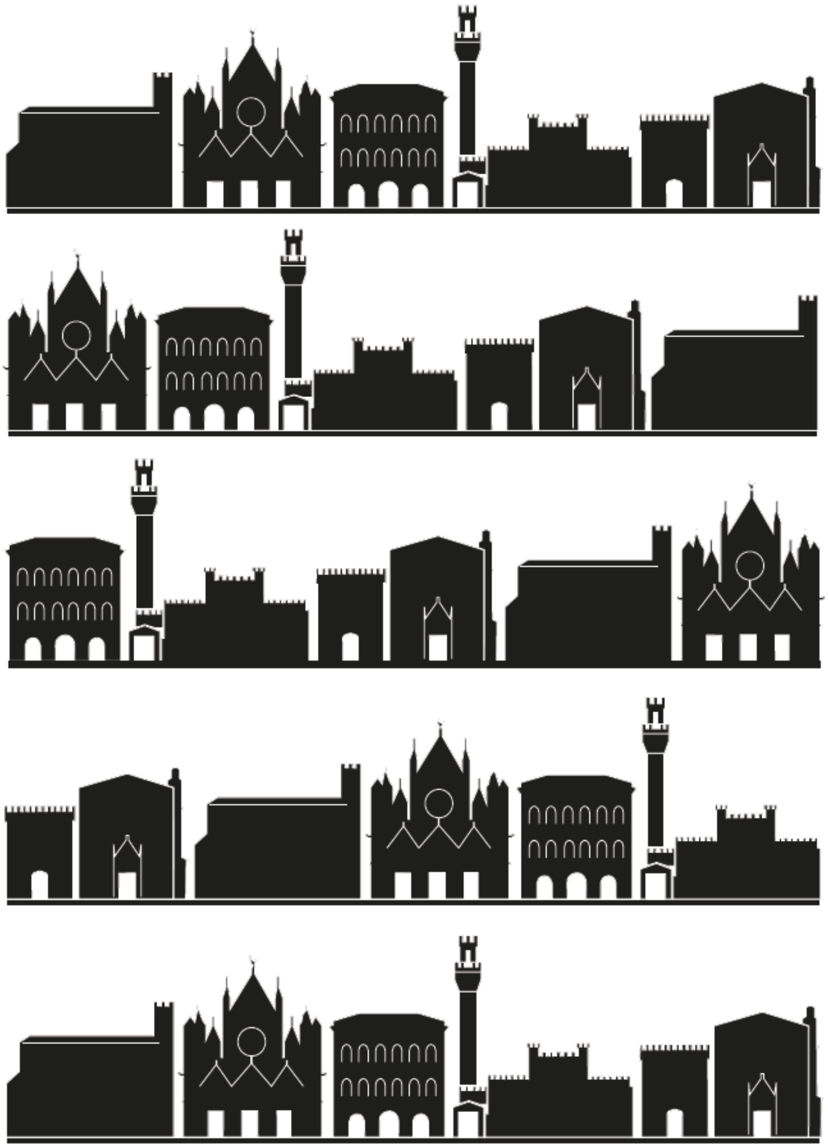
CAA2015

KEEP THE REVOLUTION GOING >>>

Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods In Archaeology

edited by
Stefano Campana, Roberto Scopigno,
Gabriella Carpentiero and Marianna Cirillo

Volumes 1 and 2



UNIVERSITÀ DI SIENA 1240

Computer Applications & Quantitative Methods in Archaeology

CAA2015

KEEP THE REVOLUTION GOING >>>

Proceedings of the 43rd Annual Conference on Computer Applications and Quantitative Methods in Archaeology

edited by

Stefano Campana, Roberto Scopigno, Gabriella Carpentiero and Marianna Cirillo

Volume 1

Archaeopress Archaeology

ARCHAEOPRESS PUBLISHING LTD
Gordon House
276 Banbury Road
Oxford OX2 7ED

www.archaeopress.com

CAA2015

ISBN 978 1 78491 337 3
ISBN 978 1 78491 338 0 (e-Pdf)

CAA2015 is availabe to download from Archaeopress Open Access site

This book is available direct from Archaeopress or from our website www.archaeopress.com

CHAPTER 4
LINKING DATA

The Labelling System:
A Bottom-up Approach for Enriched Vocabularies in the Humanities

Florian Thiery
florian.thiery@hs-mainz.de

Thomas Engel
thomas.engel@hs-mainz.de

i3mainz – Institute for Spatial Information and Surveying Technology, Mainz, Germany

Abstract: Shared thesauri of concepts are increasingly used in the process of data modelling and annotating resources in the Semantic Web. This growing family of linked data resources follows a top-down principle. In contrast, the Labelling System follows a bottom-up approach, enabling scientists working in the digital humanities to manage, create, and publish their own controlled vocabularies in SKOS (Simple Knowledge Organization System). The created concepts can then be interlinked with well-known LOD (Linked Open Data) resources, a process named the 'Labelling Approach'. The Labelling System is domain-independent, while uniting perspectives of different scientific disciplines on the same label and therefore contributing to interdisciplinary collaboration for building up cross- and inter-domain linked data communities. This paper addresses principles of the Labelling System in the light of archaeological use cases.

*Keywords: Con*trolled vocabulary, ontology, Semantic Web, Linked Open Data, SKOS

Introduction

Recently, the humanities have been aiming to interlink their areas of knowledge within the Semantic Web. This necessarily involves the implementation of a controlled and standardized vocabulary, to solve the well-known problem of ambiguity of terms like 'Roma', which may apply to the antique capital of the Roman Empire as well as to a European ethnic group. Obviously, the meaning of a term is revealed through its context. Particularly for working in cross-domain environments, a common understanding of terms is essential for collaboration, joint analysis, and data exchange.

Some institutions already provide authoritative thesauri as LOD (Berners-Lee 2009). Their hierarchically ordered terms appear very much like long-established vocabularies of natural sciences. But does this top-down approach meet the demands of the discursive character of the humanities? And if not, how can we enable humanists to build up their own vocabulary and make it accessible, transparent, quotable, and reusable?

This paper proposes solutions to overcome the vocabulary bottleneck using the Labelling Approach.

1.1 Controlled vocabularies: standards and implementations

The development of cultural and historical concepts requires the abstraction of historical realities in tags or classified items. A collection of project-specific terms are called controlled vocabularies. Such a vocabulary can be defined as a managed set of terms in one or more languages designed for a particular purpose. This is incorporated in multiple standards like ISO 25964 as well as the W3C standardized SKOS (Miles and Bechhofer 2009). SKOS is a data model for knowledge organization systems like thesauri to represent them in RDF, the Resource Description Framework (Klyne *et al.* 2014), the technical basis of the Semantic Web (Eckert 2011).

Designed as general-purpose tools, there are some applications for creating, editing, and sharing SKOS vocabularies like TemaTres (TemaTres 2015), OpenSKOS (Picturae 2013), VocBench (Food and Agriculture Organization of the United Nation, n.d.), and CultuurLINK (CultuurLINK, n.d.). From a technical point of view, they do not provide the usability required for humanist domains. Existing systems present a high barrier in terms of conceptual knowledge in Linked Data techniques. Furthermore, they do not provide simple collaboration features to strengthen the scientific discourse during the creation process. Hence, a different approach is necessary.

1.2 Reference thesauri in the Semantic Web

Today, a growing number of controlled vocabularies in machine-readable formats are available. Most of them are part of the Semantic Web and accessible in open standardized formats, particularly modelled in RDF as LOD. Pioneers in this field are the natural sciences, especially biomedical sciences (Momtchev *et al.* 2009). Research in the humanities is project-specific and terms are often a subject of scientific discourse, but some organizations, such as SKOS thesauri, are starting to model their knowledge for special disciplines. In this context we refer to LOD vocabulary examples of the Getty Research Institute (Getty Research Institute 2015) as well as Historic England (Heritage Data 2015).

1.3 From keyword lists to enriched thesauri: the term 'Roma' as an example

The development of a controlled vocabulary can be divided into three crucial steps: analysis, construction, and maintenance. While building up a set of terms and their relations, an 'evolutionary process' is set into motion: starting from a 'naive term' as a keyword, evolving to a meaningful 'intelligent term' formed by descriptions via strings or links in the web.

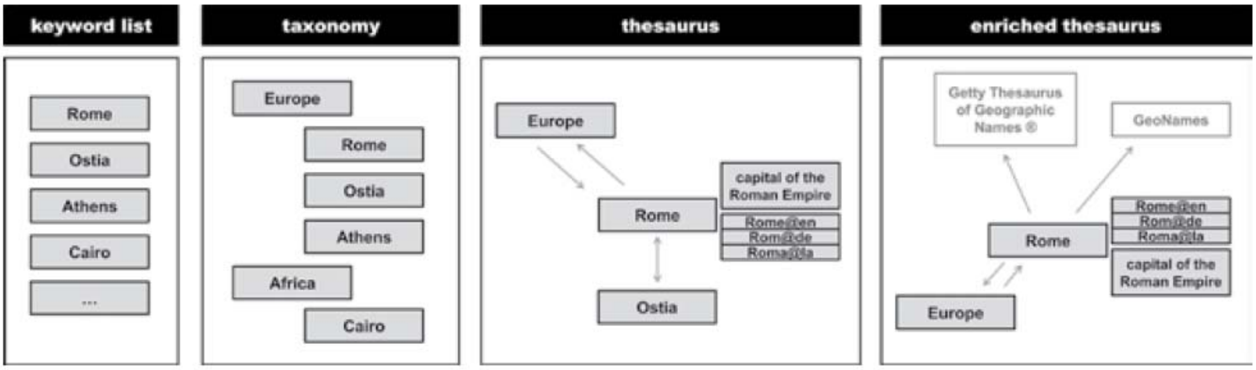


FIG. 1. EVOLUTIONARY PROCESS USING THE TERM 'ROME'.

Figure 1 shows the development of a thesaurus containing central cities in the Roman Empire. First, 'naive' terms are collected in a keyword list. Second, these terms are organized in a hierarchy (taxonomy), for example according to geographical regions. Third, multilingual concepts containing labels are created and organized relationally and can therefore be used for indexing and searching (thesaurus). Finally, these thesauri concepts are linked to reference thesauri in the worldwide (semantic) web. These 'intelligent terms' organized in enriched thesauri can be understood and classified by scientists of any discipline because of their linkage to globally defined concepts. As an example, the term 'Roma' can be linked to authoritative thesauri like GeoNames (Geonames, n.d.) or the Getty Thesaurus of Geographic Names (TGN 2013) defining an administrative unit instead of an ethnic group.

1.4 Bottom-up approach

Within humanities research it is impossible to define controlled vocabularies that cover all conceivable applications and are generally accepted. Therefore, a top-down approach developing authoritative reference thesauri as described in section 1.3 can be just part of the solution. So how can the problem of broad rather than project-specific thesauri be solved? Building up self-defined terms linked to reference thesauri can be a way out of this bottleneck. This is why we propose the bottom-up 'Labelling Approach' (Section 2).

Take as an example the term 'potter': in a particular research question, such as the distribution of trade networks of Samian Ware, a 'potter' can be defined as a human being or organization that lived in the Roman Empire and was responsible for producing Samian Ware. The Getty AAT term 'potters' describes them as makers of vessels, tableware, vases, and other ware made of ceramic (Jean Paul Getty Trust 2004). A link to that term cannot cover all definitions for a 'Samian Potter'. This can be solved using the bottom-up approach explained in the following sections.

2 Labelling Approach

The 'Labelling Approach' is based on the idea that each vocabulary term, built for a specific application, is defined by linking to one or more concepts in a reference thesaurus, available as an HTTP-URI (Berners-Lee *et al.* 2005) (Fig. 1). A label can be understood as a vocabulary term, represented by a particular set of concepts. The approach is generic and can be applied to any scientific domain. It is motivated by the fact that vocabulary construction and semantically defining terms is hard for domain experts and even harder in a research context, where the vocabulary is part of the research process (Piotrowski *et al.* 2014).

Labels created through the Labelling Approach are specified as concepts within the SKOS ontology, providing the flexibility to structure concepts relationally and multilingually. It provides a set of methods and ideas for designing a controlled vocabulary, helping researchers to communicate internally, as well as between disciplines using LOD (Section 4). The Labelling Approach contains the following steps:

1. Creation of detailed concepts for individual research questions: each specific research topic has its own list of concepts with an individual meaning. Concepts in authoritative reference thesauri are often broad and generic. Individual concepts, however, are assigned according to their position within the reference thesauri hierarchy. Defining a relation to an existing concept permits integration of personal labels within the Semantic Web (Section 4.1).

2. Enrichment of concepts by linking into a hierarchy of a domain expert's reference thesaurus — adoption of an expert's knowledge and hierarchical structures and relations for a label (Section 4.2 and Figure 2).

3. Defining a label to add specifications of different domains: the process of interlinking vocabulary terms to different concepts helps the user to clarify the reasoning and the layer of knowledge the label is representing (Section 4.3).

4. Linking generic tags to specific contexts: one term can have different meanings. They are specified using diverse reference concepts (Section 4.3).

3 Labelling System

The current web-based prototype (i3mainz, n.d.) of the Labelling System, implementing the Labelling Approach, is freely available under an MIT licence (i3mainz 2015). The Labelling System uses open source technologies (Section 3.1) and common LOD standards (Section 3.2). Its main purpose is to enable users to create thesauri without any knowledge

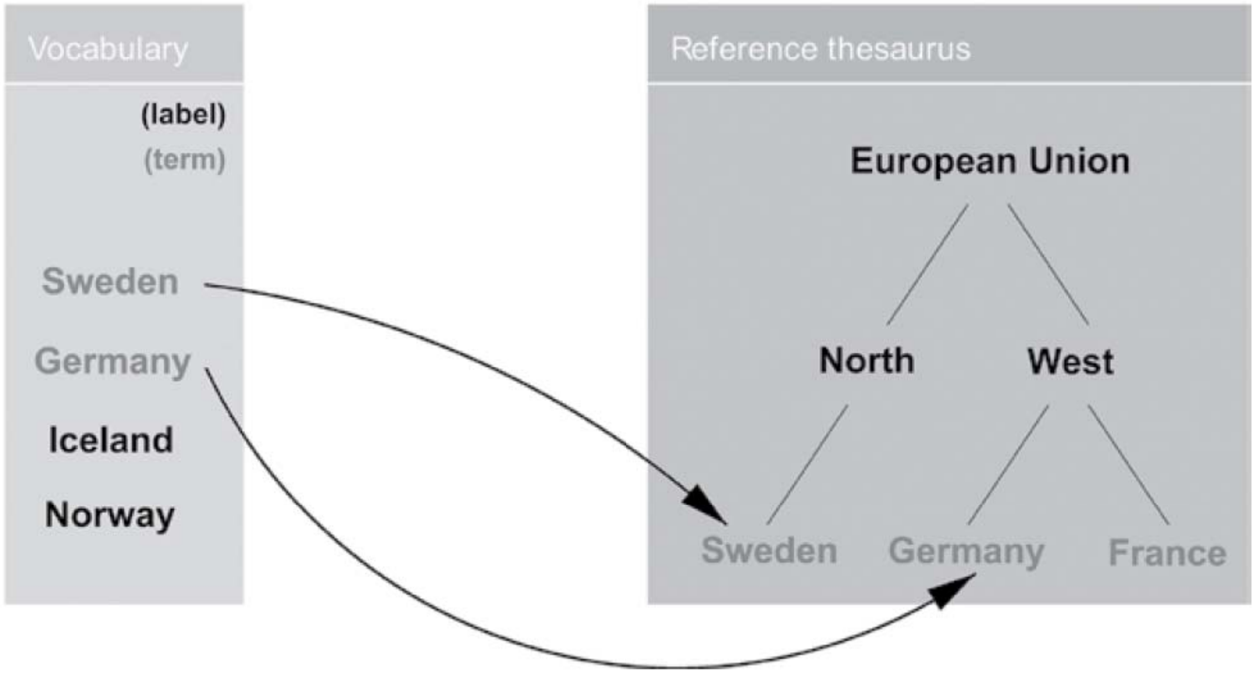


FIG. 2. LABELLING APPROACH.

of SKOS, RDF, or ontologies. Collaboration projects in vocabulary term creation are also explicitly encouraged.

To summarize, the Labelling System is a web-based tool:

- for creating controlled vocabularies;
- for building internal vocabulary term relations;
- for enriching vocabulary terms and semantically modelling their relationships to external concepts belonging to thesauri (=labelling);
- for publishing and sharing vocabularies as quotable URI.

Within the Labelling System, several user roles are defined. The permissions are structured hierarchically: higher-level roles subsume the permissions of lower-level roles. Agents are permitted to query the system for available content via the Semantic Web interfaces (Section 3.4). Users are able to create enriched controlled vocabularies. In addition, ontologists are allowed to import, for example, reference thesauri by storing the URI for a SPARQL endpoint.

All steps of the evolutionary process from a simple keyword list to enriched thesauri, described in section 1.3, can be handled within the Labelling System. Building up enriched thesauri using the Labelling Approach is as easy as the creation of a 'simple' keyword list (Figure 3).

The Labelling System is not designed to be a centralized service. In contrast, using an open-source approach, it enables creating instances in different infrastructures, producing customized URIs as a consequence.

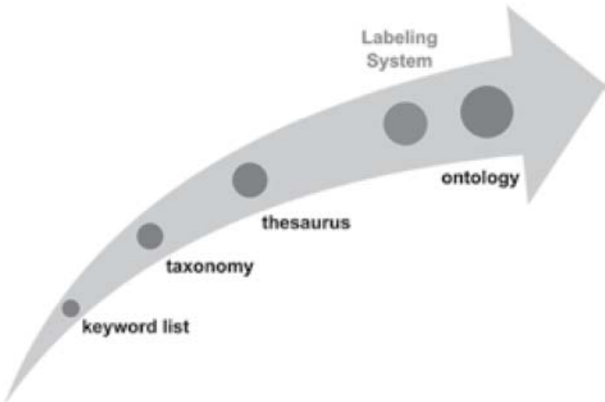


FIG. 3. LOCATION OF THE LABELLING SYSTEM WITHIN THE EVOLUTIONARY PROCESS.

3.1 Technology

The prototype is built on top of two open-source frameworks using Java, Maven, PHP, and MySQL: OpenRDF Sesame triplestore (Brockstra 2002) and Usercake management system (Usercake 2012). The current client-server application is running on Linux distributions, using Apache HTTP Server as well as Apache Tomcat. The prototypical graphical user interface (Labelling System, n.d.) is based on HTML5, CSS3, and JavaScript.

OpenRDF Sesame is an open source Java framework for processing RDF data. It offers an API that can be handled in Java servlets including libraries for managing RDF data like Apache Jena. As a consequence, all data and resources produced by the Labelling System and their inner relations are stored in an RDF ontology (Section 3.2) and are therefore queryable through the triplestore.

In contrast, UserCake uses a relational database (MySQL). It controls the administration of users and their specific roles and is extendable for LDAP based authentication.

Using technologies like HTML5, CSS3, and JavaScript enables the presentation of complex structures in a user-friendly and familiar layout. JavaScript libraries like jQuery and D3.js support displaying complex semantic data structures.

3.2 Ontologies

The Labelling System aims for scalability and strict conformity with common standards. This implies the usage of standardized Linked Data models, in particular SKOS and Dublin Core (Dublin Core 2012), as well as FOAF (Brickley and Miller 2014) and RDF-Schema (Brickley and Guha 2014).

A well-defined list of terms developed by the Dublin Core Metadata Initiative (DCMI), the 'DCMI Metadata Terms' or 'Element Set', is perfectly suited for describing metadata of Labelling System resources. In particular the properties dc:creator, dc:identifier, dc:date, dc:language, and dct:licence are used.

A SKOS vocabulary is based on concepts as units of thoughts. In the human mind concepts exist as an abstract entity. A SKOS concept is therefore used to represent items (terms, ideas, meanings, objects, or events) in a knowledge organization system. Furthermore, a SKOS concept scheme is similar to a vocabulary and a container for concepts. Labelling System vocabularies are stored as a skos:ConceptScheme. Labels as skos:Concept belong to a scheme. Each concept scheme is available as URI and is also downloadable. Conceptually, the product is comparable to big authoritative vocabularies of providers like the Getty Research Institute or Historic England.

As mentioned in section 3.1, all resources and instances of the Labelling System and their internal relations are structured in an RDF ontology (Thiery 2015a). Managing the inner structure requires a set of classes and properties defined in the 'Labelling System Vocabulary' (Thiery 2015b). The Labelling System is composed of five big classes: ls:Project, ls:Vocabulary, ls:Label, ls:SPARQLendpoint, and ls:GUI. A project is a container for a set of vocabularies. A vocabulary is a set of arbitrary terms and can be hidden or public (ls:state). A label is a vocabulary term linked to a particular set of published concepts characterized by a preference language (ls:prefLang) for visualization reasons. Each connection is created via a bidirectional ls:contains and ls:belongsTo property. Each SPARQL endpoint imported by an ontologist is stored with its properties, such as name, query, and SPARQL endpoint URL. Users are able to store personal metadata to emphasize the individual research question of their own projects, vocabularies, and labels. Properties defining individual user-GUIs exist but are not used in the current prototype. Labelling System classes are identified by HTTP addressable URIs (ls:identifier) based on the Universally Unique Identifiers (UUID) (Leach *et al.* 2005).

3.3 Labelling process

A label is independent and primarily identified with an UUID, quotable via an URI. It may belong to one or more vocabularies and projects. All these components are loosely coupled. In addition, labels provide a mandatory multilingual human-readable appellation (rdfs:label or skos:prefLabel) which represents a typical term in a controlled vocabulary. Alternative multilingual appellations (skos:altLabel) and documentation terms like notes (skos:note) and definitions (skos:definition) are optional and can be used to describe a label in free texts. Each vocabulary including labels is published under a Creative Commons Attribution (CC-BY) licence. The relations between concepts are not bidirectional and only reflect the author's personal view of his concepts. In the modelling process, legal rights of thesauri providers have to be respected. Currently, no 'code of ethics' for creating vocabularies exists. However, the Labelling System should not be used to create copies of existing thesauri to circumvent legal aspects.

The 'Labelling System Ontology' provides a well-defined canon of semantic predicates, which can be used to link a vocabulary term to an external SKOS concept or web resource as well as to generate internal relations. Applying the Labelling Approach, links to concepts of reference thesauri, or simple HTTP web resources, eight RDF properties come into play (Figure 4). SKOS supports hierarchical, associative, and mapping relationships.

For building up hierarchical relations in a vocabulary, skos:broader and skos:narrower are used. The broader and narrower terms are not transitive, but are bidirectional connected within the Labelling System Ontology. In contrast, the associative related property (skos:related) implies an unidirectional relationship. By convention, the SKOS mapping properties are expected to be asserted between concepts that belong to different concept schemes. The relationships skos:closeMatch and skos:exactMatch represent concept mappings. The skos:closeMatch term indicates that two concepts are sufficiently similar and that they may be used interchangeably in some contexts. Skos:exactMatch denotes a higher degree of similarity. Both concepts have the same meaning in all contexts. The mapping properties skos:broadMatch, skos:narrowMatch, and skos:relatedMatch are equivalent to skos:broader, skos:narrower, and skos:related but refer to concepts belonging to different concept schemes (Miles and Bechhofer 2009).

Linking to other web resources that are not modelled as SKOS concepts can be realized using the rdfs:seeAlso, rdfs:isDefinedBy, and owl:sameAs properties.

For linking resources, the following best practices for RDF properties are as follows:

- broader/broadMatch: My label A has a broader Label/Concept B;
- narrower/narrowMatch: My label A has a narrower Label/Concept B;
- related/relatedMatch/seeAlso: My label A is related in some way to Label/Concept/Resource B;
- closeMatch/isDefinedBy: My label A is similar to Concept/Resource B;
- exactMatch/sameAs: My label A is the same as Concept/Resource B.

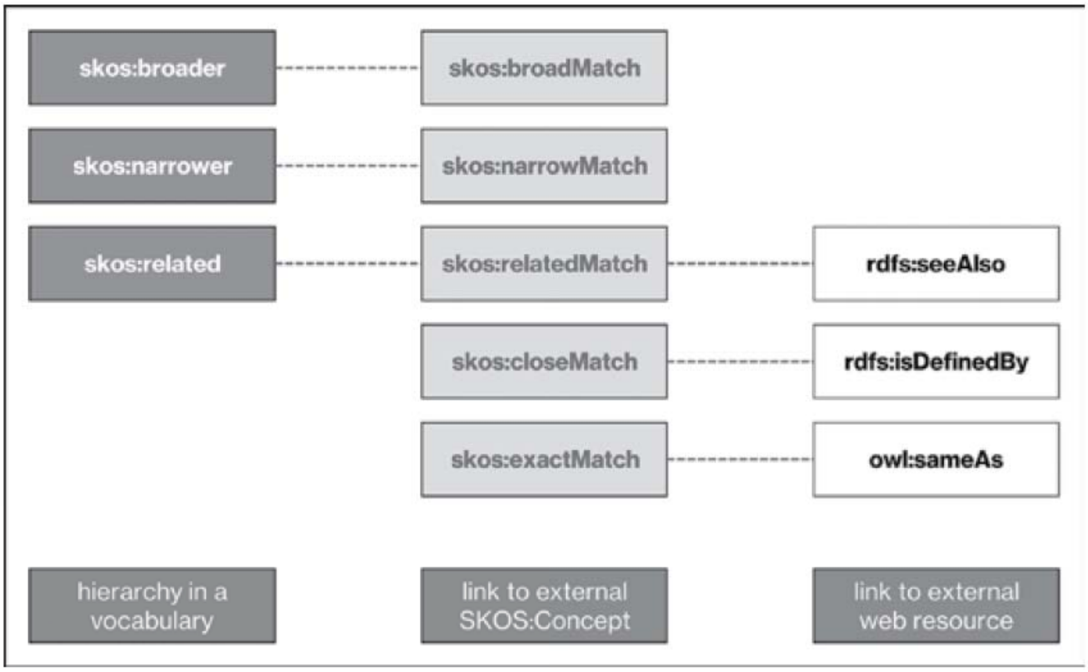


Fig. 4. Labelling System predicate canon.

The labelling process will be supported by a user-friendly GUI in the release version. Integrated SPARQL endpoints of reference thesauri providers allow searching for the existence of a particular substring in the set of concepts, and returns the resource. Uploaded SKOS concept schemes can be queried the same way. Finally, a manual entry of resource URIs provides access to the world of digital encyclopaedias (e.g. DBpedia) or gazetteers such as GeoNames or Pleiades.

The single steps of the labelling process can be combined in a semi-automatic procedure. First, appellations, descriptions, and links are placed into a CSV-file and uploaded to the server. Then the Labelling System server application validates all labels and their relations and converts them into triples aligned to the ontology. This possibility is designed for humanists who keep their data mostly in tables and do not want to change their usual methodology for the process of building enriched controlled vocabularies.

3.4 Semantic Web interfaces

The Labelling System offers two major Semantic Web interfaces for all resources and relations created in the labelling process: the REST-API (Thiery 2015c) and the SPARQL-API (Thiery 2015d).

All data and properties are stored in a well-defined RDF ontology. The SPARQL-API enables querying every resource individually with its relations and literals of triples using SPARQL (W3C 2013). Any user can access the SPARQL endpoint and include it into their own application via HTTP requests. Humanists can use a graphical.

The REST-API gives access to the Labelling System resources through an XML browsable interface. Each resource can be fetched with HTTP-GET and will be represented in well-defined and standardized linked data formats such as RDF, TTL, N3, and JSON-LD. It is also possible to download a dump of a skos:ConceptScheme. This dump could be used in other software frameworks using SKOS modelled thesauri, for example the Arches RDM (Getty Conservation Institute and World Monuments Fund 2015).

In contrast to the SPARQL endpoint, in the REST view (Richardson 2007), a label is only reachable via a particular concept scheme. Thus, a label implies a special use case.

4 Use cases in the humanities

The use of terms within the humanities can be a source of ambiguities in many ways. As stated in section 3.2, knowledge in archaeological domains is often based on the creation of theoretical concepts. Not only classic areas of ambiguities such as time concepts, space, place, or place types are the results of this process. The use of typology classifications for archaeological artefacts can also be misleading, as the meaning of concept terms may vary through time, within different regions and between authors and/or schools. Furthermore, modern archaeology as an interdisciplinary science incorporates several natural sciences as well as humanities domains. Those external domains have already defined generic vocabularies and corresponding data in many cases, creating useful authoritative resources to link to. Self-created thesauri are naturally more detailed than authoritative ones, which is reflected in very few skos:narrowMatch relations in contrast to skos:broadMatch or skos:relatedMatch. This can be different in other disciplines (Piotrowski *et al.* 2014). Creating new, more granular vocabularies located within the Semantic Web, humanists are able to publish a quotable thesaurus in the same way that text or data is published.

The following four examples have been identified as prototypical archaeological use cases.

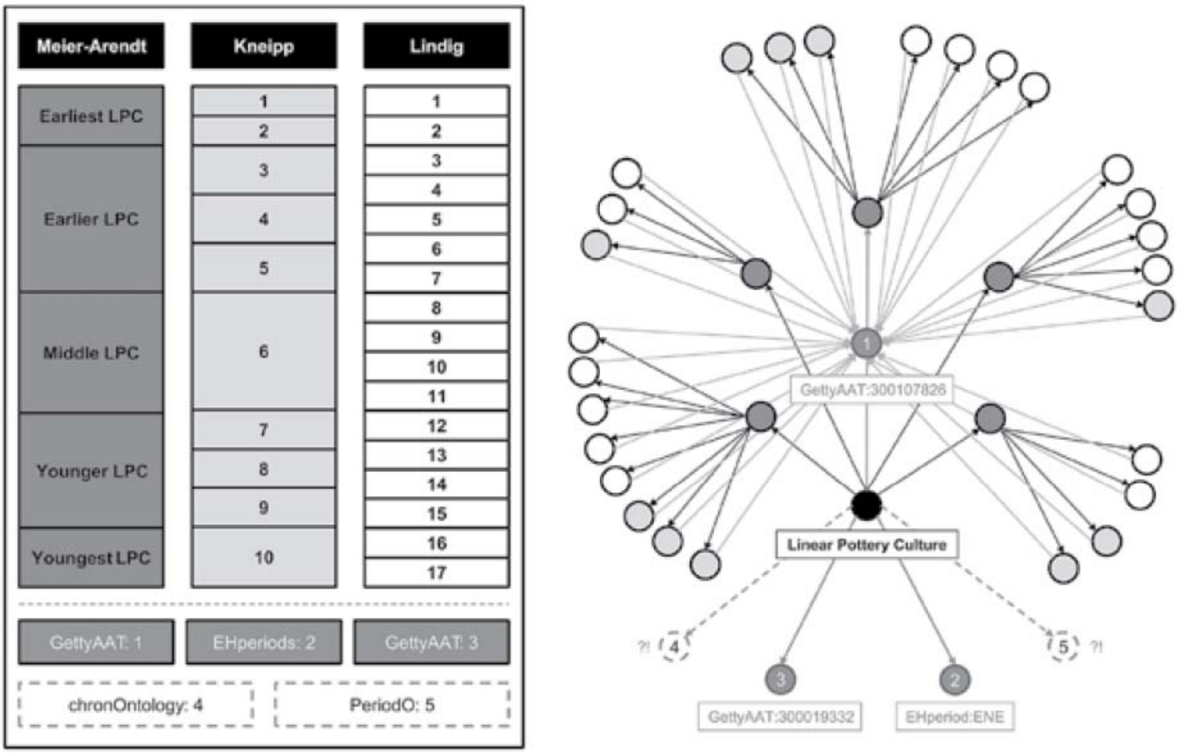


FIG. 5. USE CASE TIME CONCEPTS.

4.1 Creating concepts for individual research questions

As a key component of archaeological research, the conceptualization of time changes frequently. The same term describing a temporal phase can be used to express different ideas of it. Until now, a researcher's exact understanding of a time concept is often undefined. A researcher using temporal concepts today is able to use the Labelling System to define them by adding a relation to authoritative vocabularies and add useful metadata in order to make his/her research comprehensible. Authoritative vocabularies provide generic representations of knowledge resulting in their granularity not being sufficient for certain research questions.

Starting the conceptualization, a label is created that ideally includes all of the temporal concepts concerned and is linkable to at least one external authoritative SKOS thesaurus. In this use-case, the label is named 'Linear Pottery Culture'. Following on, for each temporal system within the linear pottery culture (Meier-Arendt 1966; Kneipp 1998; Lindig 2002) a new label is created, starting with the least granular one. This enables the researcher to build up a graph of phases that reflects the hierarchy of concepts.

Every created label gets a quotable UUID available as URI. The whole enriched thesaurus is available online and downloadable as a SKOS concept scheme. In the next step, all created concepts can be linked to external resources/thesauri using the proper relation (Figure 5; Section 3.3). In the near future, LOD gazetteers for temporal concepts will be delivered by projects like chronOntology (DFG 2015) and PeriodO (PeriodO n.d.). The latter will also offer a service to create definitions of a temporal concept. Providing a more generic approach, the Labelling System covers all domains. Consequently, scientists must decide where the creation of temporal concepts would be reasonable.

It is possible to create detailed temporal concepts sufficient for individual research questions. Furthermore, relations can be interlinked to existing temporal concepts like Getty AAT or Historic England. The relations between phases only include hierarchical and associative structures, so they are limited to skos:related, skos:narrower, and skos:broader relations. Relative chronological relationships like 'during' or 'starts with' (see Freksa 1992) cannot be implemented as they are not within the scope of the Labelling System project.

4.2 Enriching concepts by linking into hierarchies of reference thesauri

Self-created concepts concerning external scientific domains are normally not aligned to thesauri created by relevant domain experts. As an example of a natural sciences vocabulary that can be enriched by considering authoritative concepts, bones of a human skeleton were conceptualized.

After the creation of hierarchically aligned concepts for a human skeleton, these concepts are interlinked to authoritative thesauri. In this example, the concepts with the appellations 'Ulna' and 'Radius' have a skos:narrower relation to a higher-level concept 'Human Skeleton'. Furthermore, they are skos:related to each other to point out that they are both forearm bones.

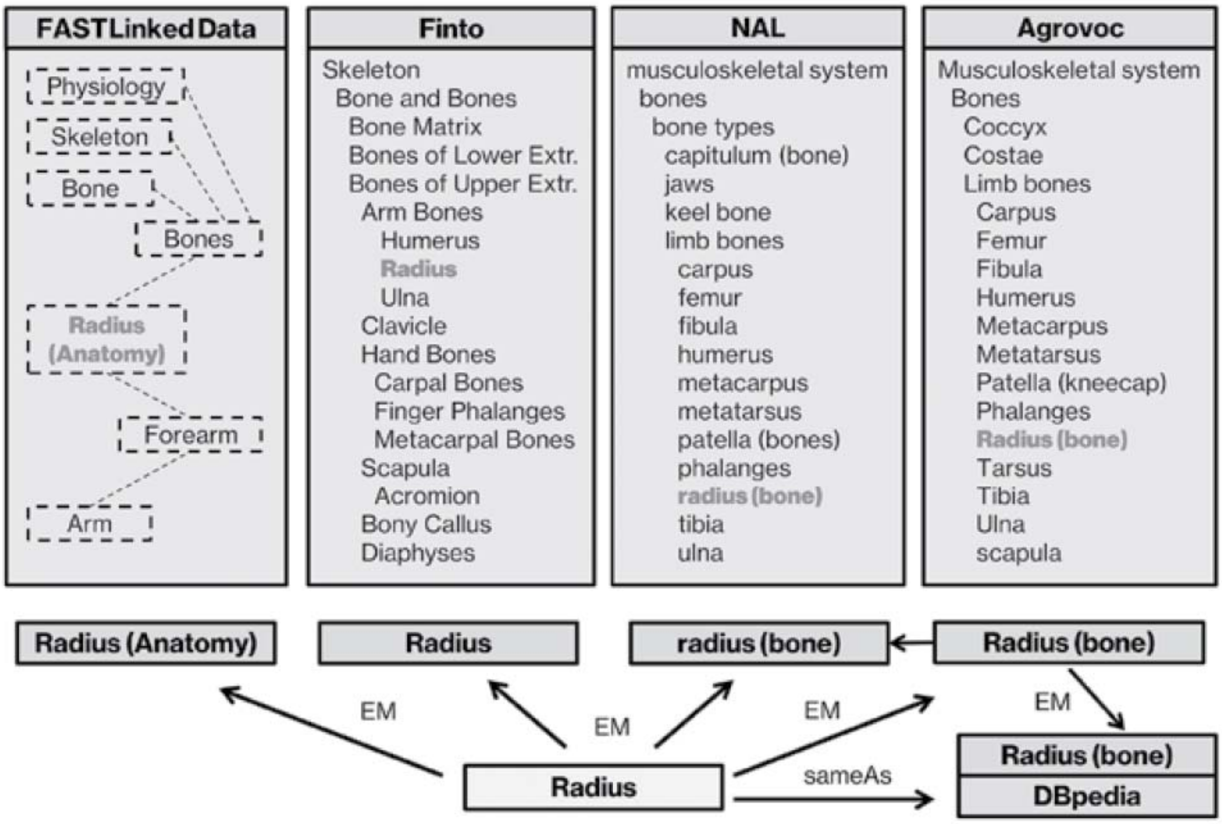


FIG. 6. USE CASE RADIUS BONE.

In a second step, external vocabularies are identified providing different concepts of bones in the human body. These were conceptualized for certain reasons, a fact that has to be considered before defining relations. Moreover, the concept must be inspected closely to identify its alignment to other interlinked vocabularies (Figure 6). If a concept expresses exactly the same as the example's 'Radius' concept a skos:exactMatch relation is necessary. Otherwise, choosing skos:closeMatch or skos:relatedMatch might be more reasonable.

By linking into vocabulary hierarchies of domain experts, the value of self-created concepts will be increased. Adopting expert knowledge, especially the structure relations, ensures a higher level of standardization. Furthermore, authoritative vocabularies are interlinked via self-created thesauri and are therefore loosely coupled. This can be seen as an alignment between existing vocabularies originating from different domains.

4.3 Defining concepts by adding properties from different domains

Appellations of place types or functions can describe different things in space, time, or culture. Generic tags for specific meanings lead to ambiguities that are difficult to resolve if their context is not transferred. There are different layers of knowledge, e.g. historical agents, socio-political contexts, or historians' interpretations. Depending on the used sources, an appellation can cover a variety of meanings.

To show the ambiguity a place type appellation can adopt, the place type 'earthwork' is created twice. As the concept names are usually associated with different things, it is crucial to define their meaning further (Figure 7). In this example, the first concept 'earthwork' is defined as a dwelling, has an agricultural aspect, and dates to the early Neolithic period. The second concept 'earthwork' is a dwelling with a shelter function used for defensive purposes. The functions defining the concept can also be related to authoritative vocabularies so that no ambiguities are generated on this second level.

A researcher is able to add specifications of space, time or culture, or any additional attributes to concepts defining a term. By doing so, generic tags are linked to specific concepts. In this example, a concept 'place type' is defined by terms expressing its functional context. The process of interlinking vocabulary terms of concepts to a functional concept, which is interlinked itself, helps humanists to clarify the reasoning and layers of knowledge. In this way, a structured, well-defined vocabulary for place types can be created, expressing exactly what it was intended for.

4.4 Synthesis of labelling methods in the world of Samian Ware

As described in section 1.4, it is currently impossible to find an existing concept for a 'Samian Potter'. The Labelling Approach can correct things by defining temporal concepts and the function as a person or organization that is responsible for

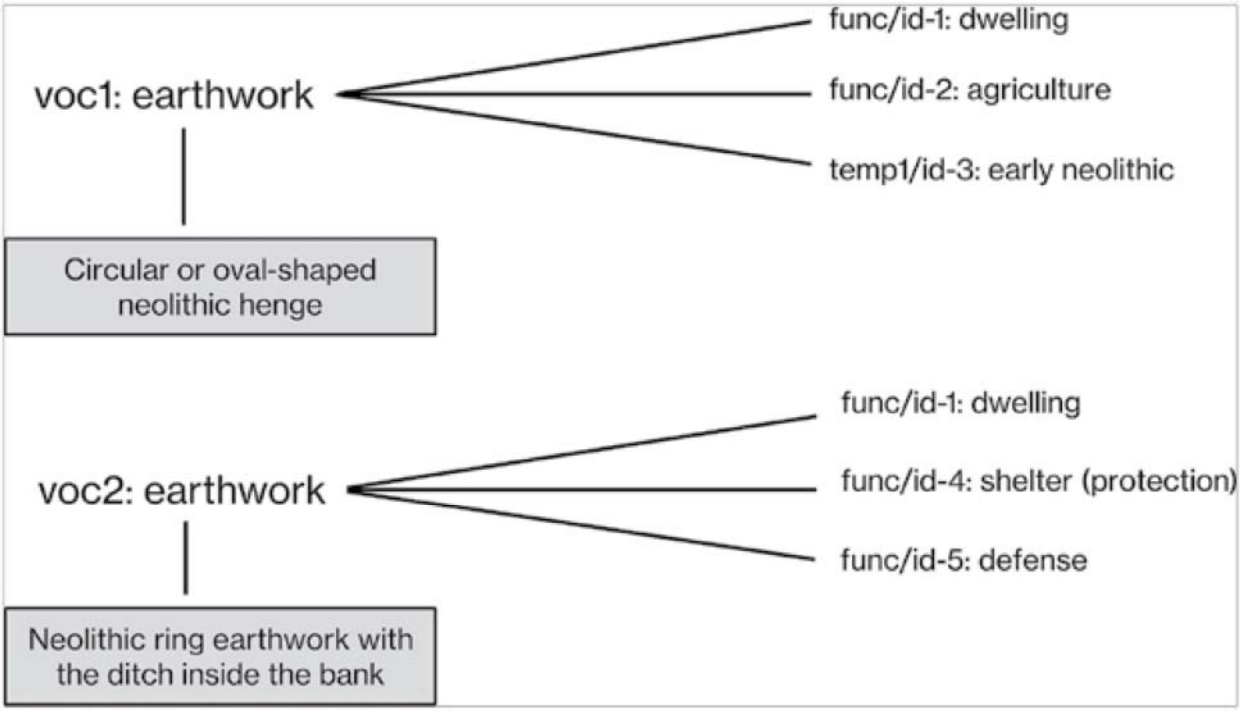


Fig. 7. Use case earthwork.

creating and trading Samian Ware. Links to Pelagios (Pelagios, n.d.) or Pleiades (Pleiades, n.d.) can locate the term 'potter' into the Linked Data Cloud as a synthesis of the mentioned use cases in the world of Samian Ware.

5 Remaining challenges

Using Linked Open Data and the Semantic Web to build up a network of concepts entails advantages, and also questions and challenges that are to be managed in the future to ensure the success of the Labelling Approach. The following issues have been identified so far:

1. It is hard to find reference thesauri that could be imported into the Labelling System. A definition for a catalogue service, which makes thesauri of all domains in the Semantic Web searchable, is not yet in sight. It might be a good idea to implement such a service in the cultural heritage domain as a prototype. A first concept is currently being discussed at i3mainz.

2. As a consequence of unidirectional relations between two vocabulary terms, it is not possible to find all resources in the Semantic Web linked to a single term. For example a label 'Roma' which is defined via unidirectional links (e.g. skos:relatedMatch) to Getty TGN and GeoNames does not inform the authoritative thesauri that the relation exists. A service analysing and displaying such relationships would identify the relevance of thesauri and their single terms through the number and provenance of links.

3. A major search engine in the LOD cloud, searching all interlinked repositories, is missing. It is usually possible to query one repository via a SPARQL endpoint and to follow the resulting links. Using gazetteers like Pleiades (AWMC, n.d.) and through multiple access points, the Pelagios Project (Pelagios, n.d.) works on basic approaches to connect repositories in the field of ancient places.

4. The open approach of the Labelling System entails the risk of building identical vocabularies (Section 3.3). This requires a proper Labelling Process as best practice: (a) to search for existing thesauri; (b) discuss them and as a consequence go through a mutual learning process; (c) adjust the individual approach.

6 Outlook

The Labelling System as the implementation of the Labelling Approach is still in a prototype phase. Use cases of various research areas will facilitate further development.

Currently, a user-friendly graphical user interface is still missing. Human-computer interaction in accordance to an intuitive GUI will help solve the major challenge of involving people outside the Semantic Web and Linked Data community in the labelling processes. Hence, the key goal for the release version is a 'labelling framework' for non-LOD experts.

For instance, the Labelling System may be used for the Arches Project (Arches 2012) and their Reference Data Manager, which is able to import SKOS thesauri. Being able to annotate point clouds via RDF triples in a web interface, the Generic Viewer (GenericViewer 2014) also supports SKOS. Furthermore, big databases placed in one organization, like TOMBA and NAVIS (RGZM, n.d.), could use Labelling System labels to interchange information on a meta-level.

Finally, humanists must be able to find and evaluate labels in the Semantic Web cloud independently. Exclusive 'islands of knowledge', published, standardized, represented as Linked Open Data, and linked to reference thesauri is just the beginning. Although additional efforts are still necessary, the Labelling Approach provides a generic solution to eliminate the semantic bottleneck.

Bibliography

Arches (n.d.) *What is Arches?*. [Online] http://archesproject.org/what-is-arches. [Accessed: 11 november 2015].

AWMC (n.d.) *About Pleiades*. [Online] http://pleiades.stoa.org/home. [Accessed: 11 november 2015].

Berners-Lee, T. 2009. *Linked Data*. [Online] http://www.w3.org/DesignIssues/LinkedData.html. [Accessed: 11 november 2015].

Berners-Lee, T., Fielding, R., Masinter, L. 2005. *Uniform Resource Identifier (URI): Generic Syntax*. [Online] https://tools.ietf.org/html/rfc3986. [Accessed: 11 november 2015].

Brickley, D., Guha, R.V. 2014. *RDF Schema 1.1*. [Online] http://www.w3.org/TR/rdf-schema. [Accessed: 11 november 2015].

Brickley, D., Miller, L. 2014. *FOAF Vocabulary Specification 0.99*. [Online] http://xmlns.com/foaf/spec. [Accessed: 11 november 2015].

Broekstra J., Kampman A., van Harmelen, F. 2002. Sesame: A generic architecture for storing and querying RDF and RDF Schema. In: Horrocks, I., Hendler, J. (eds.), *The Semantic Web – ISWC 2002. First International Semantic Web Conference Sardinia, Italy, June 9–12, 2002 Proceedings*: 54-68. Berlin, Springer.

CultuurLINK (n.d.) *Connecting Cultural Heritage. Align your vocabulary with the Dutch cultural heritage hub*. [Online] http://cultuurlink.beeldengeluid.nl. [Accessed: 11 november 2015].

DFG 2015. *chronOntology: ein Zeit-Gazetteer für die historisch-kulturwissenschaftlichen Fächer*. [Online] http://gepris.dfg.de/gepris/projekt/247900032. [Accessed: 11 november 2015].

Dublin Core 2012. *Dublin Core Metadata Element Set*. [Online] http://dublincore.org/documents/dces. [Accessed: 11 november 2015].

Eckert, K. 2011. *SKOS: eine Sprache für die* Übertragung *von Thesauri ins Semantic Web*. [Online] http://www.metadaten-twr.org/2011/01/19/skos-simple-knowledge-organisation-system. [Accessed: 11 november 2015].

Freksa, C. 1992. Temporal reasoning based on semi-intervals. *Artificial Intelligence 54*. Available from: http://cindy.informatik.uni-bremen.de/cosy/staff/freksa/publications/TemReBaSeIn92.pdf. [Accessed: 11 november 2015].

Food and Agriculture Organization of the United Nation (n.d.) *VocBench 2.0*. [Online] http://vocbench.uniroma2.it. [Accessed: 11 november 2015].

GenericViewer (n.d.) *GenericViewer*. [Online] http://www.spatialhumanities.de/en/ibr/technology/genericviewer.html. [Accessed: 11 november 2015].

Geonames (n.d.) *About GeoNames*. [Online] http://www.geonames.org/about.html. [Accessed: 11 november 2015].

Getty Conservation Institute and World Monuments Fund 2015. *Reference Data Manager (RDM). Revision e418cca2de6a+*. [Online] http://arches3.readthedocs.org/en/latest/rdm. [Accessed: 11 november 2015].

Getty Research Institute 2015 *Getty Vocabularies as Linked Open Data*. [Online] http://www.getty.edu/research/tools/vocabularies/lod/index.html. [Accessed: 11 november 2015].

Heritage Data 2015. *Vocabularies*. [Online] http://www.heritagedata.org/blog/vocabularies-provided. [Accessed: 11 november 2015].

i3mainz (n.d.) *The Labelling System*. [Online] http://i3mainz.hs-mainz.de/en/projekte/Labellingsystem. [Accessed: 11th November 2015].

i3mainz 2015. *The Labelling System*. [Online] https://github.com/i3mainz/LabellingSystem. [Accessed: 11 november 2015].

Jean Paul Getty Trust 2004. *potters*. [Online] http://vocab.getty.edu/aat/300025414. [Accessed: 11 november 2015].

Klyne, G., Carroll, J.J., McBride, B. 2014. *RDF 1.1 Concepts and Abstract Syntax*. [Online] http://www.w3.org/TR/rdf11-concepts. [Accessed: 11 november 2015].

Kneipp, J. 1998. *Bandkeramik zwischen Rhein, Weser und Main*. Universitätsforschungen zur Prähistorischen Archäologie 47. Bonn, Habelt.

Labelling System (n.d.) *The Labelling System*. [Online] http://Labelling.i3mainz.hs-mainz.de/client. [Accessed: 11 november 2015].

Leach, P. J., Mealling, M., and Salz, R. 2005. *A Universally Unique Identifier (UUID) URN Namespace*. RFC 4122, 2005. [Online] http://www.ietf.org/rfc/rfc4122. [Accessed: 11 november 2015].

Lindig, S. 2002. *Das Früh- und Mittelneolithikum im Neckarmündungsgebiet*. Universitätsforschungen zur Prähistorischen Archäologie 85. Bonn, Habelt.

Meier-Arendt, W. 1966. *Die bandkeramische Kultur im Untermaingebiet*. Veröffentlichungen des Amtes für Bodendenkmalpflege im Regierungsbezirk Darmstadt, Band 3. Bonn, Habelt.

Miles, A., Bechhofer, S. 2009. *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation, World Wide Web Consortium. [Online] http://www.w3.org/TR/skos-reference. [Accessed: 11 november 2015].

Momtchev V., Peychev D., Primov, T. and Georgiev G. 2009. *Expanding the Pathway and Interaction Knowledge in Linked Life Data*. Proceedings of International Semantic Web Challenge, 2009. Available from: http://challenge.semanticweb.org/documents/Linked%20Life%20Data-LLD%20semantic%20web%20challenge%202009.pdf. [Accessed: 11 november 2015].

Pelagios (n.d.) *About*. [Online] http://pelagios-project.blogspot.de/p/about-pelagios.html. [Accessed: 11 november 2015].

PeriodO (n.d.). *PeriodO*. [Online] http://perio.do/narrative. [Accessed: 11 november 2015].

Picturae 2013. *OpenSKOS: Simple Knowledge Organization System Repository*. [Online] http://openskos.org. [Accessed: 11 november 2015].

Piotrowski, M., Colavizza, G., Thiery, F. and Bruhn, K.-C. 2014. The Labelling 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*: 1–6. New York, ACM.

Pleiades (n.d.). *Pleiades*. [Online] http://pleiades.stoa.org. [Accessed: 29 june 2015].

RGZM (n.d.). *RGZM: Online Databases*. [Online] http://web.rgzm.de/en/research/onlinedatabases.html. [Accessed: 11 november 2015].

Richardson, L., Roby, L. 2007. *RESTful Webservices. Web services for the real world*. Sebastopol, O'Reilly Media.

TemaTres 2015. *TemaTres*. [Online] http://www.vocabularyserver.com. [Accessed: 11 november 2015].

TGN 2011. *Getty Thesaurus of Geographic Names*. [Online] http://getty.edu/research/tools/vocabularies/guidelines#tgn. [Accessed: 11 november 2015].

Thiery, F. 2015a. *Labelling System Ontology. Copeman Edition*. [Online] http://Labelling.i3mainz.hs-mainz.de/ontology. [Accessed: 11 november 2015].

Thiery, F. 2015b. *Labelling System Vocabulary. Edison Edition*. [Online] http://Labelling.i3mainz.hs-mainz.de/vocab. [Accessed: 11 november 2015].

Thiery, F. 2015c. *Labelling System REST-API. Archway Edition*. [Online] http://Labelling.i3mainz.hs-mainz.de/rest-api. [Accessed: 11 november 2015].

Thiery, F. 2015d. *Labelling System SPARQL-API. Aldgate Edition*. [Online] http://Labelling.i3mainz.hs-mainz.de/sparql-api. [Accessed: 11 november 2015].

Usercake 2012. *Home*. [Online] http://usercake.com/index.php. [Accessed: 11 november 2015].

W3C 2013. *SPARQL 1.1 Overview*. [Online] http://www.w3.org/TR/sparql11-overview. [Accessed: 11 november 2015].