

Jakob Voß

Datenqualität als Grundlage qualitativer Inhaltserschließung

1 Einleitung

Spätestens mit Beginn des 21. Jahrhunderts findet die inhaltliche Erschließung von Dokumenten praktisch ausschließlich in digitaler Form statt. Dies gilt sowohl für die fachliche Inhaltserschließung durch Bibliotheken und andere Dokumentationseinrichtungen als auch für die verschiedensten Formen inhaltlicher Beschreibung in Datenbanken – von Produktbeschreibungen im Internethandel bis zum Social Tagging. Selbst dort, wo analoge Ursprünge vorhanden sind, beispielsweise handschriftliche Notizen oder retrokonvertierte Findmittel, liegt die Sacherschließung am Ende in Form von Daten vor.

Für die konkrete Ausprägung dieser Daten gibt es allerdings viele verschiedene Möglichkeiten. Der vorliegende Beitrag soll einen Überblick darüber geben, wie unterschiedliche Praktiken der Datenverarbeitung die Qualität von Inhaltserschließung beeinflussen und wie die Qualität von Erschließungsdaten beurteilt werden kann. Der Fokus liegt also nicht auf den Inhalten von Erschließungsdaten, sondern auf ihrer Form. Die Form von Daten ist keine rein technische Nebensächlichkeit, sondern durchaus relevant: So ist eine inhaltlich hervorragende Erschließung unbrauchbar, wenn die Erschließungsdaten aufgrund inkompatibler Datenformate nicht verwendet werden können. Zur qualitativen Einschätzung von Inhaltserschließung ist es daher notwendig, sich auch darüber im Klaren zu sein, wie und in welcher Form die Erschließungsdaten verarbeitet werden.

Nach der Klärung grundlegender Begriffe rund um Datenqualität werden Faktoren dargelegt, die die Qualität auf formaler Ebene beeinflussen. Eine zentrale Rolle spielen dabei Datenformate und Anwendungen. Aus den Einflussfaktoren gehen Anforderungen hervor, um den qualitativen Umgang mit Erschließungsdaten beurteilen und sicherstellen zu können.

2 Datenqualität und ihre Dimensionen

Im Rahmen des Qualitätsmanagements bezieht sich Qualität auf die „Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen“ (ISO 8402) bezie-

ungsweise auf den „Grad, in dem ein Satz inhärenter Merkmale Anforderungen erfüllt“ (ISO 9000). Qualität ist also kein Selbstzweck, sondern muss immer danach beurteilt werden, welche Anforderungen erfüllt werden sollen. Eine umfassende Beurteilung von Datenqualität in der Inhaltserschließung erfordert also die genauere Bestimmung dessen, was mit Inhaltserschließung erreicht werden soll – kurz gesagt, die Frage nach dem Sinn von Inhaltserschließung. Um dieser durchaus berechtigten Frage aus dem Weg zu gehen, setzen wir voraus, dass Inhaltserschließung gewisse sinnvolle Zwecke erfüllen kann, die damit zusammenhängen, Dokumente gezielt aufzufinden (Retrieval) und sich einen Überblick über ihre Inhalte zu verschaffen (Bertram 2005; Gödert, Lepsky und Nagelschmidt 2012).

2.1 Datenqualität als Grundlage von Informationsqualität

Wie gut die Zwecke der Inhaltserschließung erreicht werden können, ist weniger eine Frage der Daten- als der Informationsqualität. Beide Begriffe werden trotz gelegentlicher Differenzierungsversuche meist synonym verwendet (Hildebrand u. a. 2018, Kap. 5). Deshalb ist an dieser Stelle eine genauere Eingrenzung des Datenbegriffs notwendig: Daten lassen sich aus verschiedener Perspektive als Fakten, Beobachtungen oder Dokumente auffassen (Voss 2013a). Im Rahmen dieser Untersuchung wird letztere Auffassung vertreten, nach der Daten in erster Linie digitale Dokumente sind. Daten lassen sich daher anders als Informationen unabhängig vom konkreten Inhalt in Form von Strukturen untersuchen (Voss 2013b). Die hier untersuchte Form von Datenqualität kann daher auch als *formale Datenqualität* oder *strukturelle Datenqualität* bezeichnet werden. Die Relevanz dieser Betrachtungsweise für die Datenqualität lässt sich anhand eines Datensatzes illustrieren, der eine Liste von Schlagwörtern enthält:

- Niedrige Informationsqualität ist gegeben, wenn die Schlagwörter nicht zum Dokument passen oder bessere Schlagwörter möglich wären.
- Niedrige Datenqualität ist gegeben, wenn die Schlagwörter nicht den formalen Regeln des Datenformats entsprechen, beispielsweise weil keine Liste erkennbar ist oder weil die Schlagwortliste Elemente enthält, die keine Schlagwörter sind.

Datenqualität ist demnach kein hinreichendes Kriterium für qualitative Inhaltserschließung, sondern deren notwendige Grundlage. Viele Merkmale von Datenqualität sind nicht nur für Inhaltserschließungsdaten, sondern auch für andere Anwendungen relevant; ihre Beschreibung ist deshalb auch ein allge-

meiner Beitrag zur Data Literacy. Zusammengefasst beschäftigt sich Datenqualität nicht mit dem Inhalt, sondern mit der Form. Die angestrebten Anforderungen an Daten betreffen vor allem die Lesbarkeit und Verarbeitbarkeit. Dazu ist es weniger relevant, was in den Daten ausgedrückt ist, sondern wie die Daten erstellt und verarbeitet werden.

2.2 Dimensionen der Datenqualität

Zur gezielten Erfassung und Verbesserung von Aspekten der Informations- bzw. Datenqualität werden in der Literatur verschiedene Dimensionen wie Vollständigkeit, Fehlerfreiheit, Konsistenz und Aktualität genannt. Eine einheitliche Einteilung nach Dimensionen existiert jedoch nicht (Floridi und Phyllis 2014; Batini und Scannapieco 2016).

Unter den von Franke-Maier (2018) genannten Qualitätsdimensionen im Bereich der Inhaltserschließung sind für die Datenqualität konkret die „Qualität der digitalen Werkzeuge“, die „Qualität des Datenaustauschs“ und die „Qualität der Suchwerkzeuge“ relevant. Unter den von Hildebrand u. a. (2018) formulierten allgemeinen Dimensionen lässt sich für Daten im engeren Sinne lediglich Konsistenz bestimmen: Diese basiert auf einer Menge von Konsistenzregeln, die jeweils erfüllt oder verletzt werden können. So könnte beispielsweise für bibliografische Daten festgelegt werden, dass eine Publikation keine ISBN haben darf, wenn ihr Erscheinungsjahr vor 1965 liegt.¹ Für die Sacherschließung sind solche Konsistenzregeln allerdings nur schwer anzugeben. Stattdessen haben einzelne Erschließungssysteme eigene Regeln, wie zum Beispiel die Bildungsregeln für syntaktisch korrekte Notationen der Dewey-Dezimalklassifikation (Reiner 2007).

Bei vorgeschriebener paralleler Verwendung mehrerer Erschließungssysteme (beispielsweise, wenn jeder Datensatz sowohl mit der Basisklassifikation als auch mit der Regensburger Verbundklassifikation erschlossen sein muss) und vollständig vorhandenen Normdaten-Mappings zwischen diesen Systemen² könnte eine Konsistenzregel darin bestehen, dass die gewählten Notationen bei der Systeme zumindest indirekt durch ein Mapping verbunden sein müssen.

¹ Das Standard Book Numbering System, britischer Vorläufer der ISBN, wurde 1965 von Gordon Foster entwickelt.

² Siehe dazu die Sammlung und Werkzeuge zu Normdaten-Mappings unter <https://coli-conc.gbv.de/> (17.6.2021).

3 Einflussfaktoren auf die Datenqualität

Die Faktoren, die die Datenqualität in der Inhaltserschließung beeinflussen, stammen aus drei unabhängigen Bereichen, welche in wechselseitiger Beziehung stehen. Die Bereiche beziehen sich auf unterschiedliche Aspekte von Daten in ihrer statischen Form, in ihrer dynamischen Verarbeitung und in ihrem praktischen Einsatz.

3.1 Standards und Regeln

Standards und Regeln legen fest, wie Daten idealerweise aussehen sollen. Diese präskriptiven Faktoren finden sich insbesondere in Form von Datenformaten. Für die Inhaltserschließung sind drei Gruppen von Datenformaten relevant:

- *Normdatenformate* wie SKOS, JSKOS und das MARC-21-Format für Normdaten dienen dem Austausch von Erschließungssystemen.³ Ihr Einsatz beeinflusst, wie gut konkrete Klassifikationen, Thesauri und andere Wissensorganisationssysteme genutzt werden können.
- *Bibliografische Metadatenformate* wie Dublin Core, DataCite und das PICA-Format des K10plus enthalten oft unter anderem Felder oder Attribute zur Inhaltserschließung. Die Formate beeinflussen, welche Erschließungsarten in welchem Umfang in Daten vorgesehen sind.
- Schließlich können Datenformate beider Arten mit Hilfe von *Schemaformaten* wie JSON Schema, reguläre Ausdrücke und Shape Expressions (ShEx) definiert werden. Schemata beeinflussen, wie gut sich die Einhaltung von Datenformaten kontrollieren lässt.

Die Übersicht von Wissensorganisationssystemen im Verzeichnis BARTOC⁴ zeigt, dass viele Erschließungssysteme nicht in einem offenen, standardisierten Normdatenformat vorliegen, sodass die Qualität ihres Einsatzes nur unzureichend überprüft werden kann. Bibliografische Daten sind eher verfügbar, allerdings bleibt ihre konkrete Ausgestaltung für die Inhaltserschließung oft in informellen Anwendungsregeln verborgen. Grund dafür ist auch der eher seltene Einsatz von formalen Schemas.⁵

³ Siehe <https://format.gbv.de/application/authority> (17.6.2021) für eine fortlaufend erweiterte Übersicht von Normdatenformaten.

⁴ Siehe <https://bartoc.org/> (17.6.2021).

⁵ So ist beispielsweise das Katalogformat des K10plus erst seit 2020 auch als maschinenlesbares Schema verfügbar: <https://format.k10plus.de/avram.pl> (17.6.2021).

3.2 Anwendungsprogramme

Anwendungsprogramme bestimmen, wie Daten verarbeitet werden. Diese Faktoren sind durch Algorithmen und Schnittstellen implementiert und legen fest, welche Form Daten tatsächlich annehmen können. Für die Datenqualität relevant sind praktisch alle Programme, die Erschließungsdaten verwenden, insbesondere:

- Programme oder Funktionen zur Beschreibung von Dokumenten (Dokumentenserver, Tagging-Systeme, Katalogisierungssoftware etc.)
- Programme zur automatischen Erzeugung von Erschließungsdaten
- Programme zum Retrieval oder zur Visualisierung aufbauend auf Erschließungsdaten

Der Einfluss von Anwendungsprogrammen auf die Datenqualität, insbesondere durch Usability und das Vorhandensein bzw. Fehlen von Funktionalitäten, kann nicht unterschätzt werden. So verzichten beispielsweise viele Systeme auf eine Autocomplete-Funktion, obwohl dadurch die Erschließungsqualität verbessert werden kann (Holstrom 2020). Auch wirkt die Art der Nutzung von Inhaltserschließung auf die Daten zurück: Werden beispielsweise die Daten beim Retrieval für Facettierung und Drill-Down eingesetzt, fallen Lücken und Fehler eher auf und es gibt einen Anreiz, die Qualität der Daten zu verbessern.

3.3 Praktische Faktoren

Die praktischen Faktoren ergeben sich daraus, wie Daten in der Praxis tatsächlich eingesetzt werden. Im Gegensatz zu normativen Regeln und Programmen, die lediglich eine Möglichkeit vorgeben, besteht die Datenpraxis darin, wie Menschen auf unterschiedliche Art mit Regeln und Programmen umgehen. Neben individuellen Eigenheiten gehören zu diesen Faktoren institutionelle Anforderungen und Zwänge, sowie nicht zuletzt die sogenannte Macht der Gewohnheit.

Die Einflussfaktoren auf die Datenqualität führen in den verschiedenen Bereichen oft eine Art Eigenleben: Regeln müssen erst mit Leben gefüllt werden, Implementierungen setzen eigene Beschränkungen und Möglichkeiten und aus der Praxis entstehen neue Anforderungen. Nur durch fortwährende Überprüfung kann sichergestellt werden, dass Standards, Programme und Praxis so gut wie möglich übereinstimmen und Datenqualität messbar wird. Welche Anforderungen dazu erfüllt werden müssen, ist Gegenstand des folgenden Abschnitts.

4 Anforderungen an den qualitativen Umgang mit Erschließungsdaten

Durch die Betrachtung von Faktoren, die die Qualität von Erschließungsdaten beeinflussen, lassen sich einige allgemeine Anforderungen an den Umgang mit Erschließungsdaten identifizieren.

4.1 Datenqualität erfordert Regeln

Erstens muss es Regeln geben, an denen die Datenqualität gemessen werden kann. So ist beispielsweise die formale Qualität eines Datenfeldes *Schlagwörter* nicht bestimmbar, wenn genauere Angaben dazu fehlen, in welcher Form welche Art von Schlagwörtern in diesem Feld eingetragen werden sollen. Selbst Freitextfelder beinhalten bei genauerer Betrachtung meist Regeln zu erlaubten Zeichen und Längen. Zwar können Menschen Daten oft auch ad-hoc auf Grundlage von Erfahrungswerten einschätzen; zur Sicherstellung einer einheitlichen Datenqualität ist die subjektive Bewertung von Einzelfällen jedoch nicht geeignet. Stattdessen können Expert:innen durch ihre Einschätzungen dazu beitragen, sich möglicher Anforderungen an die Datenqualität bewusst zu werden und so zu konkreten Regeln zu kommen.

4.2 Datenqualität erfordert Dokumentation

Zweitens müssen die Regeln, denen qualitative Daten folgen sollen, auch dokumentiert sein. Insbesondere müssen implizite Annahmen und Praktiken (*tritt sowieso nie auf, wurde schon immer so gemacht* etc.) explizit gemacht werden. Implizite Regeln finden sich oft im praktischen Umgang mit Daten und versteckt in Anwendungsprogrammen zur Erfassung und Verarbeitung von Erschließungsdaten. Werden beispielsweise in einem Eingabefeld mehrere Schlagwörter durch Kommata getrennt, ergibt sich daraus vermutlich die Regel, dass einzelne Schlagwörter kein Komma beinhalten können. Die Notwendigkeit einer Dokumentation ist umso höher, wenn Daten zwischen verschiedenen Einrichtungen oder Systemen ausgetauscht werden. Dabei sollte die Dokumentation nicht nur für diejenigen zugänglich sein, die Erschließung betreiben, sondern auch für all jene, die Daten nutzen oder an einer Nutzung interessiert sein könnten.

Die Anforderungen an eine gute Dokumentation für Datenregeln entsprechen in etwa den allgemeinen Anforderungen an gute technische Dokumentation (Juhl 2015). Die Regeln zur Datenqualität sollten zudem leicht auffindbar publiziert sein und Informationen zu Herausgeber:innen, Gültigkeit und Änderungen beinhalten, idealerweise mit Versionierung.

4.3 Datenqualität erfordert Validierung

Datenstandards sind Abmachungen, sie können eingehalten oder gebrochen werden. Zur Sicherstellung von Qualität müssen Regeln, nach denen Daten aufgebaut werden, auch überprüft werden. Die Überprüfung muss als automatisierte *Validierung* möglich sein. Zum einen werden Regeln damit objektivierbar und zum anderen kann ihre Einhaltung dadurch nicht nur sporadisch, sondern systematisch in großem Maßstab, wiederholt und dauerhaft überprüft werden.

Prinzipiell kann Validierung durch automatische Tests in einer beliebigen Programmiersprache umgesetzt werden. Soweit möglich sollten Regeln jedoch mit Hilfe formaler Schemasprachen ausgedrückt werden.⁶ Schemasprachen dienen der Exaktheit von Regeln, helfen dabei, dass die Validierung tatsächlich das prüft, was geprüft werden soll, und ermöglichen eine Validierung unabhängig von einzelnen Anwendungsprogrammen. Ohne Schemata kommt es nämlich dazu, dass alle Daten, die vom jeweils benutzen Programm nicht explizit als Fehler erkannt werden, fälschlicherweise für korrekt erachtet werden, statt die Daten an den dokumentierten Regeln zu messen.

Die Wahl der Schemasprache hängt insbesondere von der Datenstrukturierungssprache des zu überprüfenden Datenformats ab. So gibt es zum Beispiel JSON Schema für JSON-Formate, XML Schema für XML-Formate und ShEx für RDF-Formate. Für die feldbasierten Bibliotheksdatenformate MARC und PICA gibt es seit 2018 die Schemasprache Avram.⁷ Die meisten Schemasprachen verwenden zur Einschränkung einzelner Datenfelder Datentypen und reguläre Ausdrücke – zumindest auf diese beiden Hilfsmittel sollte bei der Umsetzung der Validierung von Erschließungsdaten zurückgegriffen werden.

⁶ Siehe <https://format.gbv.de/schema/language> für eine Übersicht von Schemasprachen (17.6.2021).

⁷ Siehe <https://format.gbv.de/schema/avram/specification> (17.6.2021).

4.4 Exkurs: Grundlegende Strukturierungsprinzipien

Die formale Definition eines Datenformates und die Angabe von Validierungsregeln mittels einer Schemasprache sind zwei Seiten einer Medaille: Beide beschäftigen sich mit den charakteristischen Strukturen von Daten. Grundsätzlich können Daten auf vielfältige Art und Weise in einzelne Elemente strukturiert und miteinander in Beziehung gesetzt werden. In der Praxis gibt es allerdings einige typische Gestaltungsmuster, die unabhängig von der konkreten Technik immer wieder verwendet werden (Voss 2013b). Zum Verständnis von Datenformaten und Standards hilft es, insbesondere auf folgenden Festlegungen für potenzielle Regeln und Schemata zu achten:

1. Einteilung in Datenelemente: Welche Bestandteile von Daten sind die elementaren Einheiten und wo sind stattdessen Binnenstrukturen vorhanden? Beispielsweise können Namen je nach Erfassungsregeln entweder atomar sein oder weiter in Vor- und Nachnamen unterteilt werden.
2. Relevanz und Normalisierung: Welche Bestandteile sind für den Inhalt der Daten relevant und welche Abweichungen können ohne Bedeutungsverlust unterschiedlich gehandhabt werden (z.B. die Verwendung von Leerzeichen)? Gibt es eine normalisierte Form ohne jede Form irrelevanter Abweichungen?
3. Vorgeschriebene und erlaubte Bestandteile: Welche Datenelemente *müssen* und welche *dürfen* vorhanden sein? So könnte ein Erfassungssystem beispielsweise vorschreiben, dass jedes Objekt mit einer geografischen Koordinate versehen sein muss oder dass die Bedeutung von Schlagwörtern mit Qualifikatoren genauer eingegrenzt werden darf. Zur Vermeidung von Missverständnissen werden Anforderungen in formalen Standards idealerweise nach RFC 2119 (Bradner und Bradner 1997) bzw. nach DIN 820-2 (DIN 820) mit groß geschriebenen Modalverben angegeben:
 - MUST (MUSS) gibt an, dass ein Element zwingend notwendig ist.
 - MUST NOT (DARF NICHT) gibt an, dass ein Element verboten ist.
 - MAY oder OPTIONAL (DARF oder KANN) geben an, dass ein Element erlaubt ist, aber auch weggelassen werden kann.
 - SHOULD (SOLLTE) und SHOULD NOT (SOLLTE NICHT) geben an, dass ein Element empfohlen wird, aber in begründeten Ausnahmefällen auch weggelassen werden kann.
4. Wiederholbarkeit: Welche Bestandteile sind wiederholbar und welche können maximal einmal in einem Datensatz vorkommen?
5. Reihenfolge: Wo spielt die Reihenfolge von Elementen eine Rolle und welche Bedeutung hat sie (beispielsweise die Gewichtung innerhalb von Schlagwortfolgen)?

6. Identifikatoren: Welche Datenelemente dienen als Identifikatoren? Neben offensichtlichen Verweisen auf externe Datenbanken sind auch einfache Listen von erlaubten Werten kleine kontrollierte Vokabulare und ihre Werte damit Identifikatoren.

4.5 Datenqualität erfordert Praxisbezug

Schließlich muss die formale Überprüfung von Datenqualität realistische Auswirkungen auf die Praxis haben. Wenn Regelverstöße keine Konsequenzen nach sich ziehen, ist davon auszugehen, dass das Ergebnis der Validierung im Zweifelsfall ignoriert wird. Welche Konsequenzen angemessen sind (einfache Mitteilungen, Aufforderung zur Korrektur, automatische Ablehnung invalider Daten etc.) hängt von Aufwand und Nutzen der Qualitätsverbesserung ab. Wenn die Überprüfung eines Qualitätsmaßes nicht geleistet werden kann oder ihr Verletzen keine Auswirkungen hat, kann es auch in der Inhaltserschließung sinnvoll sein die Maßstäbe etwas niedriger anzusetzen und statt unrealistischer Vorstellungen nur zu verlangen, was auch tatsächlich umsetzbar ist.

Grundsätzlich sollte Datenqualität nicht als rein technische Angelegenheit gesehen werden: Keine Regel ist frei von Interessen, sozialen Implikationen und möglichen Konflikten. Hier hilft es, sich nicht nur über den praktischen Einsatz der Daten (Anwendungen, Hilfsmittel, Umfang etc.), sondern auch über Aspekte wie Agierende, Interessen und Abhängigkeiten klar zu werden.

5 Zusammenfassung

Gute Sacherschließung lässt sich letztendlich nur inhaltlich und abhängig von ihrem Einsatz beurteilen. Gleichwohl können auf formaler Ebene Anforderungen identifiziert werden, deren Erfüllung für gute Datenqualität notwendig ist. Die Anforderungen ergeben sich aus den Dimensionen und Einflussfaktoren der Datenqualität, insbesondere aus Standards, Anwendungsprogrammen und der Datenpraxis. Vor allem ist Datenqualität nur mit expliziten Regeln möglich. Diese Regeln müssen angemessen dokumentiert sein und durch Validierung überprüft werden. Idealerweise geschieht die Validierung mittels formaler Schemata, in denen grundlegende Strukturierungsprinzipien der gewünschten Daten explizit gemacht werden. Dass mit diesen Mitteln auch große Mengen von Daten einer Qualitätskontrolle unterzogen werden können, zeigt die Analyse von Király (2019) auf Grundlage eines Avram Schemas für MARC 21.

6 Literaturverzeichnis

- Batini, Carlo und Monica Scannapieco: *Data and Information Quality. Dimensions, Principles and Techniques*. Cham: Springer International Publishing 2016.
- Bertram, Jutta: *Einführung in die inhaltliche Erschließung. Grundlagen – Methoden – Instrumente*. Würzburg: Ergon 2005.
- Bradner, S.: *Key words for use in RFCs to Indicate Requirement Levels*. IETF 1997 (RFC 2119). <https://tools.ietf.org/html/rfc2119> (17.6.2021).
- Floridi, Luciano und Phyllis Illari (Hrsg.): *The Philosophy of Information Quality*. Cham: Springer International Publishing 2014.
- Franke-Maier, Michael: *Anforderungen an die Qualität der Inhaltserschließung im Spannungsfeld von intellektuell und automatisch erzeugten Metadaten*. In: *ABI Technik* (2018) Bd. 38 Nr. 4. S. 327–331. <https://doi.org/10.1515/abitech-2018-4005>.
- Gödert, Winfried, Klaus Lepsky und Matthias Nagelschmidt: *Informationserschließung und Automatisches Indexieren: Ein Lehr- und Arbeitsbuch*. Berlin, Heidelberg: Springer 2012.
- Hildebrand, Knut, Marcus Gebauer, Holger Hinrichs, und Michael Mielke (Hrsg.): *Daten- und Informationsqualität*. 4. Aufl. Wiesbaden: Springer Vieweg 2018.
- Holstrom, Chris: *The effects of suggested tags and autocomplete features on social tagging behaviors*. In: *Proceedings of the Association for Information Science and Technology* (2020) H. 57, Nr. 1. <https://doi.org/10.1002/pra2.263>.
- Juhl, Dietrich: *Technische Dokumentation: praktische Anleitungen und Beispiele*. 3. Aufl. Berlin, Heidelberg: Springer Vieweg 2015.
- Király, Péter: *Validating 126 million MARC records*. In: *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage, ACM* 2019. S. 161–168. <https://doi.org/10.1145/3322905.3322929>.
- ISO 8402: *Quality management and quality assurance – Vocabulary*. 2. Aufl. International Organization for Standardization 1994.
- DIN 820: *Normungsarbeit*. International Organization for Standardization 2012.
- ISO 9000: *Quality management systems – Fundamentals and vocabulary*. International Organization for Standardization 2015.
- Reiner, Ulrike: *Automatic Analysis of Dewey Decimal Classification Notations*. In: *Data Analysis, Machine Learning and Applications*. Berlin, Heidelberg: Springer 2007. S. 697–704. https://doi.org/10.1007/978-3-540-78246-9_82.
- Voß, Jakob (2013a): *Was sind eigentlich Daten?* In: *LIBREAS* (2013) H. 23. S. 4–11. <https://doi.org/10.18452/9038>.
- Voß, Jakob (2013b): *Describing Data Patterns*. Dissertation Humboldt-Universität zu Berlin 2013. <http://aboutdata.org/> (17.6.2021).