

Mehr finden – weniger suchen

Wie Normdaten den Unterschied machen

Ein Normdaten-Forschungsvorhaben am Institut für Museumsforschung

Projekt: Nutzung von Normdaten für die Optimierung der Suche für Museumsobjekte in der Deutschen Digitalen Bibliothek

Dr. Domenic Städtler

In dem am Institut für Museumsforschung angesiedelten Forschungsvorhaben „Nutzung von Normdaten zur Optimierung der Suche nach Museumsobjekten in der Deutschen Digitalen Bibliothek“ wurde der folgenden Fragestellung nachgegangen: Wie kann die Suche in der Deutschen Digitalen Bibliothek durch eine gezieltere Einbeziehung von Normdaten verbessert werden? Im Rahmen des Forschungsvorhabens wurde eruiert, wie durch eine konsequentere Nutzung und Berücksichtigung von Normdaten die Auffindbarkeit von Museumsobjekten in der Deutschen Digitalen Bibliothek optimiert werden und die Verwendung von Suchfiltern zu einer höheren Suchgenauigkeit im Portal führen kann. Eine Besonderheit des Forschungsprojektes bestand darin, dass es jenseits der grauen Theorie direkt an der Datenquelle selbst ansetzt: der Erfassung im Museum. Ein bedeutsamer Mehrwert des Forschungsvorhabens zeigte sich darin, dass Projektergebnisse durch die enge Zusammenarbeit mit der Fachstelle Museum und anderen relevanten Akteur*innen auf die operative Arbeit der Deutschen Digitalen Bibliothek einzufließen und Projektergebnisse unmittelbar in die operative Portalweiterentwicklung einfließen konnten.

Am 20. September 2023 fand am Institut für Museumsforschung ein Workshop statt, in dessen Rahmen unter anderem diskutiert wurde, was Museumsdaten ausmacht.¹ Dabei wurden zentrale Frage zur Beschaffenheit von Museumsdaten und konkrete Folgen für den Datenaustausch mit ausgewählten Expert*innen der Museumsdokumentation und Vertreter*innen der Museumscommunity diskutiert. Museen zeichnen sich nämlich durch sehr heterogene infrastrukturelle Voraussetzungen, Arbeitsweisen und Intentionen aus,² was nicht zuletzt der uneinheitlichen fachspezifischen Qualifizierung von Museumsmitarbeiter*innen geschuldet ist;³ dies wirkt sich deutlich auf die Beschaffenheit der Objektdaten aus. Die Nutzung von Normdaten hängt auch maßgeblich von der Trägerschaft von Museen, ihrer Vernetzung untereinander, ihren personellen und finanziellen Ressourcen sowie von externer Unterstützung u. a. durch die Museumsverbände der Länder ab. Doch auch das Fehlen von Förderprogrammen und Finanzierungsmöglichkeiten für die Pflege und Aktualisierung bereits digitalisierter Bestände,⁴ für rückwirkende Kuratierung, für nachträgliche Datenanreicherungen spiegelt sich maßgeblich in

¹ Protokoll: <https://doi.org/10.5281/zenodo.14163368>.

² Siehe auch Rohde-Enslin 2024, passim.

³ Cf. Rohde-Enslin 2024, S. 28. Bestimmte Studiengänge der Museologie und/oder des Museumsmanagements mögen einschränkend als wichtige Ausnahme genannt werden.

⁴ In den Förderrichtlinien von digiS Berlin (Förderprogramm Digitalisierung von Objekten des kulturellen Erbes des Landes Berlin, https://www.digis-berlin.de/wp-content/uploads/2024/05/Foerderrichtlinie_Digitalisierung_kulturellesErbe_Berlin_2025.pdf, S. 4) wird der Begriff Digitalisierung weiter gefasst und inkludiert auch „Datenmanagement und die Datenredaktion (ggf. auch bereits vorhandener digitaler Daten)“. Das ist aber sicherlich die Ausnahme und für viele andere, nur auf die Digitalisierung neuer Bestände fokussierte Förderprogramme nicht repräsentativ.

der Qualität von Museumsdaten wieder. Die einschlägigsten Ergebnisse dieses Stakeholder-Workshops bilden die Grundlage für diese Studie.

Die Studie stützt sich zusätzlich auf einen Artikel, der im Januar 2024 auf Zenodo veröffentlicht wurde. Dieser gehört zu einer Artikelreihe von Mitarbeiterinnen und Mitarbeitern des Instituts für Museumsforschung, die sich mit den Herausforderungen, aber auch Chancen der Vernetzung von Kulturgut über verschiedene Kultursparten hinweg befassen.⁵ Im Artikel *Datenfelder im Museum sind anders! Besonderheiten von Datenfeldern der musealen Erschließung im Vergleich zu jenen in Bibliotheken und Archiven*⁶ wird erörtert, warum für die Erfassung von Museumsobjekten mitunter andere Datenfelder genutzt werden als im Bibliotheks- oder Archivbereich und warum gleichzeitig in der Museumssparte einige Datenfelder mit anderen Objektinformationen befüllt werden müssen als in anderen Kultursparten. Wer sich seriös mit der Bedeutung von Normdaten für die Suche von Museumsobjekten befasst, muss sich zwangsläufig zunächst die Frage stellen, was Museumsobjekte eigentlich ausmacht und wodurch sie sich von Kulturerbeobjekten anderer Sparten unterscheiden. Diese grundlegenden Vorüberlegungen wurden und werden in ebendieser Artikelserie aus verschiedener Perspektive angestellt und bilden das Fundament für die vorliegende Untersuchung.

Den ersten Teil der Studie bildet eine Symptomanalyse: Es wird untersucht, inwieweit bestimmte Arten der Suche (facettierte Suche, Freitextsuche etc.) für bestimmte Entitäten (Objekttyp, Person, Ort etc.) beeinträchtigt sein können. Das Hauptaugenmerk liegt dabei auf der facettierten Suche, da das Einschränken von Treffermengen über Suchfilter eine überaus wichtige Art des Auffindens von bestimmten Erschließungsinformationen darstellt. So ist etwa der Recall beeinträchtigt, wenn zusätzliche synonyme Filterwerte im Suchfilter ausgewählt und/oder allgemeinere Begriffe in den Suchschlitz eingegeben werden müssen, um alle gewünschten Treffer zu erhalten. Gleichzeitig ist die Precision gemindert, wenn Homonyme nicht als solche gekennzeichnet sind und sich hinter ein und demselben Filterwert eigentlich ganz verschiedene Orte, Personen oder Begriffe verbergen.

Darauf aufbauend zielt eine detaillierte Ursachenanalyse darauf ab, limitierende Faktoren zu identifizieren, die die Suche nach Museumsobjekten in der Deutschen Digitalen Bibliothek beeinträchtigen. Dazu gehören, direkt an der Datenquelle ansetzend, die unterschiedlichen Datenbankstrukturen von Erschließungssystemen (optionale oder obligatorische Datenfelder,

⁵ Zu diesem gemeinschaftlichen Publikationsvorhaben siehe: <https://www.smb.museum/museen-einrichtungen/institut-fuer-museumsforschung/forschung/kooperationen/museen-sind-anders/>.

⁶ Städtler 2024.

fehlende Pflichtfelder, Freitextfelder etc.), die genutzten Vokabulare und die heterogenen Erfassungspraktiken in Museen. Dazu gehört aber auch das Ausbleiben einer nachträglichen Datenaufbereitung und einer Anreicherung von Metadaten um persistente Identifikatoren.

Um diesen limitierenden Faktoren zu begegnen, werden dann mögliche Lösungsansätze identifiziert, um trotz fehlender Normierung und trotz unzureichender Lieferdaten dennoch eine möglichst erfolgversprechende Suche im Portal zu ermöglichen: Es lohnt ein vertiefter Blick auf die zentralen Stellschrauben, an denen gezielt gedreht werden kann, um den zuvor beschriebenen Beeinträchtigungen der Suche entgegenzuwirken, angefangen bei Maßnahmen der Normdatenvermittlung, über das gezielte Mappen von Normdaten in enger Abstimmung mit Aggregatoren bis hin zur nachträglichen Normdatenanreicherung, gegebenenfalls auch unter gezielter Einbeziehung von künstlicher Intelligenz.

Der Fokus in dieser Studie liegt auf der Suche in der Deutschen Digitalen Bibliothek. Dieses Kulturportal ist ein wichtiger Showcase, da es zweifellos eine zentrale Plattform für den Zugang zu digitalisiertem kulturellem Erbe in Deutschland darstellt und über 50 Millionen Kulturerbeobjekte aus Museen und anderen Kultursparten miteinander vernetzt. Die Ergebnisse der Untersuchung haben allerdings nicht nur Relevanz für die Deutsche Digitale Bibliothek, sondern sind grundsätzlich auch auf andere Sammlungs- und Forschungsportale übertragbar. Angesichts der Vielzahl an Museumsobjekten in der Deutschen Digitalen Bibliothek lässt sich in dieser Studie immer nur exemplarisch arbeiten. Es wäre verwegend, sämtliche Ursachen für alle denkbaren Probleme, die das Zusammenführen und auffindbar machen im Portal beeinträchtigen, im Rahmen einer einzelnen Studie abschließend erörtern und für jede Problemstelle gleich auch den richtigen Lösungsansatz präsentieren zu wollen. Selbstredend hängen auch gar nicht alle limitierenden Faktoren, die die Suche in der Deutschen Digitalen Bibliothek negativ beeinträchtigen, zwingend mit Normdaten zusammen; weitere ließen sich beliebig ergänzen, würden im Einzelfall aber den selbstgesteckten Rahmen dieser Untersuchung sprengen. Wenn mit dieser Studie aber ein Beitrag dazu geleistet werden kann, zumindest auf einige der gravierendsten Datenqualitätsprobleme im Hinblick auf die Suche im Portal aufmerksam zu machen, und zumindest der eine oder andere Ansatz Abhilfe schafft, hat sich die Mühe gelohnt.

Inhalt

Abkürzungsverzeichnis	10
1. Einleitung.....	11
2. Entitäten-Überblick	15
2.1 Die Entität Objekttyp.....	15
2.2 Die Entität Schlagwort.....	16
2.3 Die Entität Klassifikation.....	18
2.4 Die Entität Ort.....	19
2.5 Die Entität Person/Organisation	19
2.6 Die Entitäten Material, Technik und Funktion	21
2.7 Die Entität Periode/Epoche	22
3. Symptomanalyse zur Beeinträchtigung der Suche.....	23
3.1 Freitext-Suche.....	23
3.1.1 Objekte	24
3.1.1.1 Objekttitel.....	25
3.1.1.2 Objektbeschreibung	26
3.1.1.3 Inschrift/Beschriftung.....	26
3.1.1.4 Rollen.....	27
3.1.1.5 Material, Technik.....	28
3.1.1.6 Periode/Epoche	29
3.1.2 Personen.....	29
3.1.3 Organisation	29
3.1.4 Themen.....	30
3.2 Facettierte Suche [GLOSSAR-RELEVANT]	30
3.2.1 Geminderte Precision.....	30
3.2.1.1 Objekttyp.....	30
3.2.1.2 Schlagwort.....	31
3.2.1.3 Ort.....	32
3.2.1.4 Person/Organisation	34
3.2.2 Geminderter Recall.....	35
3.2.2.1 Objekttyp.....	35
3.2.2.2 Schlagwort.....	36
3.2.2.3 Ort.....	36
3.2.2.4 Person/Organisation	37
3.2.3 Abweichung von der Intension.....	39

3.2.3.1	Objekttyp.....	39
3.2.3.2	Schlagwort.....	40
3.2.4	Eindeutige Filterzuordnung.....	41
3.3	Suche über Objektverknüpfungen.....	41
3.3.1	Verknüpfung über Personen-/Körperschaftsseiten.....	41
3.3.2	Verbundene Objekte, ähnliche Objekte.....	42
4.	Ursachenanalyse.....	43
4.1	Limitierender Faktor „Datenbankstruktur“.....	43
4.1.1	Freitext-Erweiterungen in Thesaurus-/Stammdatenfeldern.....	44
4.1.1.1	Objekttyp.....	45
4.1.1.2	Ort.....	46
4.1.1.3	Person/Organisation.....	46
4.1.1.4	Material, Technik.....	48
4.1.2	Optionale Datenfelder für Pflicht-Angaben.....	49
4.1.3	Recherchemöglichkeiten.....	51
4.1.4	Nicht vorgesehene Datenfelder.....	52
4.1.4.1	Objekttyp.....	52
4.1.4.2	Funktion.....	53
4.1.4.3	Periode.....	53
4.1.5	Nicht vorgesehene Differenzierung.....	54
4.1.5.1	Schlagwort.....	54
4.1.5.2	Ort.....	55
4.1.5.3	Material und Technik.....	56
4.1.6	Keine ereignisbasierte Erfassung.....	57
4.1.6.1	Ort.....	57
4.1.6.2	Person/Organisation.....	58
4.1.7	Gruppenbearbeitungen – eine unterschätzte Gefahr.....	58
4.2	Limitierender Faktor „Kontrollierte Vokabulare“.....	59
4.2.1	Gemeinsame Normdatei (GND).....	59
4.2.2	Art and Architecture Thesaurus (AAT).....	61
4.2.3	Wikidata – ein “Datenhub”.....	64
4.2.4	Objektbezeichnungsdatei (OBG) und Hessische Systematik.....	65
4.2.5	Ortsvokabulare: Geonames, TGN, Pleiades.....	67
4.2.6	DDC-Sachgruppen und Iconclass.....	68
4.2.7	Weitere Vokabulare.....	70
4.2.8	Exkurs: Zum Umgang mit Normdaten in der Deutschen Digitalen Bibliothek.....	71

4.3	Limitierender Faktor „Erfassung im Museum“	73
4.3.1	Verschiedene Benennung gleicher Erschließungsinhalte	73
4.3.1.1	Fachvokabular und Alltagssprache	73
4.3.2	Variierende Granularität und Spezifität	74
4.3.2.1	Objekttyp und Objektgattung	74
4.3.2.2	Ort	74
4.3.2.3	Material	75
4.3.3	Begriffs-/Ortsketten; Begriffe in nicht-lemmatisierter Form	75
4.3.3.1	Begriffsketten	75
4.3.3.2	Ortsketten	76
4.3.3.3	Begriffe im Plural	76
4.3.4	Umgang mit Synonymen, Schreibvarianten und Oberbegriffen	77
4.3.4.1	Schlagwort	77
4.3.4.2	Ort	77
4.3.4.3	Person	78
4.3.5	Umgang mit Homonymen und gleichnamigen Personen/Orten	79
4.3.6	Diskrepanzen bei der Verschlagwortung	80
4.3.6.1	Überindexierung	80
4.3.6.2	Assoziative interpretative und sensorische Verschlagwortung	80
4.3.7	Diskrepanzen bei der Erfassung von Orten	81
4.3.8	Heterogene Vokabular-Nutzung	83
4.3.8.1	Person/Organisation	83
4.3.8.2	Material	84
4.3.9	KI-Nutzung für die Objekterfassung	84
4.3.10	Divergierende Handhabung der CARE-Prinzipien	85
4.3.10.1	Umgang mit kontrollierten Vokabularen	85
4.3.10.2	Einbeziehung von Herkunftsgesellschaften	85
4.3.10.3	Umgang mit Rezeptionsobjekten	86
4.3.10.4	Unpräzise und heikle Ortsangaben	86
4.3.10.5	Personen und Rollen	86
4.3.11	Weitere Limitationen	87
4.3.11.1	Objekttyp	87
4.3.11.2	Klassifikation	87
4.3.11.3	(Fiktive) Person	88
4.4	Limitierender Faktor „Export-Aufbereitung“	92
4.4.1	Keine Normdaten-Anreicherungen (OpenRefine)	92

4.4.1.1	Ort.....	92
4.4.1.2	Person/Organisation	93
4.4.1.3	Material und Technik.....	93
4.4.2	Diskrepanz zwischen Begriff und Bezeichnung	94
4.4.3	Keine Aufspaltung von Begriffs-/Ortsketten.....	95
4.4.4	Keine nachträgliche Typisierung.....	95
4.4.4.1	Schlagwort.....	95
4.4.4.2	Ort.....	95
4.4.4.1	(Fiktive) Person.....	95
4.4.5	Keine Datentransformationen für den Export (XSLT).....	96
4.4.5.1	Objekttyp und Objektgattung	96
4.4.5.2	Schlagwort.....	96
4.5	Limitierender Faktor „Herausforderungen und Grenzen des Portals“	98
4.5.1	Begrenzte Suchmöglichkeiten	98
4.6.2.2	Facettierte Suche.....	98
4.5.2	Keine Anzeige der (Export-)Datierung des Datensatzes.....	99
5.	Lösungsansätze.....	101
5.1	Stellschraube „Normdaten-Mapping“ in Zusammenarbeit mit Aggregatoren.....	101
5.1.1	Die Normdatenfeld-Priorisierung.....	101
5.1.2	Das kuniweb-Kategorien-Mapping.....	102
5.1.3	Feinschliff des Normdaten-Mappings	103
5.1.4	Ergebnisse und Grenzen.....	104
5.2	Stellschraube „Terminologiearbeit“	105
5.3	Stellschraube „Normdatenvermittlung“	107
5.3.1	Minimaldatensatz-Empfehlung	107
5.3.2	Normdatenfokussierter Beispieldatensatz.....	109
5.4	Stellschraube „Normdatenanreicherung“	111
5.4.1	OpenRefine.....	111
5.4.1.1	Aufsplitten von Orts- und Begriffsketten	112
5.4.1.2	Normdatenanreicherung durch den Reconciliation-Service.....	113
5.4.2	LidoRefine.....	114
6.	Ergebnisse und Ausblick [AUSSTEHEND].....	115
	Glossar.....	115
	Bibliographie.....	121
	Abbildungsverzeichnis.....	125
	Abbildungen	126