

Nomisma.org: Numismatik und das Semantic Web

Wigg-Wolf, David

david.wigg-wolf@dainst.de
Römisch-Germanische Kommission des Deutschen
Archäologischen Instituts

Tolle, Karsten

tolle@dbis.cs.uni-frankfurt.de
Johann Wolfgang Goethe-Universität Frankfurt am Main

Kissinger, Timo

Timo.Kissinger@adwmainz.de
Akademie der Wissenschaften und der Literatur | Mainz,
Deutschland

Einführung

In diesem Beitrag wird eine domänenspezifische Anwendung von Linked Open Data und Ontologien vorgestellt, die als Paradigma für den Umgang mit digitalen Corpora in der Archäologie dienen kann. Zunächst wird die Entwicklung des Projektes Nomisma.org sowie dessen Anwendung für die Verlinkung von digitalen Datenbeständen erläutert. In einem zweiten Teil wird ein Pilotprojekt vorgestellt, das darauf abzielt, textbasierte Münzpublikationen als RDF zur Verfügung zu stellen und somit prüft, inwieweit das Vokabular und die Ontologie von Nomisma.org eingesetzt werden können.

Das 2010 von der American Numismatic Society, New York, initiierte Projekt Nomisma.org¹ definiert und stellt stabile digitale Repräsentationen numismatischer Konzepte nach den Prinzipien von Linked Open Data zur Verfügung. Sie werden als http-URIs veröffentlicht², die Zugang zu weiterverwertbaren Informationen zu den Konzepten liefern, dazu noch Links zu weiteren Linked Open Data-Ressourcen (Getty Vocabularies, wikidata, viaf, GND, u.v.m.). Ferner wurde eine numismatische Ontologie³ entwickelt, die den Bedürfnissen und Arbeitsweisen der Numismatik gerecht ist und eine unkomplizierte Modellierung ermöglicht. Das kanonische Format von Nomisma.org ist RDF/XML, aber auch weitere Formate wie JSON-LD (für Geodaten geoJSON-LD), Turtle, KML und HTML5+RDFa 1.1. werden bedient.

Virtuelle Sammlungen

Zunächst lag der Schwerpunkt der Arbeit von Nomisma.org bei der römischen Numismatik und bis 2016 erfolgte die Onlinestellung der Linked Open Data-

Ressourcen *Coinage of the Roman Republic Online* (CRRO)⁴ und *Online Coins of the Roman Empire* (OCRE)⁵ als virtuelle Münzsammlungen für die Prägungen der römischen Republik und der Kaiserzeit. Die Abfrage erfolgt über RDF Dumps der Projektpartner, die zentral gehalten werden; die dazugehörigen Bilder werden on-the-fly von den Servern der Partner geholt, wodurch nicht nur das Volumen der zentralen Datenhaltung gering gehalten und die Abfragegeschwindigkeit erhöht, sondern auch eventuelle rechtliche Probleme bei der Lizenzierung der Bilder vermieden werden.

Mittlerweile befassen sich erste Projekte auch mit dem disparaten und weitaus komplexeren Stoff der griechischen Welt,⁶ z.B. *PELLA* für die Prägungen der makedonischen Dynastie der Argeaden, *Corpus Nummorum Thracorum* für Thrakien, oder *Seleucid Coins Online* für die Prägungen der Seleukiden. Des Weiteren haben Arbeitsgruppen bereits angefangen, die notwendigen Konzepte für die keltische und islamische Numismatik sowie für das Mittelalter zu definieren.

Fast 200.000 antike Münzen von insgesamt 39 Institutionen werden heute in auf Nomisma.org basierenden Online-Ressourcen veröffentlicht.⁷ Aktiv an der Entwicklungen beteiligt sind u.a. das Institute for Studies of the Ancient World, New York, das Deutsche Archäologische Institut sowie vier der weltweit bedeutendsten Münzsammlungen: Die American Numismatic Society, das British Museum, die Bibliothèque nationale de France und das Münzkabinett der Staatlichen Museen zu Berlin. Auch eine Reihe kleinerer Sammlungen sind dabei ihre Bestände in Portalen wie OCRE und CRRO online zu stellen. Darunter schon acht der 34 deutschen universitären Sammlungen, die Mitglieder im NUMiD-Verbund (*Netzwerk universitärer Münzsammlungen in Deutschland*)⁸ sind.

Die verlinkte Antike

Nomisma.org und die darauf bauenden Projekte sind in die Linked Open Data-Welt der Antike fest integriert. Mit Nomisma.org verlinkte, bzw. gemappte Münzen werden bei Ressourcen wie *Pleiades*⁹ und *Pelagios*¹⁰ angezeigt und umgekehrt können Einträge in den virtuellen Sammlungen mit kontextvermittelnden Daten aus anderen Quellen angereichert werden. Beispielsweise: Wird Literatur in der Form von iDAI.bibliographie / ZENON¹¹ URIs zitiert, können umgekehrt in Zenon Verweise auf Münzen beim entsprechenden Titel angezeigt werden. Im Abschlussbericht des Work Package 15 des EU-FP7 Projektes ARIADNE wurde Nomisma.org mehrfach als führendes Beispiel für die fachspezifische Anwendung von Linked Open Data zitiert (Geser 2016).

Nomisma.org hat numismatische Daten auf eine Weise digital zur Verfügung gestellt, die vor 10 Jahren kaum vorstellbar gewesen wäre und macht einen bedeutenden Bestand an Material der Forschung leichter zugänglich.

Durch die Verlinkung mit anderen LOD-Ressourcen sind diese Daten auch deutlich sichtbarer und stehen damit für Forschungsvorhaben in anderen Disziplinen vermehrt zur Verfügung. So kann Nomisma.org beispielsweise einen wichtigen Beitrag zur Aufhebung der immer noch weit verbreiteten Isolation der Numismatik in den Altertumswissenschaften leisten, eine Isolation, die in der (bisher oft fehlenden) Zusammenarbeit mit der Archäologie besonders deutlich erkennbar ist.

Fundmünzen und Archäologie

Konzentrierten sich Projekte wie OCRE und PELLA zunächst auf die Präsentation von Münzen aus öffentlichen Sammlungen, richtet sich der Blick mittlerweile zunehmend auf Münzen im archäologischen Kontext. Nomisma.org kompatible Standards für die Aufnahme und Veröffentlichung von Münzen aus Ausgrabungen wurden bei einem Treffen an der University of Oxford im September 2018 vereinbart, um Re-use und Interoperabilität zu ermöglichen.¹²

Nationale Projekte wie *NUMIS*¹³ für die Niederlande oder das *Inventar der Fundmünzen der Schweiz*¹⁴ veröffentlichen umfangreiche Bestände an Fundmünzen im Web, die aber nur über das jeweilige nationale Portal zugänglich sind. Eine länderübergreifende Abfrage ist nicht möglich. Eine Ausnahme bietet das Projekt *Antike Fundmünzen in Europa*¹⁵, das die Datenbanken der Römisch-Germanischen Kommission in Frankfurt und des Unternehmens *Finds of Roman Coins in Poland* auf das Vokabular und die Ontologie von Nomisma.org mappt und mittels eines D2R-Servers als RDF online stellt. Ein SPARQL-Endpoint ermöglicht die gemeinsame Abfrage der beiden Datenbestände und erleichtert damit die Analyse der Fundmünzen aus einem von der Nordsee bis zur Ukraine reichenden Raum.

Text zu RDF: Erster Versuch einer Digitalisierung

Bisher lag der Schwerpunkt der Verlinkung mit bzw. Mapping auf Nomisma.org auf in Datenbanken gehaltene Daten. In diesem Beitrag soll abschließend ein Pilotprojekt vorgestellt werden, das sich mit den Möglichkeiten der Anwendung von Nomisma.org auf textbasierte Datenbestände beschäftigt.

Grundlage

Das Projekt „Fundmünzen der römischen Zeit in Deutschland“ (FMRD) brachte 48 Bände mit weit über 300.000 erfassten Fundmünzen heraus. Im Rahmen eines Praxisprojektes im Studiengang „Digitale Methodik in den Geistes- und Kulturwissenschaften“¹⁶ wurde

als Beispieldatensatz aus FMRD der Fundkomplex Domgrabung/Liebfrauen-Areal in Trier mit 1.157 Münzen digitalisiert (M. Radnoti-Alföldi 2006: 119–206).

Das Verfahren

Die Daten zu den Münzen wurden aus dem PDF in CSV und XML extrahiert und in ein RDF-Dokument umgewandelt, um die Vernetzung und Visualisierung zu ermöglichen (Abb. 1).

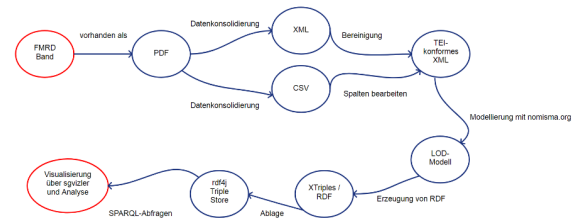


Abbildung 1. Verlaufsplan.

Für die Konvertierung aus dem PDF und die Bearbeitung wurde *Adobe Acrobat Pro DC*¹⁷, der *Oxygen*¹⁸ Editor und *Libre Office*¹⁹ verwendet. Die extrahierten Daten wurden anhand der CSV-Extraktion in ein einheitliches Spaltensystem gebracht. Über das XML-Dokument wurden anhand von regulären Ausdrücken Konvertierungsfehler etc. ausgebessert. Am Ende stand eine einzelne TEI-konforme Tabelle. Der nächste Schritt bestand darin, ein Linked-Open-Data-Modell anhand der Ontologie von Nomisma.org (Abb. 2) zu entwickeln.

```
<rdf:Description rdf:about="https://lod.academy/fmrd/id/*5,">
  <nmo:hasAuthority>Galba</nmo:hasAuthority>
  <nmo:hasDenomination>S</nmo:hasDenomination>
  <nmo:hasDate>68/69</nmo:hasDate>
  <dcterms:date>1. Jh. n. Chr.</dcterms:date>
  <nmo:hasMint>?</nmo:hasMint>
  <nmo:hasReferenceWork>RIC ?</nmo:hasReferenceWork>
  <nmo:hasPeculiarity>NaN-Wert</nmo:hasPeculiarity>
  <dcterms:description>Domgrabung A 5. Vs.: Kopf n.r. m. Lbkr. Fast völlig
    abgegriffen.</dcterms:description>
  <dcterms:bibliographicCitation>339,1</dcterms:bibliographicCitation>
</rdf:Description>
```

Abbildung 2. RDF-Modell.

Um dieses Modell automatisiert auf alle 1.157 Münzen der Trierer Domgrabung übertragen zu können, wurde der Webdienst *XTriples* verwendet²⁰, der es erlaubt aus XML ein RDF-Dokument zu erzeugen.

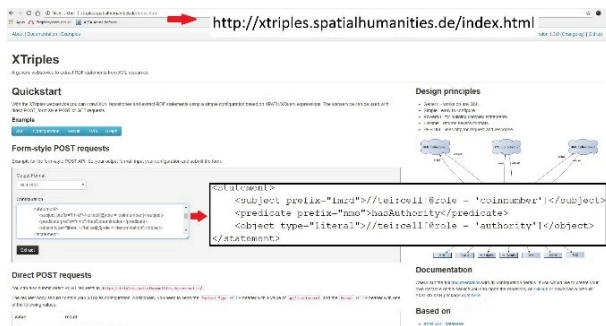


Abbildung 3. Webservice XTriples.

Das RDF-Dokument wurde auf einen Triple Store abgelegt, der über ein RDF4J-Framework verfügt und SPARQL-Abfragen erlaubt. Ein weiterer Webdienst, *sgvizler*,²¹ ermöglicht es im Anschluss die Daten zu visualisieren und zu analysieren (Abb. 4).

als CSV ausgegeben kann. Im Verlauf der Erprobung stellte sich *Tabula* als ungeeignet heraus, da immer noch zu viele manuelle Eingriffe vonnöten sind. So erkennt es manche Tabellenteile nicht (Abb. 5) und die Spalteninhalte im erzeugten CSV werden nicht richtig zugeordnet.

		EINZELFÜNDE	3003.1
1.-	Unvollständig		
2.	D? 1.-3. Jh?		09,1621
3.	Barbarisierung: Nr. 1.-4. Gall?		09,1137
4.	ME " " ?		09,1149

3. Jahrhundert n. Chr.

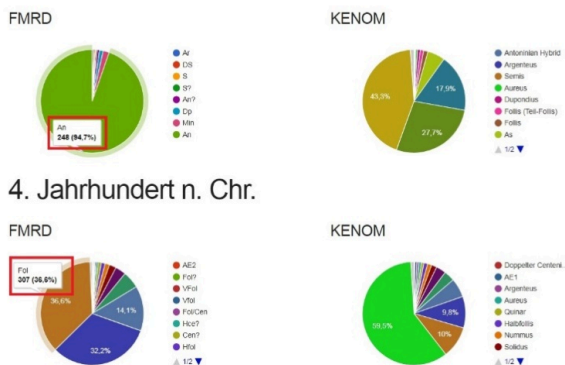


Abbildung 4. Visualisierung über sgvizler zu den Fundmünzen der Trierer Domgrabung.

Ausblick

Das oben beschriebene Verfahren wurde dann erweitert, um eine Pipeline aufzubauen, mit der alle Bände der FMRD-Reihe digitalisiert werden können. Auf manuelle Schritte und proprietäre Software soll dabei möglichst verzichtet werden.

Tabellenextraktion

In einem ersten Schritt wurden aktuelle Open-Source-Angebote verglichen, die es ermöglichen, aus der PDF-Struktur Tabellen in CSV- bzw. XML-Dateien zu extrahieren. Die Daten liegen jedoch so inhomogen vor, dass bisher erschienene Programme und Skripte/Bibliotheken schnell an ihre Grenzen stoßen. Als Beispiel soll hier das browserbasierte Tool *Tabula*²² dienen, welches Tabellen im PDF automatisch erkennt und anschließend



Abbildung 5. Tabula: Fundmünzen Nummer 1 bis 4 werden nicht erkannt.

pdfhtml

Eine andere Herangehensweise erzeugt als Zwischenschritt XML, bevor dieses in CSV umgewandelt werden kann. Für die Erzeugung des XML wird das Tool *pdf2html* verwendet.²³ Das Ergebnis lässt sich über den *pdf2xml-viewer*²⁴ überprüfen (Abb. 6).

Abbildung 6. OCR-Ergebnis.

Die Normalisierung der Spalten ist hier leichter, da auch leere Tabellenspalten erkannt werden. Jedoch wird auch hier der Spalteninhalt nicht überall korrekt erkannt (Münze 18). Die XML-Struktur erlaubt es über die Tags zu navigieren und diese mit X-Technologien zu manipulieren. Doch ist der Anteil der nicht korrekt erkannten Spalteninhalte immer noch zu hoch, um die Daten wie im Druckband abzubilden.

Reguläre Ausdrücke

Eine andere Möglichkeit an die Tabellendaten zu gelangen sind reguläre Ausdrücke. Mithilfe eines Pythonskriptes wurden diese separiert und als CSV ausgegeben. Dafür wurden beispielhaft verschiedene Münzkomplexe herangezogen und die regulären Ausdrücke um die pro Komplex neu auftretenden Sonderfälle modifiziert. Dieses Verfahren erzeugte gute Ergebnisse für die behandelten Komplexe. So waren nur noch kleinere manuelle Nachbesserungen nötig. Im Vergleich erzeugte dieses Verfahren das beste Ergebnis.

Fazit

Die Digitalisierung von „Fundmünzen der römischen Zeit in Deutschland“ erweist sich als anspruchsvoll. Der Ansatz, die Daten über reguläre Ausdrücke zu gewinnen und

zu extrahieren, erscheint bisher am vielversprechendsten. Die dabei auftretenden Abweichungen zwischen den Münzfundkomplexen lassen sich durch auf die FMRD-Struktur zugeschnittene Ausdrücke ausbessern. So erhoffen wir uns, die bisher über Nomisma.org veröffentlichten Datenbestände durch wichtige archäologische Kontexte zu erweitern und zu bereichern.

Fußnoten

1. <http://nomisma.org/>; Hinweis: Alle im Beitrag erwähnten URLs wurden zuletzt am 14.10.2018 überprüft.
2. z.B. <http://nomisma.org/id/sestertius>
3. <http://nomisma.org/ontology>
4. <http://numismatics.org/crro/>
5. <http://numismatics.org/ocre/>
6. <https://www.greekcoinage.org/portals.html>
7. <http://nomisma.org/datasets>
8. <http://www.numid-verbund.de/>
9. <https://pleiades.stoa.org/>
10. <http://peripleo.pelagios.org/>
11. <http://peripleo.pelagios.org/>
12. <https://www.greekcoinage.org/coins-in-context.html>
13. <https://www.dnb.nl/en/about-dnb/nationale-numismatische-collectie/numis/numis-database/index.jsp>
14. <https://www.fundmuenzen.ch/>
15. <http://afe.fundmuenzen.eu/>
16. <https://www.digitale-methodik.uni-mainz.de/>
17. <https://acrobat.adobe.com/de/de/acrobat/acrobat-pro.html>
18. <https://www.oxygenxml.com/>
19. <https://de.libreoffice.org/>
20. <http://xtriples.spatialhumanities.de/index.html>
21. <http://mgskjaeveland.github.io/svgvizler/>
22. <https://tabula.technology/>
23. <https://poppler.freedesktop.org/>
24. <https://github.com/WZBSocialScienceCenter/pdf2xml-viewer>

Bibliographie

G. Geser (2016): *ARIADNE WP15 Study: Towards a Web of Archaeological Linked Open Data* (www.ariadne-infrastructure.eu/.../ARIADNE_archaeological_LOD_study_10-2016.pdf, 31.10.2016).

E. Gruber: *Numishare Blogspot*: <http://numishare.blogspot.com/>.

E. Gruber / G. Bransbourg / S. Heath / A. Meadows (2014): *Linking Roman Coins: Current Work at the American Numismatic Society*, in: **G. Earle et al. (eds):** *Archaeology in the Digital Era: Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*, Southampton, 26-29 March 2012 (Amsterdam) 249–258.

E. Gruber / S. Heath / A. Meadows / D. Pett / D. Wigg-Wolf (2014): *Semantic Web Technologies Applied to Numismatic Collections*, in: **G. Earle et al. (eds):** *Archaeology in the Digital Era: Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*, Southampton, 26-29 March 2012 (Amsterdam) 264–274.

Radnoti-Alföldi, M. (2006): *Die Fundmünzen der römischen Zeit in Deutschland IV 3/2. Stadt und Reg.-Bez. Trier. Die Sog. Römerbauten (Mainz)*.

K. Tolle / D. Wigg-Wolf (2015): *Uncertainty Handling for Ancient Coinage*, in: **F. Giligny et al. (eds):** *CAA2014. 21st Century Archaeology. Concepts, Methods and Tools*. Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology (Oxford) 171–178.

K. Tolle / D. Wigg-Wolf (2016): *How to Move from Relational to 5 Star Linked Open Data – A Numismatic Example*, in: **S. Campana et al. (eds):** *CAA2015. Keep the Revolution Going*. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology (Oxford 2016) 275–281.

D. Wigg-Wolf / F. Duyrat (2017): *La révolution des Linked Open Data en numismatique: Les exemples de nomisma.org et Online Greek Coinage*. *Archéologies numériques* 1.1, 2017.