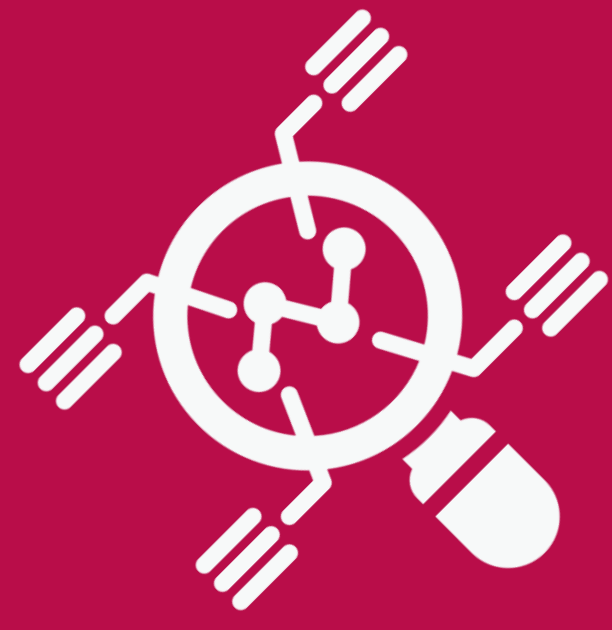


Das optimale Datenmodell: eine Spurensuche im Möglichkeitsfeld der Kodierung



Wie kann ein optimales Datenmodell entwickelt werden?

Während im Förderansuchen häufig die Nennung von XML/TEI-Kodierung für Textquellen ausreichend ist, stellt sich die praktische Arbeit meist komplizierter dar: Schon die TEI (TEI Consortium 2021) bietet zahlreiche Möglichkeiten, ähnliche Sachverhalte zu annotieren und die gewählte Strategie muss dabei auf das Material, die Forschungsfrage sowie die Archivierung und Weiterverwendung der Daten Rücksicht nehmen.

Herausforderungen bei der Datenmodellierung

- die untersuchten Merkmale möglichst genau und standardisiert in einem anerkannten Schema erfassen
- Referenzimplementationen und *best practice*-Guidelines berücksichtigen
- fachspezifische Vokabularien und Normdaten mit einbeziehen
- die Daten bereits im Entstehungsprozess mit Metadaten für die Weiterverwendung versehen
- FAIR-Dataprinzipien befolgen
- projektspezifische Merkmale und Legacy-Daten berücksichtigen
- alle Aspekte in der Projektlaufzeit erfüllen

Unser Lösungsansatz: ein Kompromiss

- Guidelines (TEI, CMIF, DTABf)
- *best practice*-Beispiele (Humboldt-Editionen)
- fachspezifische Vokabularien (Projektthesaurus)
- Normdaten (VIAF, GND, GeoNames)
- Metadaten für Weiterverwendung (ZIM-ACDH, marcrelator, dcterms)
- projektspezifische Merkmale
- Berücksichtigung der Legacy-Daten

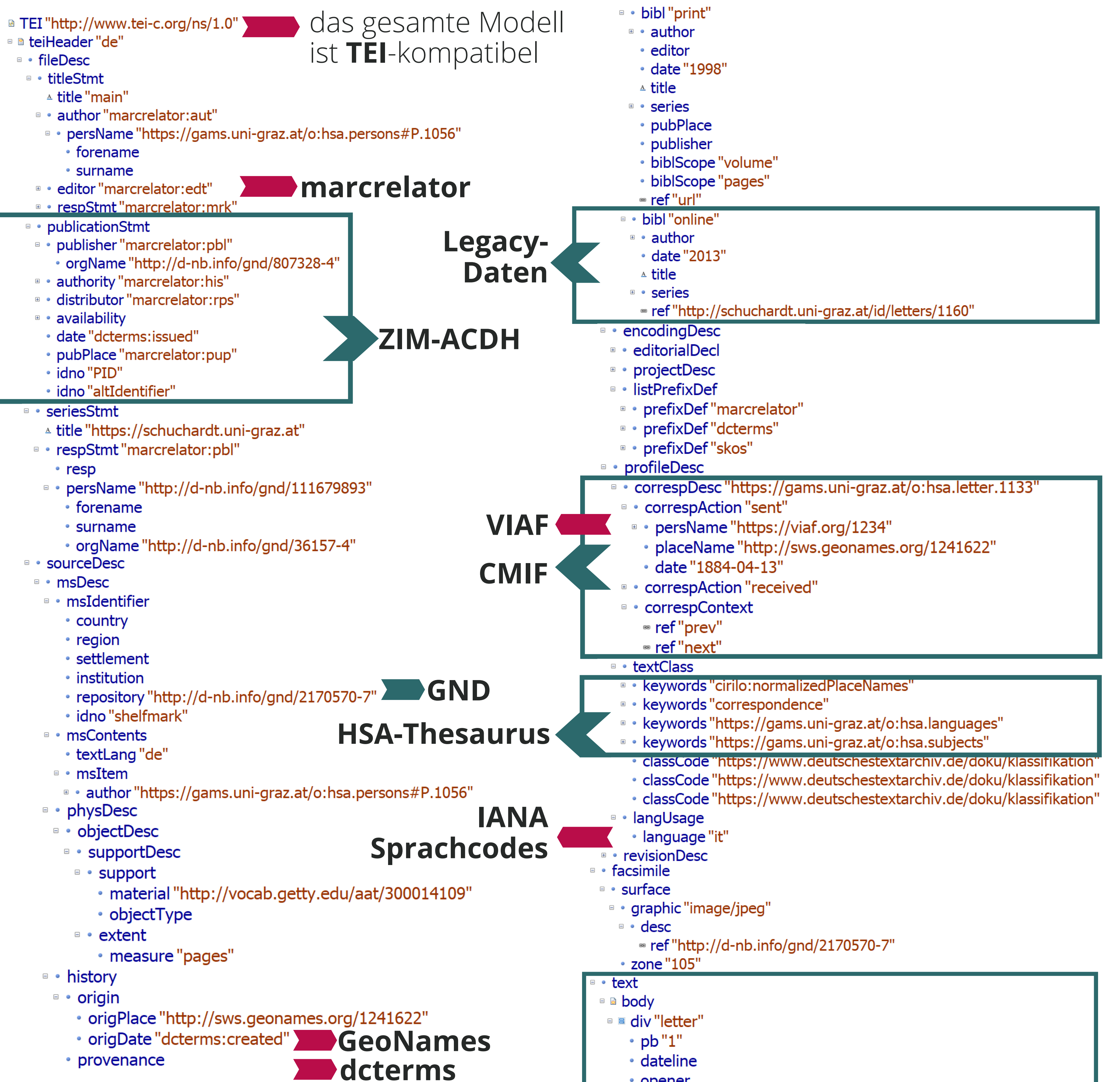
Beispiele

Hugo Schuchardt Archiv

Das „Hugo Schuchardt Archiv“ (hrsg. von Bernhard Hurch 2007-) ist ein langjähriges Vorhaben des Instituts für Sprachwissenschaft der Universität Graz. Im Mittelpunkt einer Kooperation mit dem Institut Zentrum für Informationsmodellierung – Austrian Centre for Digital Humanities steht die Migration aller Ressourcen vom Institut für Sprachwissenschaft in das Repositorium GAMS, um die Korrespondenz und andere Dokumente aus dem Nachlass zu archivieren. Derzeit läuft ein weiteres Projekt in Kooperation mit der Universität Zürich (FWF I 5076) zu Schuchardts Korrespondenz mit Gaston Paris.

Briefdatenmodell im Hugo Schuchardt Archiv

<?xml-model href="https://gams.uni-graz.at/o:hsa.odd/RNG" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?> ODD-Schema basierend auf dem DTABf



Anpassungen im ODD des DTABf waren notwendig, um zusätzliche Elemente, Attribute oder Attributswerte zu ermöglichen, wie etwa:

<idno>, <projectDesc>, <listPrefixDesc>, <zone>, <tree>, @ana in verschiedenen Elementen, @ref in <author>, @xml:lang in teilHeader, @corresp in @subtype in <div>, "print" und "online" in @type, etc.

DTABf unter Berücksichtigung der Legacy-Daten

Conclusio

Der Prozess der Datenmodellierung muss zahlreiche Einflussgrößen berücksichtigen. Er beinhaltet immer kritische Entscheidungen und Kompromisse, die sich aus dem Material, aber auch durch eingeschränkte Zeit- und Personalressourcen ergeben. Die Frage nach der bestmöglichen Kodierung kann daher nicht allgemeingültig beantwortet werden, sondern vielmehr individualisiert betrachtet. Trotzdem können konstituierende Eigenschaften für eine „gute“ Annotationspraxis beobachtet werden:

FAIR-Prinzipien befolgen und Weiterverwendung ermöglichen

Metadaten berücksichtigen, um Interoperabilität zu sichern

ODD-Schema nach Standards und *best practice*-Beispielen erstellen

ausreichend dokumentieren, um Nachvollziehbarkeit zu gewährleisten

Kontakt

Sanja Sarić sanja.saric@uni-graz.at^{1,2}
 Elisabeth Steiner elisabeth.steiner@uni-graz.at¹
 Maximilian Vogeltanz maximilian.vogeltanz@uni-graz.at²



Institut für Sprachwissenschaft

1 Institut Zentrum für Informationsmodellierung, Universität Graz
 2 Institut für Sprachwissenschaft, Universität Graz

Links

https://sprachwissenschaft.uni-graz.at
 https://informationsmodellierung.uni-graz.at
 https://gams.uni-graz.at
 http://schuchardt.uni-graz.at

