The work outlined in this paper (based on a Bachelor's thesis) is aimed at implementing multiple archaeological KOS in the Simple Knowledge Organization System[1] and create automated alignments between them. The LOD implementation and alignment permit the interoperability and more prevalent use of archaeological KOS. This feasibility study allows statements regarding the automated extension of knowledge organization at the German Archaeological Institute(DAI)[2].

Since the institute's founding in the 19th century, large thesauri have been developed, which have become the most important terminologies in this area and are still comprehensively maintained. The "Roman" thesaurus is in first place, which subsequently has been extended through the data of the Roman Germanic Commission in Frankfurt and other DAI departments and projects. Traditionally dominated by the research focused on Greece and Italy, these central Information Retrieval instruments cannot meet the now much more diversified scientific archaeological activities any more. The development of parallel, topographically and chronologically different terminologies has therefore been a necessary complement to the traditional thesauri to represent the entire research at the DAI. Especially the iDAI.gazetteer[3] gained importance with well over 100,000 place entries, but also smaller project-related and thus very specifically oriented terminologies such as the vocabulary of the East Asian Archaeology of the Chinese Archaeology Database[4] have gained importance.

With this parallel existence of KOS within a single scientific area an improvement of data interoperability is needed, eventually achieving exhaustive retrieval results through a unified data space. In this work the automated implementation of this interoperability with intellectually created KOS is investigated. A brief overview of the procedure is presented, first converting the data into SKOS and subsequently processing an alignment with the tool amalgame[5]. Finally, it shows the approach for an evaluation of the alignment results.

For conversion to SKOS, several projects can be named as guidelines, especially the documentation of the implementation of "Roman" thesaurus already carried out in the DARIAH-DE project[6] at the DAI [1], the conversion of the Thesaurus for the Social Sciences[7] (TheSOZ) at GESIS [4] [8] and the STW Thesaurus for Economics[8] at the ZBW [5].
Numerous studies serve as the basis for alignment strategies, in particular the basic work described in [2], furthermoreS the work on the TheSOZ and the STW Thesaurus for Economics [3], while on an international level [7] as well as the activities of the Ontology Alignment Evaluation Initiative[9] in general have been used for orientation purposes.

For conversion to SKOS, the Python script, resulting from the preparatory activities [1], was adapted and extended. This was necessary particularly due to the addition of the iDAI.gazetteer - and Charda data, the former in JSON format, the latter stored in a simple Excel table. Vocabulary specifics had to be considered, for example the partially duplicate descriptor designation in the Charda vocabulary, why for this vocabulary skos:altLabel was used. The problem of topography determination had to be addressed in the iDAI.gazetteer data, which resulted in the use of skos:related only for this vocabulary. An expected increase in recall spoke for this decision, however the correctness of this implementation is questionable on the semantic level.

With the use of skos:prefLabel for every term in each dataset the important multilingual character of the terminologies is supported, as no preference to any language is given and interoperability even across languages can be obtained.
While the broader term relationship is clearly indicated for the thesaurus and gazetteer data, this relationship is not explicitly shown in the Charda vocabulary. Therefore it was decided to view the Charda "categories" as a broader term, despite their heterogeneous nature (for example both 'weapon' and 'context of excavation' are mapped as concepts at the same level).
Due to the significantly enriched gazetteer records additional SKOS properties had to be used which could not be applied for the other vocabularies (e.g. natural language descriptions or links to external vocabularies such as geonames[10]). In contrast to that only three significant SKOS properties could be used for the flat term structure of the Charda data (namely skos:broader/ skos:narrower, skos:prefLabel and skos:altLabel).

The 'skosified' vocabularies were loaded into the Triplestore ClioPatria[11], and aligned using amalgame. Amalgame's focus on SKOS and the transparency of the alignments are important properties, which led to this tool's choice. Another reason is the related background within the scope of the Europeana[12]. The results were evaluated and the strategy adjusted accordingly.

1 http://www.w3.org/2004/02/skos/

2 http://www.dainst.org/

3 http://gazetteer.dainst.org/app/#!/home

4 http://charda-xplore.dainst.org/index.php/de/login

5 http://semanticweb.cs.vu.nl/amalgame/

6 https://github.com/mromanello/skosifaurus

7 http://www.gesis.org/unser-angebot/recherchieren/thesauri-und-klassifikationen/thesaurus-sozialwissenschaften/

8 http://zbw.eu/stw/version/latest/about

9 http://oaei.ontologymatching.org/

10 http://www.geonames.org/

11 http://cliopatria.swi-prolog.org/home

The string-based matches [2] were examined to investigate ambiguities, i.e. the one-to-one and many-to-many relationships. The results were represented in RDF, whereof samples were used for easier examination. The evaluation of the results on syntactic and semantic accuracy and the presentation of the results followed in tabular form.
Because no gold standard exists, the procedure for recall measurement needed to be checked. The precision measurements were estimated based on intellectual observation.

Overall, the alignments showed a 50% correctness, which is considered to be good, compared to the heterogeneous basis data. However, with a further selection of the data could more accurate results could be achieved in future alignment strategies.

Future work will consider various alignment strategies, to combine different alignment methods and tools to improve matching results [6]. In general it is planned to get the DAI data further connected, also through integrating it into the LOD cloud. The present work serves as a starting point for these aims and is intended to provide a methodology for KOS alignment in the archaeological field.

References

[1] N. Beer, K. Herold, W. Kolbmann, T. Kollatz, M. Romanello, S. Rose, & N.-O. Walkowski: *Interdisciplinary Interoperability*. (*DARIAH-DE working papers 3)*. GOEDOC, Göttingen (DE), 2014.
[2] J. Euzenat, & P. Shvaiko. Ontology matching. Springer-Verlag, Heidelberg (DE), 2013.
[3] A. O. Kempf, D. Ritze, K. Eckert, & B. Zapilko. New Ways of Mapping Knowledge Organization Systems. Using a SemiAutomatic Matching Procedure for Building Up Vocabulary Crosswalks. *Knowledge Organization*, 41(1):66-75, 2014.
[4] P. Mayr, & V. Petras. Crosskonkordanzen: Terminologie Mapping und deren Effektivität für das Information Retrieval, 2008.
[5] J. Neubert. Bringing the "Thesaurus for Economics" on to the Web of Linked Data. *LDOW* 25964, 2009.
[6] P. Shvaiko, & J. Euzenat. Ontology matching: state of the art and future challenges. *Knowledge and Data Engineering, IEEE Transactions on, 25(1)*:158-176, 2013.
[7] A. Tordai, J. van Ossenbruggen, & G. Schreiber. Combining vocabulary alignment techniques. In *Proceedings of the fifth international conference on Knowledge capture.* ACM, 2009.
[8] B. Zapilko, & Y. Sure. Converting the TheSoz to SKOS. (*GESIS Technical Report 2009/07*), 2009.

12 http://www.europeana.eu/portal/