

Datenmodellierung – Kernpunkte

IFS/RCA 11.8.2021

Im Folgenden sind die wichtigsten Kernpunkte aus der IFS-Dokumentation zur Datenmodellierung zusammengefasst.

Datensammlung – Grundlage für Publikationen – Forschungsdatenbank

Die neue IFS-Datenbank dient unterschiedlichen Zwecken:

1. Datensammlung

Dies ist der Grundauftrag des IFS. Alle Informationen zu Fundmünzen der Schweiz und Liechtensteins werden zentral erschlossen.

2. Grundlage für Publikationen

Auszüge aus der Datensammlung werden entweder in der online-Datenbank und ggf. in gedruckter Form publiziert (Monographienreihe IFS, Bulletin IFS ITMS IRMS). Die Daten sollen in der Datenbank erfasst und bereinigt werden. Formatierte Exporte gewährleisten, dass ausserhalb der Datenbank nur formale Anpassungen notwendig sind.

3. Forschungsdatenbank

Die IFS-Datenbank ist ein Forschungsinstrument, das gezielte Suchen und Auswertungen im Bereich Archäologie & Numismatik ermöglicht und unterstützt. Die Einbindung in die internationalen Netzwerke muss gewährleistet sein.

Erfassung – Auswertung

Die wichtigsten Fragen bei allen zu erhebenden Daten sind:

1. Was soll ausgewertet können?

All das benötigt Normierung und somit eine Umsetzung in Form von Normdaten bzw. mehrsprachig hinterlegten Listen.

2. Was ist individuell und muss nicht systematisch gesucht und ausgewertet werden können?

Hierfür sind Freitextfelder möglich, entweder nur in der user-Sprache oder ggf. in mehreren Sprachvarianten.

LEGO BRICK System

- Die Datenbank muss modular aufgebaut sein (ggf. muss die Entwicklung etappiert werden).
- Andocken (Datenaustausch, APIs für LinkedOpenData-Vernetzung) muss auf verschiedenen Ebenen möglich sein
Beispiele: archäologische Dienste, Auswertungen, APIs wie OCRE & Co. oder Metagrid
- In Zukunft müssen Erweiterungen möglich sein.
vgl. Bsp. Pflichtenheft: Bilder, Archivalien, Visualisierungen

Komplexität der Datenmodellierung

Die Komplexität ist bedingt durch:

- Komplexität der Numismatik
- Stand der Digital Numismatics, vor allem nomisma.org, OCRE & Co.
- Komplexität der bereits vorhandenen Daten
- Redundanzen sollen soweit möglich vermieden werden. Punktuell werden Redundanzen bewusst modelliert (Bsp. *mint* – *modern_place*, Gemeinde/Ort – *place*).

0 – n / 1 – n-Beziehungen und rekursive Beziehungen

Für die MySQL-Mock-up-Datenbank und den Datentransfer müssen alle derartigen Beziehungen aufgelöst werden. In der MySQL-Datenbank wurden dafür ausgelagerte Unter-Tabellen angelegt, welche zur Kennzeichnung die folgenden Komponenten im Namen tragen:

***_junction (Endung bei wenigen Verbindungstabellen weggelassen)
 ***_relation
 ***_part

Herauslösen von Varianten in Erweiterungstabellen ***_name

Personen und Institutionen (*organisation*), aber auch Münzherrschaften (*authority*) und Münzstätten (*mint*) können mit unterschiedlichen Benennungen erfasst sein:

- Bsp. *ruler*: 1 Person, verschiedene authorities / Namen / Regierungszeiten
 Maria Theresia als Königin von Ungarn oder als Markgräfin von Burgau
- Bsp. *modern_person*: 1 Person, verschiedene Namen, Anbindungen an verschiedene Institutionen
 ledig/verheiratet; tätig in verschiedenen Museen, teilweise überlappend oder gleichzeitig mit verschiedenen Teilzeitpensen
- Bsp. *mint*: ein geographischer Ort (*place*), verschiedene Namen in unterschiedlichen Epochen
 Rom: Antike Roma, Mittelalter/Neuzeit Rom
 modern_place Istanbul: Antike griechisch Byzantion, römisch Byzantium, 4. Jh. Constantinopolis
- Bsp. *organisation*:
 Schweizerisches Landesmuseum (SLM) wird zu Schweizerisches Nationalmuseum (SNM)

Bei diesen Tabellen wurden die Daten aufgeteilt in die unveränderlichen Kerndaten und eine Erweiterungstabelle ***_name für die Namen und weitere wechselnde Daten.

Mehrsprachigkeit im IFS

Das IFS ist auf die drei Landessprachen d/f/i ausgelegt. Grundsätzlich ist angestrebt, dass Publikationen zu Funden in der regionalen Landessprache erscheinen; Ausnahmen sind sehr gut zu begründen und mit den Partnern auszuhandeln.

Bsp. Bulletin IFS ITMS IRMS: dreisprachige Publikation, Einträge werden automatisch in der jeweiligen Gemeindesprache publiziert

Im Hinblick auf die internationale Einbindung soll Englisch vollwertig integriert werden.

Um einen möglichst grossen Teil der erfassten Daten mehrsprachig zu erschliessen – und sie zugleich auswertbar zu machen –, müssen

- möglichst viele Daten über Dropdowns/Wertelisten/Normdaten erfasst werden.
- möglichst wenige Informationen in Freitext (sprachsensitiv) erfasst werden.

Normdaten-Tabellen – Erfassungstabellen

Wir unterscheiden zwischen Normdaten (gelb, dunkelblau, hellblau) und Bereichen der freien Erfassung (grün); für die Farben vgl. den Graphen im github repository (https://github.com/IFS-ITMS-IRMS/SICF_MySQL/tree/main/Documentation/Coloured_Graph).

Die Normdaten sind hinterlegte Begriffslisten/Thesauri, z. T. abgeschlossen (gelb) und zum Teil kontrolliert erweiterbar (dunkelblau, hellblau).

Basislisten (gelb):

- in sich abgeschlossene Listen
- Ggf. greifen mehrere Tabellen auf diese Basislisten zu.

Es wird eine Datenmenge zur Verfügung gestellt, die:

- möglichst vollständig das beim Datentransfer in die neue Applikation Vorhandene abbildet
- den Grossteil der in der Schweiz und in Liechtenstein zu erwartenden Münzen abdeckt
- z. K.: Vollständigkeit aus numismatischer Sicht soll und kann nie erreicht werden!
Bsp. moderne Münzmetalle.

Änderungen an diesen Basislisten (neue Einträge, Korrekturen) sollen nur im IFS nach interner Prüfung vorgenommen werden.

harte Normdaten (dunkelblau):

Es wird eine Datenmenge zur Verfügung gestellt, die:

- möglichst vollständig das beim Datentransfer in die neue Applikation Vorhandene abbildet
- den Grossteil der in der Schweiz und in Liechtenstein zu erwartenden Münzen abdeckt
- z. K.: Vollständigkeit aus numismatischer Sicht soll und kann nie erreicht werden!
Bsp. Integration der Normdaten für die gesamte antike griechische Numismatik & Roman Provincial Coinage für die 100–300 Ex. bzw. Typen in den von uns erfassten Münzfunden

Änderung bestehender Datensätze müssen vom IFS validiert werden und sollen daher nur im IFS möglich sein.

Bei Neuerfassungen soll:

- der neu erfasste Datensatz durch den Ersterfasser sofort genutzt werden können.
- eine automatisierte Information ans IFS erstellt werden.
- eine rasche Überprüfung und Freischaltung für alle User durchs IFS vorgenommen werden.
- falls nötig eine Korrektur und Vervollständigung des Datensatzes durchs IFS vorgenommen werden.
- Für vom IFS abgelehnte Datensätze eine Begründung für die Ablehnung innerhalb oder ausserhalb der Datenbank gespeichert werden.

weiche Normdaten (hellblau):

Es wird eine Datenmenge zur Verfügung gestellt, die:

- möglichst vollständig das beim Datentransfer in die neue Applikation Vorhandene abbildet
- den Grossteil der in der Schweiz und in Liechtenstein zu erwartenden Münzen abdeckt
- z. K.: Vollständigkeit aus numismatischer Sicht soll und kann nie erreicht werden (siehe oben harte Normdaten)!

Änderung bestehender Datensätze müssen vom IFS validiert werden und sollen daher nur im IFS möglich sein.

Bei Neuerfassungen soll:

- der neu erfasste Datensatz durch den Ersterfasser sofort genutzt werden können.
- eine automatisierte Information ans IFS erstellt werden.
- eine rasche Überprüfung und Freischaltung für alle User durchs IFS vorgenommen werden.
- falls nötig eine Korrektur und Vervollständigung des Datensatzes durchs IFS vorgenommen werden.
- Für vom IFS abgelehnte Datensätze eine Begründung für die Ablehnung innerhalb oder ausserhalb der Datenbank gespeichert werden.

Erfassungstabellen (grün)

Unter Erfassung (grün) werden hier diejenigen Tabellen zusammengefasst, die im Alltag in hoher Autonomie erstellt und verändert werden: Erfassung von neuen Fundmeldungen (*find*), Bearbeitung von Fundobjekten (*find_object*, *coin*, *other_object*).

Ziel ist, dass Grunddaten über eine Schnittstelle integriert werden können, um in sensiblen Daten Tippfehler zu vermeiden :

- Informationen zum Komplex/zur Grabung (Grabungs-/Objektnummer, Geodaten, ggf. weiteres), falls bereits in Partner-Institution vorhanden.
- Informationen zum *find_object* (Inv., ggf. Gewicht, Geodaten, FK/Pos., archäologische Informationen), falls bereits in Partner-Institution vorhanden.
-

Sprachenverwendung in Datenmodellierung

Aktuell ist in den Tabellen Datenmodellierung und im Dokument Datenmodellierung – Herausforderung ein Sprachen-Mix verwendet.

- deutsche Begriffe:
Eindeutige Begriffe bzw. im IFS etablierte Bezeichnungen im Text und ggf. in der Kommentarspalte sollen beibehalten werden; ggf. mit französischer und/oder italienischer Version kombiniert.
- englische Begriffe:
Sollen soweit möglich eingesetzt werden, falls diese eindeutig bzw. nomisma.org-konform sind (nomisma.org ontology, NUDS).

In der Umsetzung der Programmierung soll nur eine Sprache verwendet werden: englisch (die Wahl einer anderen Sprache müsste gut begründet werden).

Ein Teil der Übersetzungsarbeit in den Tabellen/classes ist noch zu leisten, in enger Anbindung an die Ontologie von nomisma.org und – falls Begriffe dort noch nicht existieren – in enger Abstimmung mit der numismatischen und der archäologischen Linked Open Data (LOD) Community.

Das Dokument Datenmodellierung_Kernpunkte.docx ist aktuell ausschliesslich in Deutsch verfasst.

user

Die *user*-Modellierung ist relevant für die:

- Erfassung von Daten (inkl. Historisierung)
- Bearbeitung von Daten (inkl. Historisierung)
- gespeicherte Abfragen
- persönliche Notizen / Beschlagwortungen

Über die *id_user* werden zudem die Zugriffsrechte gesteuert.

Bei user-bezogenen Feldern setzen wir «id_user» als foreign key ein. Da die User-Verwaltung stark von der Implementierung abhängt, verzichten wir auf die Modellierung. Wir gehen aber davon aus, dass jeder user eine eindeutige id erhalten wird und einer oder mehreren *organisations* zugeordnet werden kann.

Achtung: Es gibt auch Personen, die mit keiner *organisation* verbunden sind.

Achtung: Die institutionelle Verbindung einer Person hat rechtliche Konsequenzen!

In der MySQL-Datenbank ist der user so modelliert, dass die bereits vorhandenen Daten mit ihrer Historisierung transferiert werden können (Datenherkunft, Datenersteller, letzte Mutation).

Zur Historisierung generell vgl. Pflichtenheft Kap. 3.1.1 und 3.1.10.

Felderreihenfolge in Datenmodellierung aktueller Stand

Die Felderreihenfolge ist noch festzulegen. Aktuell entspricht die Felderreihenfolge der einzelnen Tabellen der inhaltlichen Logik. Das bedeutet, dass die jetzige Felderreihenfolge

- **nicht zwingend die Reihenfolge der Felder in der neuen Datenbank sein soll**
- **nicht zwingend der gewünschten Reihenfolge im GUI/in den Masken (Erfassung, Abfragen) entspricht**

GUI/Nutzeroberfläche: verschiedene Versionen

Für die Arbeit mit der neuen IFS-Datenbank sind unterschiedliche Nutzeroberflächen notwendig:

- **einfache Erfassung/Bearbeitung/Abfrage:**
Hierbei soll nur die user-Sprache sichtbar sein. Selten genutzte Felder sind ausgeblendet und können nach Bedarf eingeblendet werden.
- **Experten-Modus:**
Weitere Sprachen können angezeigt werden, freie Textfelder können in mehreren Sprachen ausgefüllt werden. Selten genutzte Felder sind ausgeblendet und können nach Bedarf eingeblendet werden.
- **ADMIN-Modus:**
Alle Felder sind sichtbar und bearbeitbar, inkl. Administrationsfelder (Bsp. public TRUE/FALSE).

nomisma.org

Um die Anbindung der IFS-Daten in die internationale Community optimal vorzubereiten, sollen für die Feldernamen wenn immer möglich dieselben Begriffe wie in der nomisma.org-Ontologie verwendet werden.

In den Normdaten (gelbe Listen) werden – sofern vorhanden – die entsprechenden nomisma.org-IDs zu jedem Eintrag erfasst (Feld *nomisma_org_id*). Eingesetzt wird dabei nur die ID, da durch diese Vorgehensweise die Werte aktuell bleiben, auch wenn ggf. nomisma.org in Zukunft anders gehostet wird. Vgl. hierzu auch die aktuelle Diskussion des Transfers von <http://> zu <https://>.

Verweise auf weitere Normdaten-Bestände

nomisma.org enthält numismatische Konzepte. Für andere Konzepte werden andere Normdatensätze verwendet, die vorab evaluiert und dann regelhaft in der nomisma.org-Welt verwendet werden.

- Für die Literatur (*bibliography*) wird auf **DAI ZENON** verwiesen.
- Für *organisation* wird auf die **GND-ID** verwiesen (MARC21).

- Für *ruler* und *modern_person* wird auf die **GND**-ID verwiesen (MARC21).
- Für Orte (*place*, *mint*) wird auf [Pleiades](#) und/oder [GeoNames](#) verwiesen.

weitere externe Links

classes ***_external_reference

Dies sind Links zwischen einem (Norm-)Daten-Eintrag und weiteren Datenbeständen
Externe Datenbestände, die mehrfach genutzt werden und vertrauenswürdig sind, können in der class *external_reference_data_source*: hinterlegt werden.

classes ***_fremde_id

Dies sind Links zu Persistent Identifiern von Partner-Institutionen.

Bsp. IMDAS-Key von *archaeological_event* oder von *find_object*

Bsp. doi von *primary_source_documents*, vergeben von Staatsarchiven

In der Datenmodellierung verwendete Terminologie

Vorgabe «Text dort erfassen, wo er hingehört»

In der Datenmodellierung ist vorgesehen, dass Texte dort erfasst und angezeigt werden, wo sie hingehören.

Texte – Bemerkungen (remark) & Kommentare (comment) & Prosa (text)

remark

Bemerkungen zu den Erfassungen *find*, *find_object*, *coin*, *other_object*, *bibliography*, *primary_source_document* (grün)

comment

Kommentare zu Normdaten & Basislisten (gelb – dunkelblau – hellblau)

text

Prosa, Freitext

Bsp. Datierungen ausformuliert (vor, nach, ca., Meister \$\$\$, gest. \$\$\$...)

Bsp. Anzahl ausformulieren

Bsp. Kommentare ***_biblio_junction, ***_primary_source_document_junction

comment public

Dabei handelt es sich um Kommentare zu Normdaten, die in 4 Sprachen angelegt sind.

Bsp. Erfassungshilfe durch Mit-Anzeige im Dropdown-Menü: Roma – «als antike Münzstätte»

Bsp. Definitionen, Rappen/rappen – Trennung von 1 Rappen/1 centime

comment internal –

Hier wird nur 1 Feld für alle Sprachen benötigt.

remark public

Öffentliche Kommentare sollen in 4 Sprachen angelegt werden können.

Bsp. Text zur Fundstelle, zur Überlieferung, Bearbeitungsspuren, etc.

remark internal

Für interne Kommentare genügt 1 Feld für alle Sprachen.

Bsp. Kommentare, interne Angaben wie Adressen Privatpersonen, Kaufpreise, «nochmals zu überprüfen»

text

Wird in der Regel in 4 Sprachen angelegt, als Bestandteil von Feldernamen.

ggf. reicht 1 Sprache: Bsp. *bibliography* «quotation_text»

«Name» – Bezeichnungen**name** = echter Name, ausgeschriebenBsp. *mint, modern region, modern state, commune*, Personen

Bsp.

name_german
name_french
name_italian
name_english

value = Wertv. a. Inhalt von Hilfstabellen (darunter können auch Normdaten-Tabellen sein)
(effektiver Inhalt des Feldes, für Menschen lesbarer Inhalt der Tabelle)Bsp. *language, gender, region category, epoch, era, period*

Bsp.

value_german
value_french
value_italian
value_english

alternative (als Kurzform von alternative label)Dabei handelt es sich um eine andere Bezeichnung für denselben Wert (name oder value), ggf. mehrere Einträge/verschiedene Werte im selben Feld, getrennt durch "; " in Feldernamen
« **alternative** ».

Auch veraltete Namen können als «alternative label» geführt werden, als Hauptbezeichnung (preferred label) wird der neue Wert eingesetzt.

Im ersten Eintrag wird das «preferred label» eingetragen, das durchs IFS festgesetzt wird.

Bsp. *material* – «value_alternative_german» etc.Bsp. *ruler* «name_german_alternative» etc.

Die Begriffe in «***_alternative_***» stehen für die Suche zur Verfügung. Das IFS definiert – ggf. in Rücksprache mit der Community – den Begriff, der als Haupteintrag geführt wird (Diskussion pro Sprache). Für Exporte und die Bildschirmansicht wird jeweils der Inhalt des Haupteintrags angezeigt.

Abkürzung/Kürzel – Bezeichnungen

Im Deutschen ist «Kürzel» ein Sammelbegriff für Abkürzungsarten, für die im Englischen unterschiedliche Begriffe mit unterschiedlichen Bedeutungen verwendet werden. Im Englischen gibt es keinen analogen allgemeinen Überbegriff.

shorthand = benutzerdefiniertes Kürzel

inkl. intern verwendete Abkürzungen/Kürzel

Bsp. *bibliography* shorthand_SICF_german: «Divo – Tobler, 17. Jh.»Bsp. *bibliography* shorthand_authors_year: «Divo/Tobler 1987»**acronym** = Zusammenzug der AnfangsbuchstabenBsp. *modern_person* (Personenkürzel, Bsp. RCA, MNI ...)Bsp. *organisation* (Institutionsabkürzung, Bsp. HMO, BHM, HMB ...)**tag** = das wird in der Benutzeroberfläche als Kürzel angezeigtBsp. *language* (d/f/i/e), *material* (AV/AR/AL/NS ...)Bsp. *archaeological_event*, Grabungs-/Ereignis-/Objekt-/Vorgangs-Kürzel (V.008.2 [in KAAG]; 6.11 [Bsp. ABL]; 1978/23. [Bsp. ABBS]; 1978.023. [Bsp. Augusta Raurica, AATG]; 56.035. [Bsp. KASG]; KLB-RI2019-309 [SAEF]; 128.C. [Bsp. KALU] etc.)