

re3dragon – A Research Registry Resource API for Data Dragons

Florian Thiery ^{1,*} and Allard W. Mees ¹

¹Leibniz-Zentrum für Archäologie, Mainz, Germany

*Correspondence: florian.thiery@leiza.de

Presented at ARCHEO.FOSS XV 2021: Open software, hardware, processes, data, and format in archaeological research; online; November 23rd-26th 2021.

Abstract: The re3dragon (REsearch REsource REgistry for DataDragons) FLOSS tool envisages two aims: first, publication of an archaeology related open extendable LOD resource catalogue including authority-data, thesauri, gazetteers, (space-) time gazetteers as well as typologies and domain-specific resources. 2nd, re3dragon offers an API for requesting distributed LOD resources, returning resources in a standardised JSON format based on JSKOS. The re3dragon is coded in JAVA and Open Source published on GitHub.

Keywords: Linked Open Data, SKOS, JSKOS, JAVA, Thesauri Catalogue

FOSS software used and license:

1 re3dragon, MIT License;

2 Netbeans, Apache 2.0 License;

3 Apache Maven, Apache-License, Version 2.0;

4 Apache Jena, Apache-License, Version 2.0;

5 Eclipse Jersey, Eclipse Public License 2.0

Repository and license:

re3dragon on GitHub: <https://github.com/leiza-rse/re3dragon>

JSKOS+ on GitHub: <https://github.com/leiza-rse/jskos-plus>

Introduction

Cartographers in historical maps used the phrase “*Hic sunt dracones*” (historically translated as: here be dragons) to describe areas which were beyond the known world of the map creator. Today, the digital data universe is full of unknown data, which have to be made FAIR (Findable, Accessible, Interoperable and Reusable) to integrate them in archaeological research. The World Wide Web offers researchers the possibility to share research data and enables the community to participate in the scientific discourse in order to generate previously unknown knowledge. However, much of this data is neither directly comparable and alignable nor easily findable or accessible, thus resulting in so-called modern unknown Data Dragons (Thiery et al. 2019). These Data Dragons lack connections to other data sets, which leads to a lack of interoperability and, in some cases, unusability. To overcome these shortcomings, a set of techniques, standards and recommendations can be used: Semantic Web and Linked Open Data (LOD) (Berners-Lee 2006; Schmidt et al. 2022) and consequently Linked Open Usable Data¹ (LOUD). To tame and unveil the modern data dragons, the CAA Special Interest Group (SIG) on Semantics and LOUD

¹ <https://linked.art/loud/> (accessed 25/02/2022)

in Archaeology (SIG-DataDragon²) was established in 2020. Data dragons require a safe location, what we call Dragon Lair, where they can be clustered and made accessible. This data dragon LOD lair location and its machine-readable accessibility are combined in the re3dragon tool - the REsearch REsource REgistry for DataDragons. The FLOSS³ re3dragon tool envisages two aims. First, publication of an open extendable archaeology related LOD resource catalogue (the Dragon Lair) including authority data (e.g. Integrated Authority File - GND), thesauri (e.g. Getty AAT, Heritage Data Vocabularies), controlled vocabularies, gazetteers (e.g. GeoNames, Pleiades), space-time gazetteers (e.g. ChronOntology, PeriodO), as well as typologies and domain specific resources (e.g. Roman Open Data, Nomisma, Linked Open Samian Ware). 2nd, offering an API for requesting distributed LOD resources (so-called Dragon Items) providing resources in a standardised JSON format based on the JavaScript Object Notation for Simple Knowledge Organization Systems: JSKOS (Voß et al. 2016; Voß 2021). The tool is currently being implemented as part of the LEIZA digital research data infrastructure. It plays an essential role in the design of an overarching keyword registry (Meta-Index), which enables aggregating external distributed data resources in order to qualify internal data (e.g. the Meta-Index term Mainz may refer to Roman, mediaeval⁴ or modern Mainz⁵, resulting in different aggregation references). This process to transform a vocabulary term into a label by linking it to reference thesauri concepts has been described in Piotrowski et al. (2014). This approach enforces interoperability and reusability, enabling mapping on external major union systems like Europeana. Moreover, these aspects play a pivotal role in the planned German National Research Data Infrastructure (NFDI) consortium NFDI4Objects⁶. Based on creating a JSKOS-enhancement (JSKOS+), re3dragon supports the furtherance of interdisciplinary major existing knowledge hubs such as the Basic Register of Thesauri, Ontologies & Classifications (BARTOC)⁷ registry and the accompanying SKOS based mapping tool Cocoda⁸.

Implementation

The re3dragon API is coded in JAVA using Maven and Apache Jena and is published Open Source on GitHub⁹ (Thiery 2021). The re3dragon is based on the Labeling System approach (Piotrowski et al. 2014; Thiery and Engel 2016; Thiery and Mees 2023) and the re3cat API¹⁰.

The basis of re3dragon is an online catalogue of Linked Open Data resources (the Dragon Lair), which includes norm- and authority data resources (e.g. Getty AAT, Wikidata) as well as domain specific typologies and archaeological data (e.g. Terra Sigillata production centres, distribution sites, potters and ceramic typologies), stored as RDF in a RDF4J triplestore. Dragon Lairs are semantically modelled using the so called Linked Archaeological Data Ontology (LADO) which is based on the Research Software Engineering (RSE) Tools Ontology (Thiery 2019). Figure 1 demonstrates exemplary the Getty AAT Dragon Lair using the prefixes lado¹¹ and wd¹²:

² <https://caa-international.org/special-interest-groups>; <http://datadragon.link> (accessed 25/02/2022)

³ <https://www.gnu.org/philosophy/floss-and-foss.en.html> (accessed 25/02/2022)

⁴ <https://pleiades.stoa.org/places/109169>; <http://imperium.ahlfeldt.se/places/3> (accessed 25/02/2022)

⁵ <http://sws.geonames.org/2874225>; <https://www.wikidata.org/entity/Q1720> (accessed 25/02/2022)

⁶ <https://osf.io/4t29e/wiki/home/> (accessed 25/02/2022)

⁷ <https://bartoc.org/> (accessed 25/02/2022)

⁸ <https://coli-conc.gbv.de/> (accessed 25/02/2022)

⁹ <https://github.com/RGZM/re3dragon> (accessed 25/02/2022)

¹⁰ <https://github.com/mainzed/re3cat> (accessed 25/02/2022)

¹¹ lado prefix: <http://archaeology.link/ontology#>

¹² wd prefix: <http://www.wikidata.org/entity/>

property	value
rdf:type	lado:DataDragonLair
owl:sameAs	wd:Q611299
rdfs:label	Getty AAT
lado:description	controlled vocabulary used for describing items of art, architecture, and material culture
lado:author	Getty Research Institute
lado:sparqlendpoint	http://vocab.getty.edu/sparql
lado:hasLegalType	lado:ResearchInstitution
lado:language	lado:en

Figure 1. Dragon Lair example as RDF (Getty AAT). [Florian Thiery, CC BY 4.0]

The re3dragon offers three types of API services. 1st, a search API service for LOD resources with string and distance similarity; 2nd, an API resolving service for LOD resources related to specific URIs. 3rd, a catalogue API for Dragon Lairs. Service types 1 and 2 offer results according to the JSKOS+ format and HTML, and service type 3 provides an interoperable output integrable in research applications.

/re3dragon/rest/search?repo=gettyaat&q=lion
<pre> [{ "displayLabel": { "en": "Panthera leo (species)" }, "type": ["http://www.w3.org/2004/02/skos/core#Concept"], "lair": { "scheme": "Getty AAT", "legaltype": "lado:ResearchInstitution", "id": "ULBU3XXM", "type": "Thesaurus", "wikidata": "wd:Q611299", "quality": "lado:qualityHigh", "group": "lado:chreferencethesauriGroup" }, "uri": "http://vocab.getty.edu/aat/300310388", "displayDesc": { "en": "Large, powerful species of cat that is well-muscled, with a large head, short legs, size and appearance that varies considerably between the sexes, and is unique among the cats in living in family groups or prides. [...]" }, "similarity": { "levenshtein": 20.0, "dameraulevenshtein": 20.0, "jarowinkler": 0.52, "normalizedlevenshtein": 0.91, "searchstring": "lion", "lairstring": "Panthera leo (species)" } }] </pre>

Figure 2. JSKOS+ output for a Getty AAT search string "lion". [Florian Thiery, CC BY 4.0]

The JSKOS format defines a JavaScript Object Notation (JSON) structure to encode Knowledge Organisation Systems (KOS) (Voß 2021). JSKOS+ provides a standardised (Geo-)JSON(-LD) data model according to JSKOS which can be used in response to an API request. Figure 2 demonstrates the JSKOS+ JSON results for Getty AAT requests, showing the original JSKOS items in red and the added JSKOS+ items in blue.

The re3dragon applied in a research application

The re3dragon tool has been applied in the ARS3D project¹³ (Thiery and Rokohl 2021; Thiery et al. 2023), in which it enriched data related to late Roman African Red Slip Ware (ARS). As an example, an ARS vessel is generically described as a “bowl” with decorative motives of “Hercules” and “Victoria” (O.39446¹⁴), which can be annotated using Getty AAT, IconClass and Wikidata items. The potform can be described as a bowl using Getty AAT item 300203596 and Wikidata Item Q15398¹⁵, cf. fig. 3. The two features can be described as follows: Hercules: IconClass item 94L¹⁶ (cf. fig. 4) and Wikidata item Q240679 according to (Zu Löwenstein 2015) “B / FT III” (Q110892402), (Armstrong 1993) “8.109” (Q110892540), and (Atlante 1981) “135” (Q110892520); Victoria: IconClass item 96A5 (VICTORIA) and Wikidata item Q308902 according to (Zu Löwenstein 2015) “N / FT III (Victoria)” (Q110892417) and Armstrong (1993) “8.100” (Q110892537).



Figure 3. left: JSON+ Output for Wikidata item “Q15398” (bowl); right: scanned bowl O.39446 with Hercules and Victoria. [left: Florian Thiery, CC BY 4.0; right: ARS3D Project / RGZM / i3mainz, CC BY-SA 4.0]

¹³ <https://ars3d.rgz.de> (accessed 25/02/2022)
¹⁴ <https://ars3d.rgz.de/object.htm?id=ars3do:eab38a5a-aaa2-41a1-b17b-0b91cbab006c> (accessed 25/02/2022)
¹⁵ [host]/re3dragon/rest/item?uri=http://www.wikidata.org/entity/Q153988
¹⁶ [host]/re3dragon/rest/item?uri=http://iconclass.org/94L

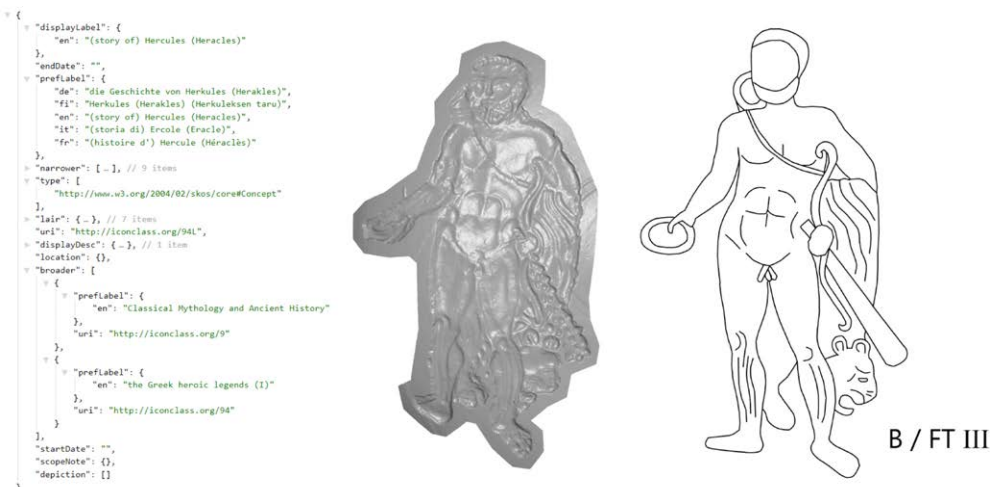


Figure 4. left: JSON+ Output for IconClass item “94L” (Hercules); middle: extracted Hercules feature from bowl O.39446; right: Figure B / FT III according to S. von Löwenstein [left: Florian Thiery, CC BY 4.0; middle ARS3D Project / RGZM / i3mainz, CC BY-SA 4.0; right: Sophie zu Löwenstein (2015, p.459)]

Outlook

During its usage in archaeological data enrichment projects, the re3dragon tool can be easily and demand driven enhanced by adding more Dragon Lairs. The tool is already embedded in the ongoing development of the so-called Meta-Index of the Römisch-Germanisches Zentralmuseum Mainz (Thiery et al. 2018; Thiery and Mees 2021).

This procedure enables future enhancements of the JSKOS standard, e.g. semantic location descriptions, mapping properties as well as space and time typologies (Dimensionally Extended 9-Intersection Model, Allen’s Interval Algebra, etc.).

References

- Armstrong, M.A., ed., 1993. *A Thesaurus of Applied Motives on African Red Slip Ware*. New York University: ProQuest.
- Berners-Lee, T. 2006. “Linked Data - Design Issues”. Last modified 18/06/2009. <https://www.w3.org/DesignIssues/LinkedData.html>.
- Istituto della Enciclopedia Italiana. 1981. *Atlante delle forme ceramiche: Ceramica fina romana nel bacino mediterraneo*. Roma: Istituto della Enciclopedia Italiana.
- Piotrowski, M., Colavizza G., Thiery F. and Bruhn K-C. 2014. “The Labeling System: A New Approach to Overcome the Vocabulary Bottleneck”. In *DH-CASE II: Collaborative Annotations on Shared Environments: Metadata, Tools and Techniques in the Digital Humanities - DH-CASE '14, (1-6 September 2014)*. Fort Collins, CA, USA: ACM Press. <https://doi.org/10.1145/2657480.2657482>.
- Thiery, F. 2019. “The RSE Tools Ontology”. Zenodo. <https://doi.org/10.5281/zenodo.3541559>.
- Thiery, F. 2021. *Re3dragon - REsearch REsource REgistry for DataDragons (version v0.1)*. Zenodo. <https://doi.org/10.5281/zenodo.5338740>

- Thiery, F. and Engel T. 2016. “The Labeling System: The Labelling System: A Bottom-up Approach for Enriched Vocabularies in the Humanities”. In *CAA2015. Keep the Revolution Going. Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology.*, edited by Campana S., Scopigno R., Carpentiero G. , and Cirillo M., 259–68. Oxford: Archaeopress.
- Thiery, F. and Mees A. 2021. “RGZM Meta-Index and Community-Driven Vocabularies in the LOD Cloud: Examples from the Department of Scientific IT”. <https://doi.org/10.5281/zenodo.5764570>.
- Thiery, Florian and Allard Mees. 2023. “Taming Ambiguity - Dealing with doubts in archaeological datasets using LOD” in *Human History and Digital Future. Proceedings of the 46th Annual Conference on Computer Applications and Quantitative Methods in Archaeology*. <http://dx.doi.org/10.15496/publikation-87762>
- Thiery, F., Mees A. and Heinz G. 2018. “RGZM-Meta-Index: A Central Linked Data Hub for Aligning Distributed Databases”. <https://doi.org/10.5281/zenodo.2222237>.
- Thiery, F. and Rokohl L. 2021. “African Red Slip Ware Digital (ARS3D) - The Portal”. <https://doi.org/10.5281/zenodo.5646897>.
- Thiery, F., M. Trognitz, E. Gruber, and D. Wigg-Wolf. 2019. “Hic Sunt Dracones! The Modern Unknown Data Dragons”. <https://doi.org/10.5281/zenodo.3345711>.
- Thiery, F., J.Veller, L. Raddatz, L. Rokohl, F. Boochs and A. W. Mees. 2023. “A Semi-Automatic Semantic-Model-Based Comparison Workflow for Archaeological Features on Roman Ceramics” *ISPRS International Journal of Geo-Information* 12, no. 4: 167. <https://doi.org/10.3390/ijgi12040167>
- Schmidt, S. C., F. Thiery and M. Trognitz. 2022. “Practices of Linked Open Data in Archaeology and Their Realisation in Wikidata” *Digital* 2, no. 3: 333-364. <https://doi.org/10.3390/digital2030019>
- Voß, J. 2021. “JSKOS Data Format for Knowledge Organization Systems”. Github. <https://gbv.github.io/jskos/jskos.html>.
- Voß, J., Ledl A. and Balakrishnan U. 2016. “Uniform Description And Access To Knowledge Organization Systems With Bartoc And Jskos’”. <https://doi.org/10.5281/zenodo.438019>.
- Zu Löwenstein, S. 2015. *Mythologische Darstellungen Auf Gebrauchsgegenständen Der Spätantike: Die Appliken- Und Reliefverzierte Sigillata C3/C4*. *Kölner Jahrbuch* 48.