Maintaining a Linked Data Cloud and Data Service for Second World War History

Mikko Koho[1 [0000−0002−7373−9338]], Esko Ikkala[1,2 [0000−0002−9571−7260]], Erkki Heino[1], and Eero Hyvönen[1,2 [0000−0003−1695−5840]]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland
[2] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
http://seco.cs.aalto.fi, http://heldig.fi

Abstract. One of the great promises of Linked Data is to provide a shared data infrastructure into which new data can be imported and aligned with, forming a sustainable, ever growing Linked Data Cloud (LDC). This paper studies and evaluates this idea in the context of the WarSampo LDC that provides a data infrastructure for Second World War related ontologies and data in Finland, including several mutually linked graphs, totaling ca 12 million triples. Two data integration case studies are presented, in which the original WarSampo LDC and the related semantic portal were first extended by a dataset of hundreds of war cemeteries and thousands of photographs of them, and then by another dataset of over 4450 Finnish prisoners of war. As a conclusion, lessons learned are explicated, based on hands-on experience in maintaining the WarSampo LDC in a production environment.

1 Introduction

This paper studies the fundamental process of building the Web of Data [6] by incrementally aggregating and aligning new datasets into a Linked Data Cloud (LDC). The focus is in particular on publishing and using Cultural Heritage Linked Data on the Semantic Web [8].

We first overview previous research related to the problem of maintaining ontologies and linked data. Based on this, a typology of change propagation in interlinked Resource Description Network (RDF)[3] graphs is presented. Two practical case studies are discussed where a new dataset is integrated into the WarSampo LDC [9], which contains a dynamic ontology infrastructure and a collection of Linked Open Data about Finland in the Second World War (WW2). In both cases, change propagation scenarios are discussed, with lessons learned explicated. As a conclusion, guidelines for integrating a new dataset into an LDC are outlined.

The main contribution of this paper is to address the linked dataset maintenance problem on an LDC level. The paper contributes also by explaining how the new datasets can be shown to the end user as new application perspectives

[3] https://www.w3.org/RDF/

and through enriching other existing application perspectives with additional data.

WW2 data is of great interest not only to historians, but to potentially hundreds of millions of citizens globally whose relatives participated in the war, creating a global shared trauma. However, data about the WW2 is scattered in various organizations and countries, written in multiple languages, and represented in heterogeneous formats. WarSampo [9] provides a novel infrastructure for publishing WW2 data as LOD. The infrastructure supports integrating new datasets into WarSampo, by extending both the DOs and the MDSs. Published in 2015, WarSampo is to our best knowledge the first large scale system for serving and publishing WW2 LOD on the Web. WarSampo is a part of the global LOD cloud[4], and was awarded with the LODLAM Challenge Open Data Prize in 2017.

The data is served on an open data service[5], which enables anyone to build applications that use the data via standard APIs. The WarSampo semantic portal uses the data service to provide different perspectives to the WW2 LOD as customized web applications. New perspectives can be added in a flexible way to provide views to new data, or to answer new research questions with existing data.

The War Cemetery perspective is an in-use application on the Semantic Web: it was published in November 2017 and got 57 000 users in one week after that. The Prisoners of War perspective will be published later in 2018. In total, the WarSampo data service was used by 130 000 different users through the WarSampo semantic portal[6] in 2017.

2 Related Work

The problem of maintaining ontologies and linked data have been studied extensively, but mostly from a point of view of editing and managing evolving ontologies and data, not on an LDC level as in this paper. Early works on this line of research include, e.g., [11,15]. In [20,3], the problem of managing a set of interlinked hierarchical RDFS thesauri is discussed. Ontology evolution, and the propagation of changes caused by it, has been discussed in [23] and [25].

Umbrich et al. [24] have surveyed solutions to detect, propagate and describe changes in Linked Open Data resources and datasets. Requirements and approaches are studied for different use cases, e.g. link maintenance and vocabulary evolution. These linked data dynamics are explored also in [2,16]. Handling broken links in Linked Data is discussed in [22].

In addition to the global LOD cloud, other LDCs, like the Lexvo [17] and the MIDI LDC [18] have been previously studied.

A framework for integrating heterogeneous OpenCourseWare data repositories into a Linked Data publication is presented in [21]. A framework and tool for

[4] http://linkeddata.org
[5] http://www.ldf.fi/dataset/warsa
[6] https://sotasampo.fi/en/

data fusion, conflict resolution, and quality assessment of Linked Data graphs is presented in [19]. Knoblock et al. [12] presented lessons learned in integrating heterogeneous data from 14 museums into a Linked Data publication, harmonizing data with CIDOC CRM[7].

An overview of the WarSampo data service and semantic portal has been presented in [9]. A core dataset of WarSampo, the casualties of war, and its application in digital humanities research is presented in [14]. Using the war cemetery data in prosopographical research is discussed in [10]. Overview of the Prisoners of War case study with preliminary results have been published [13], with a comparison of different online publishing approaches. Named entity linking in WarSampo was studied in [7]. This paper provides a new view to this line of research from an LDC management point of view.

3 Anatomy and Maintenance of Linked Data Clouds

An LDC consists of a set of graphs. Data is interlinked across graphs by mappings and direct references to URIs in other graphs. We differentiate the graphs into two major categories based on their usage: *metadatasets* (MDS) and *domain ontologies* (DO). MDSs describe objects or other things in an application domain in terms of a metadata schema [4], such as Dublin Core or CIDOC CRM. Collection metadata in libraries, museums, and archives, or their harmonized aggregated versions are typical examples of MDSs. DOs define the basic concepts used in populating the MDSs and are shared by them. DOs include, e.g., ontologies for subject matter concepts (keyword thesauri), places, people, times, and events. The generic, domain independent structure and semantics of DOs and MDSs are defined by a set of shared domain independent vocabularies, such as RDF(S), SKOS, and OWL. Data linking in an LDC is based on making references to shared domain independent vocabularies, domain specific DOs, and mappings.

We call a set of DOs used for populating a set of MDSs in an application domain the *ontology infrastructure*. In many cases, DOs, MDSs, mappings, and domain independent vocabularies are published as one homogeneous triple mass. If there is no separation of DOs and MDSs into graphs, the distinction between them can be vague. An example of this is DBpedia[8], in which resources are separated by namespaces, but this distinction is insufficient, since typically one graph can use a variety of different namespaces. A key observation underlying this paper is that from a data management point of view, DOs, MDSs, and mappings are different from each other, and it makes sense to keep them separate in order to support different kind of maintenance operations.

An important property of a graph is *independence*: we define a graph independent if it does not make a reference to (i.e. links to) resources in other graphs. For example, SKOS keyword thesauri are often independent DOs mak-

[7] `http://cidoc-crm.org`

[8] `http://dbpedia.org`

ing only `skos:broader/ narrower/related` references to concepts within the same concept scheme.

A Typology of Change Propagation A graph can change through changes in its resources. The following three change types are the most fundamental: 1) *Addition*. A new resource is added into the graph. 2) *Modification*. A resource is modified in terms of its properties. 3) *Removal*. A resource is removed from the graph. Based on the primitive changes, more complex changes can be modeled as sequences of more primitive ones, such as moving a resource from a graph into another. The primitive changes may occur in a DO or an MDS, and may have an effect in related DOs or MDSs [23]. We have identified the following four principal cases of change propagation needs between graph types. Here the notation $X \rightarrow\ Y$ means that a change in a graph X creates a potential need for a change in a graph Y that makes a reference to X.

1. **DO→MDS**. In all cases, linkage based on probabilistic entity linking, from an MDS to the DO, needs to be revalidated. **Addition:** An addition in the DO usually doesn't create a need for change propagation to MDSs. However, when a new DO resource is introduced in an MDS, the linkage from the MDS to the DO is broken since the new resource is not there in the DO. **Modification:** no additional effect. **Removal:** The MDSs can get corrupted by having URI references to removed URIs. The affected MDSs need to be fixed.
2. **DO→DO**. If the changed DO is independent, there are no change propagation needs. Otherwise change propagation is needed as in case DO→MDS.
3. **MDS→DO**. **Addition:** If DOs cover the values used by the MDS, there is no effect. Otherwise the DOs may need to be updated accordingly. **Modification:** usually no effect. If a new value not in a DO would be needed as a property value in the MDS, the DO may need to be updated accordingly. **Removal:** no effect, unless a DO makes a reference to the MDS. This may happen, e.g., when an event ontology makes a reference to an artifact collection database.
4. **MDS→MDS**. Changes between MDSs are propagated as in MDS→DO.

Practical examples of the change propagation scenarios are presented in the use cases in Sections 5 and 6.

4 Maintaining the WarSampo Linked Data Cloud

Creating the WarSampo ontology infrastructure has been a dynamic process, involving several people working with up to seven datasets at the same time. The metadatasets and domain ontologies have been constantly evolving, which often causes existing entity matching to be invalidated.

Fig. 1 shows the main MDSs and DOs of WarSampo, after the data model changes caused by the case studies presented in this paper. Each MDS and DO shows the number of individual entities belonging to the corresponding class(es). The arrows depict the direction of linkage, which is normally from the MDSs

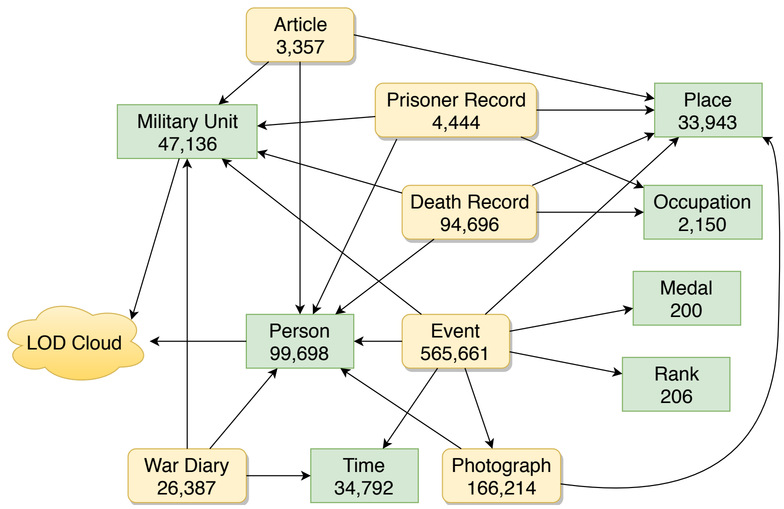

Fig. 1. The main metadatasets (yellow, rounded corners) and domain ontologies (green rectangular boxes) of WarSampo.

towards the DOs that have been used in annotating the entities. There is also linkage to the global LOD cloud.

The WarSampo LDC is centrally maintained, even if it is based on data from distributed sources. This means that DOs can adapt to changes that are needed when integrating new datasets into the LDC, and should do so to better represent their domain. The ontology infrastructure is extended as needed.

Maintaining the WarSampo LDC is different from maintaining the global LOD cloud, where `owl:sameAs` and related mappings between datasets are created, but changes are seldom propagated across the datasets. This is not feasible in our case, since a new piece of information in one graph of the service may require changes in other graphs, too. For example, if a new place or a person is introduced in a new or existing MDS, the Place DO or the Person DO has to be extended before the new data can be aligned.

Integrating data into a dynamic LD environment is challenging. As a DO becomes more complete, covering its domain more accurately, MDSs using that DO may need to redo their entity linking process, to get more accurate linkage. Failing to do so can cause structurally or semantically erroneous annotations [22] to be used. WarSampo employs plenty of probabilistic entity linking, e.g., to the Person, Place and Military Unit DOs, in which the DO is not expected to cover all of the information about its domain. The usual case of change propagation in the WarSampo context relates to the invalidation of the entity linking.

Because of the complex change propagations in the dynamic LDC, maintaining the DOs and MDSs directly in RDF format is too laborious and error-prone to implement in practice, especially in a way that a domain expert with little Linked Data expertise could make changes. This is especially true in the case of person instances as they are linked, directly or indirectly, to everything in WarSampo. Modeling even just the basic information of a person entails, e.g., multiple events, such as birth, death, promotions, an so on. It was decided that the domain experts directly maintain the datasets in their native formats (usually spreadsheet files), which can then be easily transformed and integrated into WarSampo, as needed.

5 Case 1: War Cemeteries

In our first case study, a war cemetery dataset was produced and integrated into the WarSampo LDC [10]. Since Finnish soldiers who perished in WW2 were transported back to their hometown for burial whenever it was possible, the local cemetery is a natural starting point for studying the common characteristics and events of the residents of one's hometown in the turmoil of the war.

Starting Point & Source Data A complete listing of war cemeteries in Finland was not available, but the Casualties MDS, that was previously integrated into WarSampo, includes the name of the cemetery and/or the municipality in which the person is buried. However, the lack of uniform naming conventions and missing coordinates of the cemeteries made it difficult to locate them and to specify the people buried there.

In 2016–2017 the Memorial Foundation of the Fallen and the Central Organization of Finnish Camera Clubs (Suomen Kameraseurojen Liitto ry, SKsL) carried out a project called "War Cemeteries in Finland". Its goal was to locate, photograph, and collect data about all known war cemeteries in Finland. In total 615 war cemeteries were found, accompanied by 2500 photographs.

Workflow A representative of the SKsL manually harmonized the data entry sheets and filenames of the photographs sent by the camera clubs, and organized them into one table. Finally the table was converted into WarSampo compatible RDF by using a Python data processing pipeline[9], which 1) handles the matching of existing cemetery names found in the WarSampo death records to cemetery names in the source table, 2) creates new URIs for cemeteries not found in WarSampo, and 3) creates photograph and photography instances according to the WarSampo data model. Whenever there is a need to update the cemetery data, the source table can be edited and the data processing pipeline can be run again to produce new RDF files for WarSampo.

To avoid errors in the data integration, the "War Cemeteries in Finland" project was instructed to start with the same cemetery name listing that was used when the death records were collected into a database. A challenge here was that some of the cemeteries mentioned in the Casualties MDS were unambiguous.

The structure of the project's output table was agreed on beforehand, so that information about one individual cemetery was gathered in one row, with values separated on columns, easing the RDF conversion process. The cemetery data processing pipeline was run multiple times in order to enhance data quality, and

[9] `https://github.com/SemanticComputing/cemeteries-csv2rdf`

a listing of spelling errors, missing photograph files, etc. was sent back to SKsL for making manual corrections to their table.

The cemetery data was integrated into the Place DO, and the cemetery photograph data into the Photograph MDS. The photographs are generally linked to the photography places via a photography event, which has created the photograph. However, photographs of war cemeteries represent the cemeteries, which are modeled as a subclass of the place class. Some military units and people are mapped to entities in the global LOD Cloud, i.e., Wikidata and DBpedia.

Change Propagation With regard to maintenance, the basic data about the cemeteries remains independent in the Place DO. If the cemeteries in Place DO change, the linkage from the death records in the Casualties MDS need to adjust for the change as according to the DO→MDS case in Section 3.

However, the information about the people buried in a cemetery is stored in the Casualties MDS which makes references to the Place DO. Thus, the changes in Casualties or Photograph MDS related to the cemeteries must be propagated to the Place DO according to the MDS→DO case in Section 3.

Semantic Portal Changes The new War Cemeteries Perspective[10] showcases how the integration of cemetery data enriches the existing WarSampo data and vice versa. The perspective has been developed to gain new insights from the casualties data based on the community-level aspect provided by the cemeteries. This approach is useful, because there is not enough data about the casualties to construct detailed life stories of individual soldiers as biographies, but the amount of individuals is large enough to study the data as groups of people using, for example, visualizations.

The user interface of the Cemetery perspective is presented in Fig. 2. The user can browse all cemeteries, or search the cemeteries by name and narrow the results by using the filters on the left. The results can be viewed as a table with basic information about the cemeteries, or on a map which provides a global view of the cemeteries.

A concrete example of the data integration results can be seen in Fig. 2, where the "Number of graves" column is based on the data of the "War Cemeteries in Finland" project, whereas the "Buried people" column shows the total number of death records (collected in the 1980s) that make a reference to the cemetery. The numbers are equal only with 27 % of the cemeteries although ideally they should be equal with every cemetery. This gives valuable insight to the data providers to set the records straight.

When the user clicks the name of a cemetery, an information page opens, showing basic information, photographs, and various visualizations based on the property values of the buried people.

[10] `https://www.sotasampo.fi/en/cemeteries/`

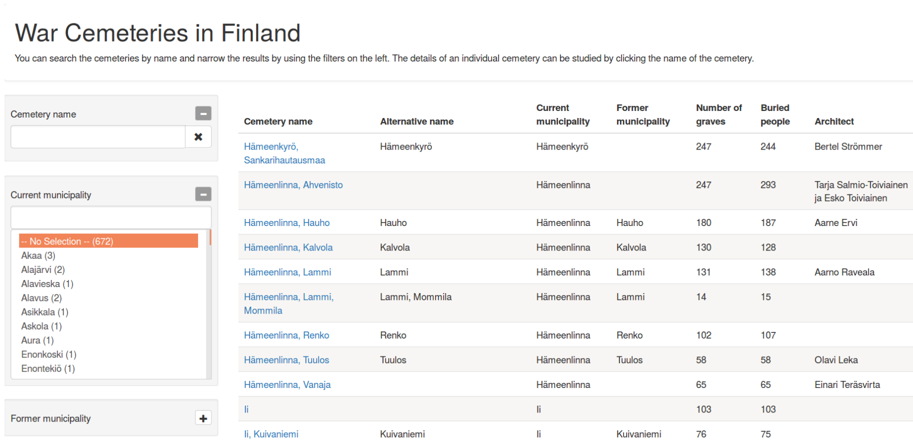

Fig. 2. Cemetery search in the WarSampo cemetery perspective.

6 Case 2: Prisoners of War

Some 4450 Finnish soldiers were captured as prisoners of war (POW) in WW2 by the Soviet Union. This case study concerns integrating the POW data into the WarSampo LDC.

Starting Point & Source Data The POW dataset was originally published in a book [1]. Recently, the dataset has been extensively extended, cleaned, and validated by domain experts. A collaboration was set up to publish the data as part of the WarSampo, which was chosen as the primary data publication platform by the stakeholders, which include the National Archives of Finland, and the Association for Cherishing the Memory of the Dead of the War.

The core of the dataset is a register of the Finnish prisoners of war in WW2. The register is formatted as a spreadsheet file, with additional spreadsheet files presenting data about POW camps and hospitals, as well as the primary data sources. The POW dataset contains sensitive information about the individual soldiers, some of whom are still alive. There is an ongoing process to evaluate what information can be published, by the legal experts at the National Archives of Finland. The data will be published in the autumn 2018, at which point the privacy issues should be resolved.

Workflow The data formatting evolved as a collaboration between the domain experts maintaining the original dataset, and the WarSampo team of Linked Data experts. A data processing pipeline was created[11], that handles data transformation, validation, linking, and harmonization. The pipeline transforms the spreadsheets into RDF, mapping the spreadsheet columns to RDF properties, with possibly multiple values per property, and containing annotations for primary information sources. Automatic linking processes then link the

[11] `https://github.com/SemanticComputing/WarPrisoners`

records to WarSampo DOs of military ranks, units, occupations, people, and places.

The prisoner records were modeled in a way similar to the previously published Casualties MDS [14], and they share common super classes and properties. However, the process workflow was different: the casualty data was received as a static data dump, whereas the POW dataset was constantly evolving during the project.

The original POW register is maintained in spreadsheet format, which can be easily integrated into WarSampo with our automated transformation process when the spreadsheet is updated, provided that the structure stays the same.

For most of the original data, the spreadsheet format is a natural way to represent the information, with each row of the POW register expressing information about one individual soldier, and each column representing a different property of a soldier, like his name, occupation, and date of capture.

As the data comes from multiple sources that can have contradictory information, there is a need to collect all different values for a single property, along with references to the primary data sources. For this purpose, a special cell data format is used that enables to present multiple values and source references in the spreadsheet. The cell formatting is validated during the data transformation process. Also other simple data validation rules are applied to find anomalies during data conversions.

Change Propagation The POW data introduces the main MDS of POWs, and a DO of war-time occupations. The WarSampo person DO is updated with about 3,000 new person instances. POW camps and hospitals are modeled as part of the Place DO.

The original dataset contains source references for separate pieces of information, which are used in the RDF data model by employing RDF reification for the prisoner records. This is a standard approach to modeling this kind of provenance information on an RDF triple level.

The DOs of military ranks, military units, places (e.g. municipalities, camps and hospitals), occupations and persons provide values for populating the POW MDS. Their linking uses probabilistic entity linking, while also original values are stored as literals. All changes in the DOs would require repeating the corresponding entity linking process as according to the case DO→MDS in Section 3. I.e. if a new understanding about the historical war-time Occupation DO (cf. Fig. 1) cause two occupations to be merged into one, resulting in the removal of the obsolete one, any linking to the obsolete resources need to be updated.

Adding a new property value in the MDS can propagate the change to related DOs, if the value doesn't exist there (cf. Section 3, case MDS→DO). For example, the new value could be a new military rank or a new occupation. When a new POW record is added to the registry, the changes will propagate to the Person DO, either through the linking to an existing person, in which case the person instance is enriched, or through the creation of a new person instance.

The POW records are mapped to the Person DO using probabilistic record linkage [5], where each POW's information is compared with the information in the WarSampo person instances to find matches that have high enough similarity. As the record linkage needs to be able to adapt to changing input dataset, as well as to the changes in the Person DO, a machine learning approach was used, which employs logistic regression based on weighted comparisons of a set of predefined attributes. The weights are calculated based on training data, which is initially acquired from a previous, simpler record linkage implementation, based on manually defined fuzzy matching, and updated manually during linkage iterations. With the machine learning approach, the entity linking process automatically adapts to changes in the POW MDS and Person DO. The linking process needs to be redone when the POW MDS changes.

New person instances are then created for the unlinked POW records and added into the Person DO. With the probabilistic linkage, it is possible that a record is not mapped because there is not enough information about either the POW record, or the person instance, to create a mapping between them. Modifying the information in either the MDS or the DO means that the whole record linkage process should be redone.

Semantic Portal Changes A new application perspective has been added to WarSampo to explore, analyze, and visualize the information contained in the POW metadataset. The perspective is similar to the earlier casualties perspective, which is used to show information from the death records to the user.

In addition, integrating the prisoners of war data into WarSampo has caused several necessary changes to other parts of the semantic portal. Allowing multiple values for properties with provenance data changes how the information can be presented in a person's home page and how to visualize the data. People's home pages in WarSampo were updated to show information combined from multiple sources (death records, prisoner records, Wikipedia) with source information next to each piece of information.

7 Conclusions

A key lesson learned in our work is that one should make all data transformations and linking into **repeatable, automated processes** to be able to handle change propagation automatically. In the early stages of building WarSampo, the importance of this was not obvious, and for some early WarSampo datasets, the transformation processes were never completely automated. Automating them now would require considerable effort because the datasets have gone through undocumented processes that are not easily repeatable.

The transformation processes should be built using a modular structure, to make the processes **maintainable**, and to enable the reuse of code for other data integration. In a dynamic LDC, the entity linking processes need to be able to **adapt** to common changes in all of the graphs.

Maintenance of an LDC using a complex data model, such as CIDOC CRM, is difficult natively in RDF format. For complex DOs and MDSs, it is easier to update the data in simpler formats, such as Dublin Core, and maintain the transformation processes that build the graphs of the LDC. The complexity of the transformation processes grows as they need to handle the creation or updating of missing or uncertain resources in incomplete DOs shared by multiple MDSs. Simple, independent DOs (e.g. military units, occupations) can be maintained directly in RDF format, whereas more complex DOs like Persons require a different approach.

DOs differ from each other by nature. For example, covering and disambiguating all military ranks is clearly a simpler task than performing the same task with all wartime places. In general, it is not realistic to assume that the DOs completely cover their domain.

Integrating data into a LDC is more laborious than simpler ways of publishing the data in independent data silos. However, the result is an interlinked knowledge base, a Linked Data Cloud, where the interlinked graphs enrich each other, creating a whole that is greater than the sum of its parts.

Acknowledgements Our work was funded by the Association for Cherishing the Memory of the Dead of the War, the Memory Foundation for the Fallen, the Finnish Ministry of Education and Culture, and the Academy of Finland. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

References

1. Alava, T., Frolov, D., Nikkilä, R.: Rukiver. Suomalaiset sotavangit Neuvostoliitossa. Helsinki: Edita (2003)
2. Auer, S., Dalamagas, T., Parkinson, H., Bancilhon, et al.: Diachronic linked data: towards long-term preservation of structured interrelated information. In: Proceedings of the First International Workshop on Open Data. pp. 31–39. ACM (2012)
3. Frosterus, M., Tuominen, J., Pessala, S., Hyvönen, E.: Linked Open Ontology cloud: managing a system of interlinked cross-domain light-weight ontologies. International Journal of Metadata, Semantics and Ontologies **10**(3), 189–201 (2015)
4. Gartner, R.: Metadata. Shaping Knowledge from Antiquity to the Semantic Web. Springer–Verlag (2016)
5. Gu, L., Baxter, R., Vickers, D., Rainsford, C.: Record linkage: Current practice and future directions. CSIRO Mathematical and Information Sciences Technical Report **3**, 83 (2003)
6. Heath, T., Bizer, C.: Linked Data: Evolving the web into a global data space. Synthesis Lectures on The Semantic Web: Theory and Technology, Morgan & Claypool Publishers, Palo Alto, USA (2011)
7. Heino, E., Tamper, M., Mäkelä, E., Leskinen, P., Ikkala, E., Tuominen, J., Koho, M., Hyvönen, E.: Named Entity Linking in a Complex Domain: Case Second World War History. In: Proceedings, Language, Technology and Knowledge (LDK 2017). pp. 120–133. Springer-Verlag (June 2017)
8. Hyvönen, E.: Publishing and Using Cultural Heritage Linked Data on the Semantic Web. Synthesis Lectures on the Semantic Web: Theory and Technology, Morgan & Claypool, Palo Alto, USA (2012)
9. Hyvönen, E., Heino, E., Leskinen, P., Ikkala, E., Koho, M., Tamper, M., Tuominen, J., Mäkelä, E.: WarSampo data service and semantic portal for publishing linked open data about the second world war history. In: The Semantic Web – Latest Advances and New Domains (ESWC 2016). pp. 758–773. Springer-Verlag (2016)

10. Ikkala, E., Koho, M., Heino, E., Leskinen, P., Hyvönen, E., Ahoranta, T.: Prosopographical Views to Finnish WW2 Casualties Through Cemeteries and Linked Open Data. In: Proceedings of the Workshop on Humanities in the Semantic Web (WHiSe II). CEUR Workshop Proceedings (October 2017)
11. Klein, M.: Change Management for Distributed Ontologies. Ph.D. thesis, Free University, Amsterdam (2004)
12. Knoblock, C.A., Szekely, P., Fink, E., Degler, D., Newbury, D., Sanderson, R., Blanch, K., Snyder, S., Chheda, N., Jain, N., et al.: Lessons Learned in Building Linked Data for the American Art Collaborative. In: International Semantic Web Conference. pp. 263–279. Springer (2017)
13. Koho, M., Heino, E., Ikkala, E., Hyvönen, E., Nikkilä, R., Moilanen, T., Miettinen, K., Suominen, P.: Integrating Prisoners of War Dataset into the WarSampo Linked Data Infrastructure. In: Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018). CEUR Workshop Proceedings (March 2018), `http://www.ceur-ws.org/Vol-2084`, vol 2084
14. Koho, M., Hyvönen, E., Heino, E., Tuominen, J., Leskinen, P., Mäkelä, E.: Linked Death - Representing, Publishing, and Using Second World War Death Records as Linked Open Data. In: The Semantic Web: ESWC 2017 Satellite Events. pp. 369–383. Springer-Verlag (2017), `https://doi.org/10.1007/978-3-319-70407-4_45`
15. Maedche, A., Motik, B., Stojanovic, L., Studer, R., Volz, R.: An infrastructure for searching, reusing and evolving distributed ontologies. In: Proc. of the twelfth international conference on World Wide Web. pp. 439–448. ACM Press (2003)
16. Meimaris, M., Papastefanatos, G., Pateritsas, et al.: Towards a Framework for Managing Evolving Information Resources on the Data Web. In: Proceedings of the 1st International Workshop on Dataset PROFIling & fEderated Search for Linked Data. CEUR Workshop Proceedings (March 2014), vol 1151
17. de Melo, G.: Lexvo. org: Language-related information for the linguistic linked data cloud. Semantic Web **6**(4), 393–400 (2015)
18. Meroño-Peñuela, A., Hoekstra, R., Gangemi, A., Bloem, P., de Valk, R., Stringer, B., Janssen, B., de Boer, V., Allik, A., Schlobach, S., et al.: The MIDI linked data cloud. In: International Semantic Web Conference. pp. 156–164. Springer (2017)
19. Michelfeit, J., Knap, T., Nečaskỳ, M.: Linked data integration with conflicts. arXiv preprint arXiv:1410.7990 (2014)
20. Pessala, S., Seppälä, K., Suominen, O., Frosterus, M., Tuominen, J., Hyvönen, E.: MUTU: An Analysis Tool for Maintaining a System of Hierarchically Linked Ontologies. In: ISWC 2011 - Ontologies come of Age Workshop (OCAS). CEUR Workshop Proceedings (2011), vol 809
21. Piedra, N., Tovar, E., Colomo-Palacios, R., Lopez-Vargas, J., Alexandra Chicaiza, J.: Consuming and producing linked open data: the case of Opencourseware. Program **48**(1), 16–40 (2014)
22. Popitsch, N.P., Haslhofer, B.: Dsnotify: handling broken links in the web of data. In: Proceedings of the 19th international conference on World wide web. pp. 761–770. ACM (2010)
23. Stojanovic, L., Maedche, A., Motik, B., Stojanovic, N.: User-driven ontology evolution management. In: International Conference on Knowledge Engineering and Knowledge Management. pp. 285–300. Springer (2002)
24. Umbrich, J., Villazón-Terrazas, B., Hausenblas, M.: Dataset dynamics compendium: A comparative study (2010)
25. Zablith, F., Antoniou, G., d'Aquin, M., Flouris, G., Kondylakis, H., Motta, E., Plexousakis, D., Sabou, M.: Ontology evolution: a process-centric survey. The knowledge engineering review **30**(1), 45–75 (2015)