

VON LOKALEN NORMDATEN ZUR DATENDREHSCHEIBE – NEUE NORMDATENDIENSTE DER VZG

Tobias Helms

Abteilung Digitale Bibliothek
Verbundzentrale des GBV (VZG)
Platz der Göttinger Sieben 1
37073 Göttingen

KURZDARSTELLUNG: Die Verbundzentrale bietet für kulturelle Einrichtungen verschiedene Erfassungsdatenbanken auf Basis der Software easydb an. Diese Datenbanken nutzen an vielen Stellen kontrollierte Vokabulare, Thesauri und Normdaten, welche auf nationalen und internationalen Normdatenquellen basieren und auf diese verlinken. Auch verschiedene Instanzen der Repository-Software MyCoRe und Portalanwendungen benötigen stabile Normdatendienste. Im Folgenden werden verschiedene Neuheiten bezüglich Normdatenverknüpfungen in Erfassungsdatenbanken bei der Verbundzentrale aufgezeigt und die Strategie zum weiteren Einsatz erörtert.

1. EINFÜHRUNG

Die Verbundzentrale bietet für kulturelle Einrichtungen verschiedene Erfassungsdatenbanken auf Basis der Software easydb an. Unter anderen bedient „kuniweb“ kulturwissenschaftliche, „naniweb“ naturwissenschaftliche und „KENOM“ numismatische Belange. Diese Datenbanken nutzen an vielen Stellen kontrollierte Vokabulare, Thesauri und Normdaten, welche auf lokalen, nationalen und internationalen Normdatenquellen basieren und auf diese verlinken. Auch verschiedene Instanzen der Repository-Software MyCoRe und Portalanwendungen benötigen stabile Normdatendienste. Im Folgenden werden verschiedene Neuheiten bezüglich Normdatenverknüpfungen in Erfassungsdatenbanken bei der Verbundzentrale aufgezeigt und die Strategie zum weiteren Einsatz erörtert.

für Kartographie und Geodäsie oder der Gemeinsame Verbundkatalog (GVK). Unbekanntere, aber in Deutschland dennoch häufig genutzte, eher spartenspezifische Vokabulare sind zum Beispiel „Hessische Systematik“, „Bayerische Oberbegriffsdatei“, „ITIS“, „CHEBI“ oder „WORMS“

Durch die Teilnahme verschiedenster Einrichtungen an den Erfassungsdatenbanken sind die Bedürfnisse sehr heterogen. Für bestimmte Einrichtungen sind Spezialvokabulare oder die Anbindung spezifischer externer Dienste obligatorisch für die Nutzung der Erfassungsdienste. Deshalb wurden und werden zusätzlich zu den „Standardnormdaten“ weitere (Norm-) Datenverknüpfungen entwickelt. In den letzten Monaten sind folgende drei Verknüpfungen neu entwickelt worden.

2.1 ADABWEB

2. NEUE NORMDATENDIENSTE IN ERFASSUNGSDATENBANKEN FÜR MUSEALE SAMMLUNGEN

Bekannte Datenquellen bei der Erfassung musealer Bestände sind geonames, iconclass, GND, die Ortsinformationen des Bundesamtes

In der Erfassung für museales Kulturgut kuniweb können Objekte nun über eine Schnittstelle live mit einem Eintrag aus der ADABweb verknüpft werden. Die ADABweb ist das Denkmalpflege-Informationssystem der Länder Niedersachsen und BadenWürttemberg.

Dort werden zum Beispiel durch die MitarbeiterInnen des Niedersächsischen Landesamtes für Denkmalpflege in Hannover alle archäologischen Fundstellen in Niedersachsen und Baden-Württemberg verzeichnet. Mittels der neuen Verknüpfung kann beispielsweise ein in kuniweb erfasster Tonkrug / Faustkeil / Dolch mit der entsprechenden Fundstelle in der ADABweb verknüpft werden.

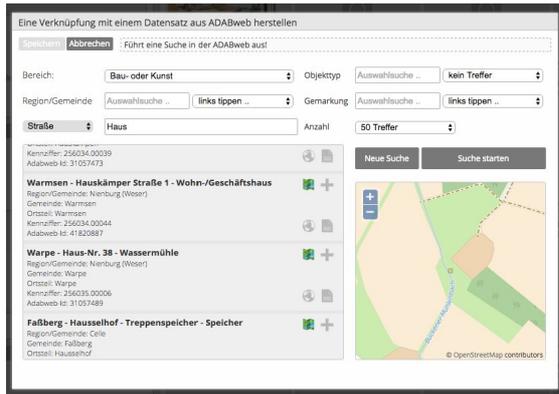


Abb. 1: ADABweb-Suche mit Kartendarstellung

Neben der eindeutigen ID werden die wichtigsten Kerndaten aus der ADABweb zur Volltextindexierung nach kuniweb übernommen. Für die Ansicht der Gesamtdaten wird immer die ADABweb live angefragt und die entsprechenden Informationen in kuniweb angezeigt.

2.2 NUMISMATICS.ORG

Mit der numismatischen Erfassungsdatenbank KENOM werden Münzen, Medaillen, Geldscheine, Siegel und Varia detailliert inventarisiert. Große Bereiche der numismatischen Objekte sind in Standardwerken gelistet und typisiert. In den Projekten „Coinage of the Roman Republic“ (CRRO) und „Online Coins of the Roman Empire“ (OCRE) wurden von der „American Numismatic Society“ und dem "Institute for the Study of the Ancient World" über <http://numismatics.org> die numismatischen Standardwerke "Roman Imperial Coinage" und "Roman Republican Coinage" online verfügbar gemacht. Diese sind über gut dokumentierte Schnittstellen frei nutzbar. In der Erfassungsdatenbank KENOM gibt es nun die Möglichkeit eine Live-Suche durchzuführen und die entsprechende Münze so mit der Typenbezeichnung und der URI zur Quelle zu versehen.

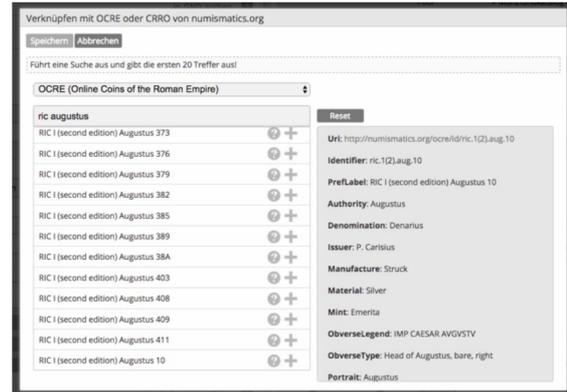


Abb. 2: Suche in OCRE

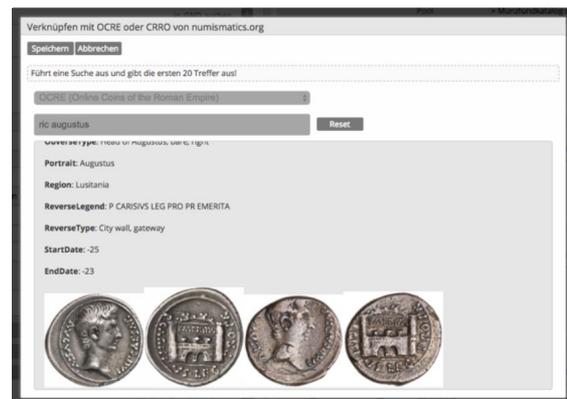


Abb. 3: Überprüfung des Treffers mit Bildern anderer Einrichtungen

2.3 GFBIO

Die Datenbank naniweb erlaubt naturwissenschaftlichen Sammlungen die Inventarisierung ihrer Bestände. Die verschiedenen naturwissenschaftlichen Disziplinen und Fächer nutzen eine große Bandbreite an Klassifikationen und Systematiken zur Kategorisierung ihrer Objekte. Das Vorhalten und regelmäßige Updates der vielen einzelnen Vokabulare in der Datenbank selbst wäre sehr aufwendig.

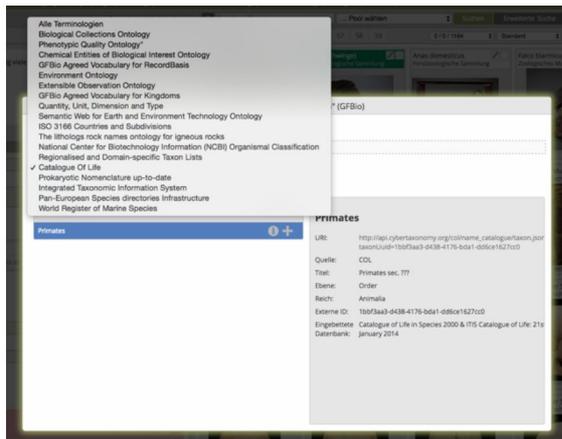


Abb. 4: Suche über verschiedenste naturwissenschaftliche Vokabulare

Dank der „German Federation for Biological Data“ (GFBio) werden die wichtigsten naturwissenschaftlichen Normdaten über einen kostenlosen Terminologieserver mit entsprechenden Schnittstellen bereitgestellt. naniweb wird aktuell mit einer live-Schnittstelle zu diesen Daten ausgestattet.

3. LOKAL VS. ZENTRALISIERUNG

Die oben genannten neuen Verknüpfungsdienste greifen über Schnittstellen live auf externe Datenbestände zu. Diese Technik wird in den Erfassungsdatenbanken auch für weitere Datenquellen verwendet. Zum Beispiel werden Literaturverknüpfungen mit dem Verbundkatalog live gesucht und verknüpft.

Andere Normdaten liegen bisher noch in den jeweiligen Erfassungssystemen lokal vor. Vorteile (zum Beispiel hinsichtlich der Performance) hat die lokale Haltung kaum, die Erfassungsdatenbanken sind wie die Live-Schnittstellen Webanwendungen.

Die Ausfallquote der Live-Datendienste ist gering und quasi vernachlässigbar. Eine zentrale Haltung der bisher lokalen Normdaten vereinfacht die Pflege, Wartung und Updates der teilweise mehrfach lokal existierenden Vokabulare um ein Vielfaches.

4. VOKABULARSERVER DATENDREHSCHLEIBE

Lokale Vokabulare werden in Zukunft über den Vokabularserver DANTE (Datendreh-scheibe für Normdaten und Terminologien) verwaltet und über zentrale, öffentliche

Dienste den einzelnen Nutzungsanwendungen bereitgestellt.

Mittels der Software easydb5 der Firma programmfabrik aus Berlin können die komplexen Anforderungen umgesetzt werden. Das Datenmodell orientiert sich an den Anforderungen der Getty-Vokabulare, der GND-Ontologie und nutzt an vielen Stellen SKOS-Prinzipien. Alle bisher bekannten Anforderungen sind umgesetzt und alle in Frage kommenden Typen von Vokabularen und Normdaten können nach DANTE importiert werden.

Die Software bietet die Möglichkeit für bestimmte Vokabulare und Nutzergruppen verschiedene Masken bereit zu stellen. Ein einfaches Vokabular wird mit einer Basismaske ausgeliefert, die Personenmaske unterscheidet sich grundlegend von der Ortsmaske. Zu Grunde liegt aber immer das gleiche Datenmodell.

Über die sogenannten „Tags“ und „Workflows“ sorgen automatische Aktionskontrollen unter anderem dafür, dass nur bestimmte Nutzergruppen Datensätze redaktionell freigeben oder Datensätze zusammenlegen dürfen.

Die easydb5 –Oberfläche wird komplett über eine HTTP-Restful-API gesteuert. Jede Kommunikation mit dem Server läuft über die API. Dazu zählen sogar Änderungen des Datenmodells.

Diese bestehende API-Funktionalität erlaubt das einfache Einbinden des Vokabularservers in unsere bisherigen Normdatendienste.

Damit die für die easydb4 entwickelten Live-Anbindungen auch für die easydb5 funktionieren, werden gemeinsam mit programmfabrik und anderen Kunden Plugins entwickelt und auf github unter LGPL veröffentlicht.[1]

5. JSKOS

Um die verschiedenen Metadatenformate der weltweiten Quellen besser miteinander in Verbindung zu bringen, entwickelte die Stabsstelle „Forschung und Entwicklung“ das Metadaten austauschformat JSKOS für Wissensorganisationssysteme. [3]

JSKOS definiert eine JavaScript Object Notation (JSON)-Struktur, um Wissensorganisationssysteme (KOS) zu codieren, also Klassifikationen, Thesauri und

Norm-vokabulare. 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 moderne und dynamische Webanwendungen nicht performant nutzbar. JSKOS ist das zentrale Ausgabeformat des Vokabularservers.

Die Suggest- und Autocomplete-Ergebnisse werden nach dem OpenSearch-Standard ausgeliefert.[3]

6. SCHLUSS

Die Zentralisierung von Normdatendienste ist eine komplexe Aufgabe. Leider bieten viele Datenlieferanten zwar Datendumps, aber keine offenen Schnittstellen, sodass diese Daten bei der Verbundzentrale „zentral, lokal“ vorgehalten werden müssen.

Auch die Zentralisierung „kleiner Vokabulare“ wie Material, Technik, Stil/Epoche oder Münznominale verspricht eine Arbeits-erleichterung und erlaubt vor allem eine unkomplizierte Weiternutzung durch und Bereitstellung an Dritte. Der Austausch von

Normdaten mit anderen NutzerInnen wird durch JSKOS vereinfacht, denn die Verbundzentrale wird alle Normdaten in einem einheitlichen Format zur Verfügung stellen. Das Mapping der Terminologien untereinander wird zudem erheblich vereinfacht.

6. QUELLEN / VERWEISE

1. <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>
2. Voss, Jakob:
<https://gbv.github.io/jskos/jskos>
3. http://www.opensearch.org/Specifications/OpenSearch/Extensions/Suggestions/1.1/Draft_1