

DANTE – DATendrehscheibe für Normdaten und Terminologien

Tobias Helms – (Verbundzentrale des GBV [VZG], Göttingen)



Die Abteilung „Digitale Bibliothek“ der Verbundzentrale des GBV (VZG) betreibt verschiedenste, immer browserbasierte Erfassungsdatenbanken für materielles Kulturgut. Dazu zählen zum Beispiel **kuniweb** für kulturwissenschaftliche Objekte, **KENOM** für numismatische Objekte oder **naniweb** für naturwissenschaftliche Objekte. In der Datenbank **LaZAR** werden regionalwissenschaftliche Forschungsdaten erhoben, im Projekt **PrizePapers** werden historische Prisenpapiere erschlossen, welche auf Kaperungen zurückgehen, die im Verlauf der Frühen Neuzeit zur Kriegsstrategie der Seekriegsmächte zählten. Über **goobi** werden Digitalisierungsworkflows abgebildet und bibliothekarisches Material erfasst. Daneben gibt es eine Vielzahl von Repositorien, Dokumentenservern und Zeitschriftenservern auf Basis der freien Software **MyCoRe** über die hauptsächlich bibliografische Daten erfasst werden. So ist zum Beispiel **OpenAgrar** das gemeinsame Repository der Ressortforschungsinstitute sowie weiterer Bundesbehörden und Einrichtungen im Geschäftsbereich des Bundesministeriums für Ernährung und Landwirtschaft (BMEL).

Die von der Verbundzentrale betriebenen Erfassungsdatenbanken werden oft von Portalen für die frei zugängliche Präsentation im Internet ergänzt.

Um die Datenqualität merklich zu erhöhen und die Verwendung, Auffindbarkeit und Weiternutzung zu verbessern, werden möglichst viele Felder der Eingabemasken mit nationalen und internationalen Normdatenquellen verknüpft.

Dafür wurden bei der Verbundzentrale zentrale Dienste geschaffen, die in Echtzeit beispielsweise fertige biografische Angaben aus dem Verbundkatalog liefern, welche wiederum für die Literaturangaben zu den Objekten genutzt werden. Personen, Institutionen und Schlagworte werden live mit der Gemeinsamen Normdatei (GND) der Nationalbibliothek verknüpft. Orte werden mit geonames oder den Daten des Bundesamts für Kartographie und Geodäsie verlinkt.

Nicht alle externen Einrichtungen, welche Normdaten zum Download anbieten, stellen auch (ausreichende) Online-Schnittstellen bereit. Zudem wurden und werden viele kleinere Vokabulare in den jeweiligen Erfassungsdatenbanken von den Bearbeiter*innen selbst gepflegt. So gibt es unter anderem jeweils in **kuniweb**, **naniweb** und **kenom** ein Technikvokabular und ein Materialvokabular. In jeder Datenbank gibt es ein Vokabular für Währungen nach ISO 4271 und Sprachen nach ISO 639-1 usw. Selbst kleine Auswahllisten wie „Geschlecht“, „Namenspräfix“ oder „Identifier-Typ“ werden als Normdatei geführt, da sich erfahrungsgemäß mit Erfassungsrichtlinien für Freitextfelder auf Dauer keine Einheitlichkeit erzielen lässt. Diese parallele Haltung, Pflege und Aktualisierung der gleichen oder ähnlicher Terminologien ist aufwendig und bindet Ressourcen.

Seit 2017 werden immer mehr dieser Terminologien, Vokabulare, Normdaten und Listen in einem zentralen Vokabularserver mit dem Namen DANTE gepflegt. Mit der Zentralisierung der Normdaten soll die bisher noch verbreitete, oftmals mehrfache lokale Haltung von Normdaten und Vokabularen in unseren Systemen aufgelöst werden. Die Zentralisierung vereinfacht die Pflege, Wartung und Updates enorm. Der Vokabularserver wurde auf Basis der Software **easydb5** umgesetzt, das Datenmodell orientiert sich an den Anforderungen der Getty-Vokabulare, der GND-Ontologie und nutzt an vielen Stellen Prinzipien des Simple Knowledge Organisation System (SKOS). Damit können alle bisher infrage kommenden Typen von Vokabularen und Normdaten nach DANTE importiert werden.

Sollen in DANTE eingepflegte Vokabulare in einer anderen Anwendung, also zum Beispiel in der Erfassungsdatenbank **kuniweb**, genutzt

werden, müssen sie veröffentlicht und mit einer Lizenz versehen werden. Ab diesem Zeitpunkt sind sie damit öffentlich einsehbar und bei Einhaltung der entsprechenden Lizenz selbstverständlich nachnutzbar.¹

DANTE stellt für bestimmte Vokabulare und Nutzergruppen ein festes Set von Masken bereit. Ein einfaches Vokabular wird mit einer Basismaske erfasst und ausgeliefert, die Maske für Personenkonzepte unterscheidet sich grundlegend von der Maske für Ortskonzepte. Zugrunde liegt aber stets das gleiche Datenmodell. Über die sogenannten „Tags“ und „Workflows“ sorgen automatische Aktions- und Zugriffskontrollen beispielweise dafür, dass nur bestimmte Nutzergruppen Datensätze redaktionell freigeben können.

Die easydb5-Oberfläche wird komplett über eine HTTP-Restful-API² gesteuert. Jede Kommunikation mit dem Server läuft über diese API. Dazu zählen sogar Änderungen des Datenmodells. Diese bestehende API-Funktionalität der Software erlaubt das einfache Einbinden des Vokabularservers in unsere bisherigen Normdatendienste. Um aus DANTE heraus Live-Abfragen an andere Normdatenlieferanten machen zu können, entwickelt die VZG gemeinsam mit der Firma Programmfabrik und anderen Kunden Plugins, welche auf github unter einer MIT-Lizenz veröffentlicht werden.³

JSKOS

Um die verschiedenen Metadatenformate der oft internationalen Normdatenquellen besser miteinander zu verbinden, entwickelte die Stabsstelle „Forschung und Entwicklung“ das Metadaten austauschformat JSKOS für Wissensorganisationssysteme.⁴

JSKOS definiert eine JavaScript Object Notation (JSON)-Struktur, um Wissensorganisationssysteme (KOS) zu codieren, also Klassifikationen, Thesauri und Normvokabulare. Der Hauptteil von JSKOS ist mit dem Simple Knowledge Organization System (SKOS) und der JavaScript Object Notation für verknüpfte Daten (JSON-LD) kompatibel. JSKOS kann jedoch ohne Erfahrung mit einer dieser Technologien eingesetzt werden.

Gerade die JSKOS-Nutzung von JSON als kompaktes, einfach lesbares Datenformat macht diesen neuen Standard überlegen. XML ist für dynamische Webanwendungen nicht performant nutzbar. JSKOS ist das zentrale Ausgabeformat der Schnittstelle des Vokabularservers⁵. Die Suggest- und Autocomplete-Ergebnisse werden nach dem OpenSearch-Standard⁶ ausgeliefert, welcher z. B. in der Adresszeile jedes modernen Browsers genutzt wird.

Ausblick

Die Zentralisierung von Normdatendiensten ist eine komplexe Aufgabe. Leider bieten viele Datenlieferanten zwar Datendumps, aber (noch) keine offenen Schnittstellen an, sodass diese Daten bei der Verbundzentrale „zentral“, aber dennoch „lokal“ vorgehalten werden müssen. Durch die Verwendung der bereitgestellten URIs ist die Umstellung auf einen möglicherweise zukünftig folgenden Schnittstellendienst der datengebenden Einrichtung allerdings problemlos.

Auch die Zentralisierung „kleiner Vokabulare“ wie Material, Technik, Ethnie oder Münznominal verspricht eine Arbeitserleichterung und erlaubt vor allem eine unkomplizierte Weiternutzung und Bereitstellung an Dritte. Der Austausch von Normdaten mit anderen Nutzer*innen wird durch JSKOS stark vereinfacht, denn die Verbundzentrale stellt so alle Normdaten in einem einheitlichen Format zur Verfügung. Die Veröffentlichung der Vokabulare macht diese überhaupt erst Dritten zugänglich und kann anderen Projekten durch die (Nach-)Nutzung erheblich Arbeit abnehmen. Durch gemeinsames Arbeiten an Vokabularen können Synergien nun sehr viel einfacher genutzt werden. Das Mapping der Terminologien untereinander wird zudem erheblich vereinfacht, sobald sich beide in DANTE befinden. Zurzeit werden beispielsweise Normdaten zweier numismatischer Verbände aufeinander gemappt.

Die Verbundzentrale nutzt Normdateien in einer Vielzahl von Erfassungsdatenbanken und Portalen. Die Etablierung dieser Datendrehzscheibe war und ist zwar mit viel Aufwand verbunden, welcher sich aber auf Grund der Vielzahl der Anwendungen bereits mit der Zentralisierung der ersten Normdaten amortisiert hat.

1. <https://dante.gbv.de> [letzter Zugriff bei allen Fußnoten: 02.02.2018].
2. <http://5.docs.easydb.de/docs/technical/api/?node=7>.
3. <https://github.com/programmfabrik/easydb-custom-data-type-gnd>; <https://github.com/programmfabrik/easydb-custom-data-type-gvk>; <https://github.com/programmfabrik/easydb-custom-data-type-geonames>; <https://github.com/programmfabrik/easydb-custom-data-type-gn250>; <https://github.com/programmfabrik/easydb-custom-data-type-getty>; <https://github.com/programmfabrik/easydb-custom-data-type-georef>.
4. <https://gbv.github.io/jskos/jskos>.
5. <http://api.dante.gbv.de/>.
6. http://www.opensearch.org/Specifications/OpenSearch/Extensions/Suggestions/1.1/Draft_1.