qualität inhaltlicher Erschließung umfasst mindestens die folgenden Qualitätsdimensionen: Sie resultiert aus den übergeordneten Grunddimensionen *Transparenz* und *Verlässlichkeit*, aus der *regelbasierten Produktion von Normdaten*, aus *Verwendungsregeln für die Ressourcenbeschreibung*, aus der *transparenten Auswertung für Retrieval und Anzeige* und aus der *Öffnung der Daten für die Nachnutzung in anderen Wissenschaftskontexten jenseits des klassischen Retrievals*.

## 2 Grunddimensionen: Verlässlichkeit und Transparenz

Als Grunddimensionen von Qualität gelten generell die Anforderungen *Verlässlichkeit* und *Transparenz*. Dabei versteht sich Verlässlichkeit im Sinne von Konsistenz der Daten, der Regeln und des Retrievals. Transparenz beinhaltet die Offenlegung von Regeln und Standards, die genaue Herkunftskennzeichnung von Daten (Erschließungsmethode, Erschließungslevel, Konfidenz) sowie das Verdeutlichen von Erschließungslücken und heterogenen Erschließungssituationen. Durch die optimale Nutzung der Stärken der unterschiedlichen Erschließungssysteme und -methoden wird ein hohes Maß an Benutzungsfreundlichkeit erzeugt. Differenzierte Recherchemöglichkeiten verbessern die *discoverability* von Bibliotheksbeständen. Weiterhin dienen die beiden Grunddimensionen in zunehmendem Maße auch der Nachnutzung dieser Daten in anderen Wissenschaftskontexten jenseits des klassischen Retrievals, wie z. B. für die Digital Humanities, oder als Grundlage für lernende Verfahren.

## 3 Qualitätsdimension I: Regelbasierte Produktion von Normdaten

Voraussetzung für eine qualitativ hochwertige inhaltliche Erschließung, sei sie intellektuell oder maschinell, sind Normdateien, Thesauri, Klassifikationen und Ontologien. Gegenüber einer reinen Verstichwortung hat die Nutzung solcher

Wissensorganisationssysteme Vorteile, z.B. die Herstellung von Kontext. Insofern bedarf es einer Regelung für die Erstellung und Verwendung von Normdaten.

Dabei sind Anforderungen an Qualität:

- eindeutige Bestimmung des Begriffsumfangs der einzelnen Konzepte
- Disambiguierung im Verhältnis zu anderen Konzepten
- spezifisch für die Gemeinsame Normdatei (GND): Kritische Überarbeitung der Zerlegungskontrolle, vor allem für solche Normdatensätze, welche mit Verfahren der Named Entity Recognition exakt zugeordnet werden können
- konsistente terminologische Kontrolle und Anreicherung durch Synonyme
- Relationierung und Hierarchisierung der Datensätze
- Gebräuchlichkeit der Terminologie, die beim Retrieval sowohl den Bedarfen von Lai:innen im jeweiligen Fach auf der einen Seite bis hin zu Expert:innen auf der anderen Seite gerecht wird
- Aktualität und Wissenschaftsnähe der Terminologie
- Ausrichtung an internationalen Standards, wie z.B. der Thesaurus-Norm ISO 25964 oder Resource Description and Access (RDA)
- Maschinenlesbarkeit aller Attribute der einzelnen Normdatensätze und geeignete Aufbereitung für das Semantic Web
- Verknüpfung mit anderen Normdaten (Linked Data, semantische Dichte), um logische Schlussfolgerungen durch Mensch und Maschine sowie die Navigation in den Daten zu ermöglichen

## 4 Qualitätsdimension II: Verwendungsregeln für die Ressourcenbeschreibung

In einem kooperativen Erschließungskontext sind unterschiedliche Bibliothekstypen (Spezialbibliotheken, Universalbibliotheken, öffentliche Bibliotheken), Bibliotheksverbünde und ggf. nicht-bibliothekarische Agierende aktiv, die untereinander Daten austauschen. Ein gewinnbringender Datenaustausch ist dann möglich, wenn sich die einzelnen Mitwirkenden des Netzwerks auf gemeinsame Standards für die inhaltliche Ressourcenbeschreibung einigen. Weiterhin werden z.T. konsortiale Softwarelösungen großer Firmen zur Recherche eingesetzt, die sich für die Optimierung der Suchwerkzeuge an klaren Vorgaben ausrichten sollten.

Anforderungen an die Inhaltserschließung können unterschiedlich sein – in Abhängigkeit einerseits von der Fachdomäne, der Institution und den Be-

nutzer:innen, andererseits vom Dokumenttyp. Verwendungsregeln für die Ressourcenbeschreibung sind vor diesem Hintergrund zu flexibilisieren, um den unterschiedlichen Bedürfnissen gerecht zu werden.

Für das Gelingen dieser Flexibilisierung und den optimalen Einsatz der besonderen Stärken der unterschiedlichen Verfahren und Systeme für Retrieval, Anzeige, Facettierung und Nachnutzung der Daten bedarf es im Sinne der Grunddimension *Transparenz* einer standardisierten und maschinenlesbaren Kennzeichnung. Die eingesetzten Verfahren und Wissensorganisationssysteme sowie die Erschließungstiefe sind idealerweise aus den einzelnen Datensätzen auslesbar.

Anforderungen an Verwendungsregeln sind:

- Verständlichkeit
- gut nachvollziehbare Gliederung und übersichtliche Gestaltung des Richtlinientextes
- Ermöglichung einer flexiblen Handhabung für unterschiedliche Szenarien durch die Einführung von abgestuften Erschließungsleveln mit unterschiedlichen Präzisionsgraden unter Berücksichtigung eines Minimalstandards
- Einsatz international anerkannter, sich weiterentwickelnder Wissensorganisationssysteme (GND, Dewey-Dezimalklassifikation, Regensburger Verbundklassifikation, Basisklassifikation etc.)
- präzise inhaltliche Beschreibung einer Ressource in ihrer Gesamtheit
- Ermöglichung der Beschreibung von Teilaspekten (*minor terms*)
- Ermöglichung einer weniger granularen Beschreibung (*broader terms*), die ggf. auch maschinell aus der präzisen inhaltlichen Erschließung erzeugt werden kann
- maschinenlesbare Kennzeichnung der Erschließungstiefe
- Ausrichtung der Verwendungsregeln auch an der Eignung der Ergebnisse als Trainingsdaten für maschinelle Verfahren

## 5 Qualitätsdimension III: Transparente Auswertung für Retrieval und Anzeige

Ziel der Inhaltserschließung ist unter Orientierung an den Bedarfen unterschiedlicher Zielgruppen die optimale Auffindbarkeit von Ressourcen im Zusammenspiel von *recall* und *precision*. Metadaten der Inhaltserschließung sind dabei die Grundlage für die thematische Suche mit iterativer Herangehens-

weise. Differenzierte Recherche- und Navigationsmöglichkeiten der Suchinstrumente verbessern die *discoverability* der Bestände.

Insofern bedarf es einer benutzungsfreundlichen Visualisierung sowie der Offenlegung und der nachvollziehbaren Dokumentation der Erschließungssituation auch im Suchwerkzeug. Informationen zum Einsatz unterschiedlicher Wissensorganisationssysteme, zu Erschließungslücken sowie zu unterschiedlichen Erschließungstiefen müssen auch für Nutzer:innen einsehbar sein.

Qualitätsanforderungen für die Gestaltung von Suchwerkzeugen sind:

- Überschaubarkeit, Vollständigkeit, Präzision von Treffermengen
- Verständlichkeit der Anzeige inhaltlicher Metadaten
- Orientierung an den IFLA LRM User tasks
- sinnvolle Normalisierung und Indexierung
- Auswertung von Cross-Konkordanzen und Relationierungen zur Homogenisierung der Vielzahl unterschiedlicher inhaltlicher Metadaten (z. B. in Discovery-Systemen)
- optimale Auswertung inhaltlicher Metadaten für das Relevanz-Ranking
- kontextsensitive Ergebnismengen auf Basis von User-Profilen bzw. fachlichen Bedarfen
- Einbindung alternativer inhaltlicher Erschließungselemente (Links zu Klappen- und Volltexten, Inhaltsverzeichnissen etc.)
- Angebot von weiterführenden Funktionalitäten auf Basis inhaltlicher Metadaten: Suchausweitung mit Hilfe von Oberbegriffsrelationen oder verwandten Begriffen, Verlinkung in Drittsysteme wie Wikipedia oder andere Referenzwerke
- Anzeige der Relationen als Graph, Themenkarte o. Ä. und Ermöglichung einer Navigation über Beziehungen innerhalb der Normdaten
- einfach zu nutzende thematische Sucheinstiege jenseits des Google-Schlitzes (z. B. Browsen, Karte, Zahlenstrahl, Warenkorb für Suchbegriffe),
- sinnvolle Anordnung und Abfolge der vorhandenen Facetten sowie Entwicklung neuer Facetten auf Basis von zeitlichen und geografischen Angaben in Metadaten
- gute Usability und kontextsensitive Hilfe

## 6 Qualitätsdimension IV: Öffnung der Daten für die Nachnutzung in anderen Wissenschaftskontexten jenseits des klassischen Retrievals

Inhaltsbeschreibende Normdaten bzw. Metadaten sind nicht nur für das klassische Retrieval im bibliografischen Kontext interessant. Die Vernetzung der GND mit anderen Wissensorganisationssystemen, z. B. mit den Library of Congress Subject Headings (LCSH), dem Virtual International Authority File (VIAF) oder mit Wikidata, und die Nachnutzung der GND im Bereich von Forschungsprojekten und Digital Humanities nimmt zu. Gleichzeitig kann das in Metadaten abgelegte vernetzte Wissen selbst zum Forschungsobjekt werden oder die Bearbeitung von Forschungsfragen unterstützen. Die im Kontext von Forschungsdaten entstandenen FAIR-Prinzipien<sup>3</sup> können ein Ausgangspunkt für die Definition von Anforderungen dieser Dimension sein. Für die Öffnung der bibliothekarischen Datenwelt und die Bereitstellung von *Linked Open Library Data* sowie für die Vernetzung in der *Linked Open Data Cloud* sind die genannten Anforderungen an Qualität daher ebenfalls von Relevanz.

## 7 Messbarkeit

Nachweisbare Qualität liegt dann vor, wenn der Erfüllungsgrad von Anforderungen objektiv messbar ist. Die Messbarkeit von Qualität der Inhaltserschließung hängt davon ab, ob geeignete Metriken entwickelt werden können, um die Umsetzung der hier angestellten Überlegungen zu überprüfen.

Ansatzpunkte für die Messung der Qualität können sein:

- Grad der Abdeckung inhaltlich erschlossener Medien in bibliografischen Metadaten-Pools
- Anteil der Datensätze mit eindeutiger Kennzeichnung der Erschließungsmethode etc.
- Zahl der Relationierungen zwischen Normdatensätzen im selben Wissensorganisationssystem, Zahl der Beziehungen zwischen dem eigenen System und anderen Systemen
- Anteil der tatsächlich in der Recherche genutzten Informationen in Normdatensätzen im Vergleich zu den dort enthaltenen Informationen (ohne Berücksichtigung solcher Informationen, die nur für Erschließer:innen von Bedeutung sind)

---

<sup>3</sup> GO FAIR Initiative: FAIR-Principles: <https://www.go-fair.org/fair-principles/> (18.1.2020).

- Nachvollziehbarkeit des Zustandekommens von Ergebnissen bzw. Treffern nach einer Recherche
- Auswertung von Sucherfolgen im Kontext thematischer Suchanfragen (z. B. Auslösen von Aktionen wie Ausleihe, Aufruf von Volltexten, Download bibliografischer Daten) oder Nachnutzung der angezeigten inhaltlichen Metadaten für weiterführende Suchen
- im Bereich der maschinellen Erschließung werden klassische Maße wie *recall*, *precision* und F-Maß berechnet
- Nachnutzbarkeit und Nachnutzung inhaltlicher Norm- oder Metadaten für die Entwicklung und den Einsatz darauf aufbauender IT-Verfahren in anderen Wissenschaftskontexten (Metadaten als Grundlage für Forschung oder für Digital Humanities).

## Literaturverzeichnis

- Arbeitsstelle für Standardisierung: Protokoll der 34. Sitzung des Standardisierungsausschusses vom 10. Dezember 2018. <https://www.dnb.de/SharedDocs/Downloads/DE/Professionell/Standardisierung/Protokolle/Standardisierungsausschuss/pSta20181211Abg.pdf> (18.1.2021).
- International Organization for Standardization: ISO 25964: Information and documentation – Thesauri and interoperability with other vocabularies. Geneva: ISO.
- International Federation of Library Associations: IFLA Library Reference Model (LRM). <https://www.ifla.org/publications/node/11412> (18.1.2020).
- GO FAIR Initiative: FAIR-Principles: <https://www.go-fair.org/fair-principles/> (18.1.2020).