

# World War 1 as Linked Open Data

Eetu Mäkelä<sup>a,\*</sup>, Juha Törnroos<sup>a</sup>, Thea Lindquist<sup>b</sup>, and Eero Hyvönen<sup>a</sup>

<sup>a</sup> *Semantic Computing Research Group (SeCo), Aalto University and University of Helsinki, Finland*

<sup>b</sup> *University of Colorado Boulder, USA*

**Abstract.** The WW1LOD dataset is primarily a reference dataset meant to link collections dealing with the First World War. For this purpose, the dataset gathers events, places, and agents related to the war from various authoritative sources. Additional information on the entities is also recorded, in order to be able to answer more complex questions relating to them. The approach is being evaluated using an existing WW1 online collection. In addition, discussions are ongoing with several other organizations about making use of the dataset.

Keywords: linked data, historical data, digital humanities, modeling, dataset

## 1. Introduction

Historical interest in the First World War (WW1) is on the rise due to the recent centenary of its outbreak (2014–2018). Several projects have been initiated for publishing related content on the web, such as Europeana Collections 1914–1918<sup>1</sup>, 1914–1918 Online<sup>2</sup>, WW1 Discovery<sup>3</sup>, Out of the Trenches<sup>4</sup>, CENDARI<sup>5</sup>, and Muninn<sup>6</sup>.

The WW1LOD dataset presented in this paper is a Linked Data reference dataset intended to link disparate historical collections and data publications related to WW1. The dataset is additionally structured to provide common points of entry into such collections, resulting in improved access to and context for them. As a case study, the dataset has been applied to a collection of WW1 primary-source documents.

In the following article, we first describe the reasons for creating the dataset and major technological choices made. After this, the dataset is presented and access to it is described. Then, the dataset is contextu-

alized with related work and projects, after which usage examples are given. In conclusion, contributions of this work and directions for future research are discussed.

## 2. Reasons for Creating the Dataset

The impetus for creating a Linked Data reference dataset sprang from user needs research undertaken at the University of Colorado Boulder (CU) to improve users' experience in working with digitized collections of historical primary sources (i.e., documents contemporary to or reported by those who experienced historical events).

To better understand the problems humanities researchers encounter in utilizing these collections, researchers conducted 21 semi-structured interviews with CU faculty and students [12]. The major user needs identified were better support for: 1) locating documents and data relevant to a particular topic within distributed online collections; and 2) contextualizing the content, for instance, to gauge author bias or simply become familiar with the places, events, and people mentioned in the documents.

In addition, problems were identified with documents being in unfamiliar languages, as well as with ambiguities and variations in names, such as place

---

\* Corresponding author. E-mail: eetu.makela@aalto.fi

<sup>1</sup><http://www.europeana-collections-1914-1918.eu/>

<sup>2</sup><http://www.1914-1918-online.net/>

<sup>3</sup><http://ww1.discovery.ac.uk/>

<sup>4</sup><http://www.canadiana.ca/en/pcdhn-lod>

<sup>5</sup><http://www.cendari.eu/research/>

[first-world-war-studies/](http://www.first-world-war-studies.org/)

<sup>6</sup><http://blog.muninn-project.org/>

names changing over time or a person being referenced by different names in different documents.

To respond to such needs, reference vocabularies are needed that go far beyond the broad and often minimal subject headings currently commonly used in library cataloguing. In thinking about ways to interlink online historical collections, the project chose to follow the path highlighted by the ISO standard CIDOC-CRM [5], which has become widely accepted by cultural heritage institutions as a basis for integrating sources.

The core idea of CIDOC-CRM is to link collection items to their real-world context through events they reference. These events, and the people, places, and timeframes related to them then provide a contextual framework that links the items. This approach seemed to fit well with primary sources, as their value is precisely in how they view and relate to the historical events they describe.

However, CIDOC-CRM only indicates that common events, actors (e.g., people and organizations), places, and timeframes are important and provides a framework for how they can be described. To achieve interoperability between different datasets, these entities' identifiers still need to be shared. The real work, then, is in creating suitable reference vocabularies from which to source those identifiers, for example, for individual battles, historical places, and people and army units involved.

Further benefits are derived if these reference vocabularies are also themselves richly interlinked, allowing for inferencing and navigation among the actors, places, and events identified.

To test these ideas, a focused case study was needed. For concrete evaluation, a newly available collection of WW1 primary-source documents at CU<sup>7</sup> [11] was selected, in part due to the aforementioned interest in the WW1 domain.

In creating the reference dataset, much thought was given to incorporating authoritative, high-quality sources of information. This decision stems from needs expressed by the intended users of the system. For example, while Wikipedia contains numerous articles related to military history, many historians remain sceptical of it as a resource for either themselves or their students [20]. Indeed, while Wikipedia seems reassuringly accurate compared to traditional reference

sources in the field of science [7], there is evidence to support their scepticism in the historical domain. In a field in which scholarly accountability is a core tenet, historians' questions largely focus on Wikipedia editors' lack of accountability, the unevenness of the quality of its articles, and particularly for historical articles, a lack of references to academic sources. [13,17]

In light of this belief, the WWI LOD project sought to incorporate data from as many sources as possible that historians would regard as authoritative and quality-controlled. On the other hand, in order to become a part of and integrate with the established Linked Open Data (LOD) cloud, work has also gone into creating equivalency relationships between the project's data sources and relevant parts of the cloud.

### 3. Dataset Description

The main object types found in the WWI LOD dataset are described in Table 1 along with their instance counts, while the core data model is depicted in abstract form in Figure 1. As stated, the actual data is encoded using the much more complex CIDOC-CRM RDFS schema<sup>8</sup>.

Table 1  
Core classes in the dataset

Type	Nr.	Example
Event	689	"Battle of the Aisne, 1914"
Event type	36	"Naval Operations"
Place	1366	"Aisne (River)"
Actor	651	"13th Cavalry Brigade"
Time	461	"02/01/1917 - 06/09/1918"
Keyword	25	"Prussian Militarism"
Theme	5	"Naval history"

The instances included in the dataset come from a variety of sources, as detailed in Table 2. First, to provide a useful common base for linking among datasets, general events spanning the war years were included. For this, an authoritative framework of 326 top-level wartime events was provided by the Imperial War Museum (IWM)'s First World War Centenary Partner-

<sup>7</sup>The collection is available at <http://libcudl.colorado.edu/wwi/index.asp>

<sup>8</sup>Available at [http://www.cidoc-crm.org/rdfs/cidoc\\_crm\\_v5.0.4\\_official\\_release.rdfs](http://www.cidoc-crm.org/rdfs/cidoc_crm_v5.0.4_official_release.rdfs). More detailed information on the schema and dataset content can be found on the WWI LOD dataset homepage at <http://ldf.fi/ww1lod/>

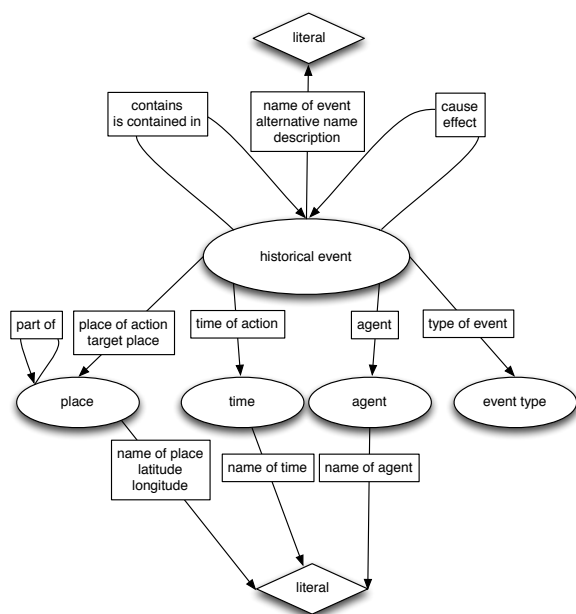


Fig. 1. The core data model of the dataset

ship<sup>9</sup>. The IWM is considered the principal cultural heritage institution in the English-speaking world relating to the war. Thus both historians and cultural heritage professionals should consider the sourced vocabulary authoritative, and it is likely to be re-used by others who are preparing datasets in this domain.

The IWM events, however, did not contain place or actor information. To overcome this limitation, a separate catalogue of some 250 events selected by domain experts was built for richer description, including annotation of places, participating actors, and temporal relationships. The information is drawn from various sources, including approved terminologies from the IWM and the British Army's Battle Nomenclatures Committee, as well as a custom term list on Belgium and WW1 created expressly for this purpose<sup>10</sup>.

These events were manually linked to the top-level events where appropriate, resulting in 46 *owl:sameAs* links. In addition, all events have been automatically linked to DBpedia [10], with a little over 100 *owl:sameAs* relationships. The latter links were vali-

<sup>9</sup>This timeline was principally derived from the official British series on the history of the war, the *History of the Great War Based on Official Documents*, particularly the volume *Principal Events, 1914–1918* [16]

<sup>10</sup>Derived in part from Patrick Lefevre's standard bibliography on this topic published by Belgium's Musée Royal d'Armée [6]

Table 2  
Dataset content

<b>Top-level events (326)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/iwm/">http://ldf.fi/ww1lod/iwm/</a>
Properties:	Name of event, Time of action, Description, Theme
Source:	IWM First World War Centenary Partnership
<b>Rich events (253)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/main/">http://ldf.fi/ww1lod/main/</a>
Properties:	Name, Alternate names, Description, Agent, Time of action, Place of action, Is contained in, Contains, Cause, Effect
Source:	IWM First World War event list [1], CU Libraries
<b>Atrocity events in Belgium (101)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/atr/">http://ldf.fi/ww1lod/atr/</a>
Properties:	Name, Agent, Time of action, Place of action, Combat related, Deportations, Human shields used, Panic, Destroyed buildings, Killings
Source:	[8]
<b>German army structure (473)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/iga/">http://ldf.fi/ww1lod/iga/</a>
Properties:	Name, Unit type, Part of
Source:	[18], [2]
<b>Other actors (181)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/main/">http://ldf.fi/ww1lod/main/</a>
Properties:	Name, Alternate names, Organizational information, Relationship information
Source:	CU Libraries
<b>Geography of Belgium and France (1312)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/main/">http://ldf.fi/ww1lod/main/</a>
Properties:	Name of region, Part of, Coordinates
Source:	IWM Western Front geographic keywords, GeoNames, CU Libraries
<b>Belgian statistical data 1914–1918 (12)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/bsta/">http://ldf.fi/ww1lod/bsta/</a>
Properties:	Population of both genders
Source:	[3]
<b>Polygons of Belgian provinces 1914–1918 (56)</b>	
Graph:	<a href="http://ldf.fi/ww1lod/bpol/">http://ldf.fi/ww1lod/bpol/</a>
Properties:	Name of region, Part of, Polygon
Source:	HISSTAT (Universities of Ghent, Brussels, Louvain-la-Neuve, and State Archives of Belgium)

dated by domain experts using the Silk link discovery framework [19].

Finally, details on the “German atrocities” that took place in Belgium in 1914 were included. This information was based on a vast study by John Horne and Alan Kramer [8], considered the standard work on the topic. Included data, for example, are the place and date of each incident, the involved army unit, the number of killings, and the number of destroyed buildings.

The focus on Belgium apparent in this source was based on awareness that adequately responding to particular research questions would require more detailed context than available in general thesauri and registries. In analyzing CU’s document collection, a rich pocket of material was identified related to the German occupation of Belgium, and particularly the violence perpetrated by German soldiers against Belgian civilians. Thus, to demonstrate how a more specialized context such as this one can fit into the general framework and be utilized, this particular subtopic was selected as a focus and sources related to it were particularly emphasized when building the dataset.

### 3.1. Actors

Actor information in the dataset comes primarily from two sources. First, details on the structure of the Imperial German Army were included. This information primarily comes from Georg Tessin’s *Deutsche Verbände und Truppe* [18], which is recognized in the field as a standard, trusted reference source for this information. The reason for inputting the German army in particular was that this allowed linking the units mentioned in the atrocity data to their organizational hierarchy for further contextualization and reasoning.

Additionally, actor information has been input by CU domain specialists in conjunction with enriching the event network. During this work, actors have not only been linked to events, but also to each other and the organizations to which they belong.

In expressing organizational and relationship information, in addition to CIDOC-CRM, the project makes use of multiple generally accepted vocabularies for such data: the W3C Organization Ontology<sup>11</sup>, the RELATIONSHIP ontology<sup>12</sup>, the FOAF schema<sup>13</sup>, and the schema.org vocabulary<sup>14</sup>.

<sup>11</sup><http://www.w3.org/TR/vocab-org/>

<sup>12</sup><http://vocab.org/relationship/>

<sup>13</sup><http://xmlns.com/foaf/spec/>

<sup>14</sup><http://schema.org/>

### 3.2. Historical Places

Regarding the locations of wartime events, a key point is that places are temporal instances, that is, they can change over time. Current geographical datasets such as GeoNames are thus not directly applicable to a historical case like WWI, as the documents and events often refer to the place names used at the time. Additionally, for example the locations of battles often refer to geographical features such as ridges or rivers, as opposed to for instance villages or administrative units of the same name found more commonly in modern datasets.

As a core source of contemporary place name data, locations related to the war were gathered from the IWM’s WWI geographical keyword vocabularies. These vocabularies also contain a partonomy structure for the locations, which was brought into the dataset. For example, the dataset contains that Pommeroeul (village) *is part of* Hainaut (municipality) *is part of* Belgium (country). Places missing from these vocabularies were annotated manually during the event and actor enrichment process.

Coordinate information for these places was sourced through two means. First, 1248 contemporary place names were automatically determined to directly match their modern equivalents in GeoNames based on identical names and hierarchy information. These links, encoded as *skos:closeMatch* triples, were then evaluated by domain experts for accuracy.

Second, wartime boundaries of Belgian provinces were obtained from HISSTAT, a collaborative project of the Universities of Ghent, Brussels, and Louvain-la-Neuve, and the State Archives of Belgium. Their geographies are highly accurate, penetrating to the municipal level. This provincial coverage, in combination with information from GeoNames and other sources, will allow the placing of items in their geographical context using boundaries accurately defined for the war years. To encode the border polygons, the GeoRSS vocabulary<sup>15</sup> was used.

### 3.3. Time

Historical context and uncertainty also creates complexity for temporal modeling, as it is often difficult to say with absolute precision when a certain event took place. For example, actors on opposing sides of a battle

<sup>15</sup>[http://www.georss.org/rdf\\_rss1.html](http://www.georss.org/rdf_rss1.html)

tend to see the extent of that battle differently [16, explanatory notes], depending on if they include preparatory attacks leading up to the battle or the consolidation of gains following a successful attack in the time-frame.

For this reason, CIDOC-CRM's temporal representation supports a level of uncertainty in the encoding. In other words, this means that it is possible to present a timestamp, for instance, for "at the beginning of the year 1917" by specifying four temporal points: the earliest possible start time, latest possible start time, earliest possible end time and latest possible end time. By using such timeframes, analyses and visualizations of the temporal relationships between war events do not miss events with uncertain or ambiguous dates.

### 3.4. Keywords and Themes

Supporting the core classes – events, agents, places, and times – are themes and keywords. Themes are used to categorize historical events into major classes, mostly for user interface purposes, while keywords exist to provide non-event thematic foci for linking. This need was identified early on in indexing the primary sources. Often the sources would reference events in general, for example, talking about general wartime hunger and malnutrition, or the resistance of the Belgian Catholic Church to German occupation. It was thus often useful to index the documents as talking not only about singular events, but also about event types or more general concepts. In these cases the links are less direct (by one level), but still allow the discovery of related items of probable relevance. For instance, the keyword "agriculture" may be used to link documents dealing with agriculture to events affecting it.

### 3.5. Statistics

To round out the coverage on wartime Belgium, population statistics of the Belgian provinces during the war years were sourced from annual figures published by the Belgian Interior Ministry [4]. To model these statistics, the project utilizes the W3C Data Cube vocabulary<sup>16</sup>.

### 3.6. Interlinking

While the project itself strove to include only information humanities scholars would consider author-

itative, it was also desirable to link to the LOD cloud at large. To facilitate this, events and actors were linked to their counterparts in DBpedia, while places were associated with GeoNames, where appropriate. An overview of all these links is provided in Table 3. They were manually verified for accuracy by domain experts at CU.

Table 3  
Equivalency links in the dataset

Target	Nr.
IWM and other events (internal)	46
DBpedia events and actors	152
Out of the Trenches events	29
GeoNames places (skos:closeMatch)	1248

## 4. Dataset Access

The dataset is published at the LDF.fi data publication service [9], which provides browsing, editing, and visualization services on top of standard SPARQL and Linked Data browsing APIs. The main page for the project is at <http://ldf.fi/ww1lod/>, which describes both the API endpoints as well as the add-on service links in detail. Individual instances in the dataset are also defined in the <http://ldf.fi/ww1lod/> namespace using unambiguous computer-generated identifiers (e.g., <http://ldf.fi/ww1lod/a74d369d> for Viscount Bryce) and support direct Linked Data browser access with content-type negotiation.

The dataset is published under a CC-BY-SA 4.0 Creative Commons license<sup>17</sup>, which allows sharing and remixing of the dataset with attribution to the original licensors. All organizations mentioned in Table 2 should be referenced in case of sharing the dataset. The license also allows altering or redistributing the dataset with the same or similar license.

The dataset is being regularly updated, enriched, and maintained by CU researchers live on the endpoint using the SAHA web-based collaborative metadata editor for RDF data [15]. For this reason, actual instance counts or even data model specifics may not match those cited in this paper, which represents a snapshot at the time of submission. The most current information, however, is always available from the dataset's home page.

<sup>16</sup><http://www.w3.org/TR/vocab-data-cube/>

<sup>17</sup><http://creativecommons.org/licenses/by-sa/4.0/>

## 5. Related Work

As stated, the primary purpose of the dataset is as a reference vocabulary to which institutions can link their collections dealing with the First World War. To avoid duplicate work and promote uptake, the project has striven to interface with as many of the other projects in the field as possible, both those wanting to publish reference vocabularies and those publishing primary data.

First, the Trenches to Triples project has published Linked Data vocabularies of WWI events, places, and actors<sup>18</sup>. However, in actuality these vocabularies do not parse as valid RDF, and even corrected, their data model is not wholly consistent. For example, the same concepts are sometimes referred to using literals, and at other times using object references. Further, the event, actor, and place registries are separate and contain no references to one another, literal or not. Nevertheless, this data has been corrected, parsed, and loaded into a staging area at <http://ldf.fi/ww1/> for further study and possible integration into WWI LOD.

Second, Europeana 1914–1918 has created a simple vocabulary for WWI that includes a subset of the Library of Congress Subject Headings (LCSH) along with 81 additional concepts<sup>19</sup>. This is, however, still preliminary, as the added concepts occur in a localhost namespace, and currently do not link to one another. Further, the thesaurus (as the LCSH) is quite general and does not refer to individual units or events but rather to general keyword concepts such as "War Crimes" or "East Front Campaigns". Finally, the vocabulary shares the pre-Linked Data behavior, inherited from LCSH, of joining orthogonal information with keywords, such as having separate concepts named "World War, 1914-1918 – Social aspects – Great Britain" and "World War, 1914-1918 – Social aspects – Germany". Conversations are ongoing with Europeana about how the WWI LOD dataset could be aligned with their work.

The Muninn project<sup>20</sup> has published information pertaining particularly to Canada in the First World War. As of this writing, however, the primary data dump available consists of the results of a SPARQL query binding triples rather than the N-Triples mentioned on its page. Thus, the dataset has not yet

been parsed and evaluated by the WWI LOD project. Muninn does maintain a number of well-thought-out ontologies relevant to the domain. These ontologies were consulted in deciding upon the data models to be used for WWI LOD, but in the end the decision was made to use the CIDOC-CRM and related ontologies as a better supported albeit more general option.

A second Canadian project of interest is Out of the Trenches<sup>21</sup>, which in addition to publishing primary collection data also includes a reference vocabulary containing 64 richly described events and 324 actors relevant to the collection. While the actors are almost exclusively Canadian and thus did not appear in the WWI LOD data, 29 event equivalencies were discovered between the datasets and recorded in the WWI LOD project.

The CENDARI project<sup>22</sup> has a stated aim to provide WWI historians with tools by which to contextualize, customize, and share their research. The project is only starting to ramp up its vocabulary integration work, but has expressed an interest in using the WWI LOD vocabulary for this purpose.

Finally, Historypin<sup>23</sup> has expressed an interest in the WWI LOD data. Their intention is to use the project's event timeline to highlight events that occurred one hundred years ago as part of their centennial coverage.

## 6. Usage Examples

Currently, only the CU WWI Collection has been directly linked to the dataset, but discussions regarding collaboration with other institutions are ongoing. At present, other collections can be linked to in a limited manner through the equivalency links included in the project.

With regard to CU's digital collection, current work focuses on a contextual reader application<sup>24</sup>, designed to aid students in intellectually situating the primary sources they are working with and discovering further relevant resources. In the application, concepts from multiple vocabularies are identified in the CU primary source documents, with the intention of then showing the user contextual information relevant to that topic. For example, places mentioned in the texts can be

<sup>18</sup>Available at [http://data.aim25.ac.uk/about\\_t3.php](http://data.aim25.ac.uk/about_t3.php)

<sup>19</sup>Viewable at [http://skos.europeana.eu/api/find-concepts?q=inScheme:\\*&rows=1000](http://skos.europeana.eu/api/find-concepts?q=inScheme:*&rows=1000)

<sup>20</sup><http://rdf.muninn-project.org/>

<sup>21</sup><http://www.canadiana.ca/en/pcdhn-lod>

<sup>22</sup><http://www.cendari.eu/research/first-world-war-studies/>

<sup>23</sup><http://historypin.com/>

<sup>24</sup>A prototype is available at <http://demo.seco.tkk.fi/ww1/>

placed on maps, and short glosses can be given on the events, people and specialized vocabulary indicated. After further collections and vocabularies have been linked, the user interface would also then allow users to see any linked resources, for example, in Europeana or the Digital Public Library of America (DPLA).

In preliminary tests with this application, the WWI LOD vocabulary seems to perform significantly better in contextualizing the texts, for instance, than DBpedia, VIAF, LCSH or the Europeana vocabulary. The primary reasons for this seem twofold. First, the vocabulary is both more focused and rich in its field, and therefore has both a significantly higher recall as well as precision for concept location. Secondly, there has been a conscious attempt in the WWI LOD project to include alternative names and spellings for concepts, further increasing recall in the texts.

As for more direct use of the data, some example SPARQL queries are given at the dataset home page. The queries show for example that of the units in the German 3rd Army, the 100th to 106th Infantry Regiments engaged in a disproportionate number of atrocity incidents. On the other hand, there does not seem to be a clear relationship between population change in Belgian provinces during the war years as related either to the number of atrocities or to the total number of events. Another example shows results for all documents in the Colorado collection relating to themes associated with the Prussian general Friedrich Adolf Julius von Bernhardt (e.g., "Prussian militarism at work", a letter written by Henry Cleary, the Roman Catholic Bishop of Auckland, on the theme of "Prussian militarism").

A final example further highlights both the possibilities as well as current limits in querying multiple data sources. A query for items related to events that happened in West Flanders is able to bring in content from all of Europeana, the CU WWI Collection, and Out of the Trenches. However, the example had to be carefully selected. In reality, in the Out of the Trenches data items make reference only to two events: the event for the Battles of Ypres, 1917 and the event for the whole war. In Europeana on the other hand, only three individual battles are mentioned as subject annotations, while again many more items are related to the war as a whole.

## 7. Discussion and Future Work

The WWI LOD dataset currently contains a relatively rich, quality-controlled framework of events,

agents, and places related to the First World War. Thought has also been given to how collections can link not only to these entities, but also to more general categories, be they unit types, keywords, etc. As such, it currently represents the most comprehensive linked open reference available related to WW1.

Discussions are ongoing with various institutions on how the WWI LOD vocabulary could be integrated with other collections and vocabularies. Some resources already exist to which the project can map, such as Out of the Trenches, Europeana 1914–1918, and DPLA, and others will be available soon, like 1914–1918 Online.

In these integrations however, a problem that appears is that existing collections are often only sparsely annotated, and reference very general keywords. In connection with the Colorado WWI Collection, this gap was bridged not only with manual work, but also with newly developed tools for automatic annotation [14]. A similar approach should be feasible at least for other collections containing mostly textual items.

The WWI LOD dataset itself is also still a work-in-progress. While the project has striven to include many different sorts of material in order to demonstrate the potential of this approach, its coverage of all types but events is sporadic. Further work would be needed to include geographies of places other than Belgium, the structures of armies other than the German, and so on. The hope is that researchers who could benefit from inclusion of this information will add it, thereby enriching the dataset.

To facilitate this outcome, the project should reach out to more institutions holding WWI collections. To demonstrate the usefulness of the approach to potential partners, the contextual reader interface and other means of access require further development.

## Acknowledgements

We would like to thank Michael Ortiz (CU) and Tuomas Palonen (Aalto University) for annotating the resources, Michael Dulock (CU) and Holley Long (CU) for their aid with the digital collection and metadata, and Martha Hanna (CU), Patrick Tally (CU), Alan Kramer (Trinity College Dublin), Sophie de Schaepe drijver (Pennsylvania State University) and Tammy Proctor (Wittenberg University) for their expert opinion on the content.

## References

- [1] Great Britain. Army Council. War Office. Battles Nomenclature Committee. *Official names of the battles and other engagements fought by the military forces of the British Empire during the Great War, 1914-1919*. London, 1921.
- [2] Hermann Cron. *Geschichte des Deutschen Heeres im Weltkriege 1914-1918*. Biblio-Verlag, Berlin, 1937.
- [3] Belgium. Ministère de l'Intérieur et de l'Hygiène. *Annuaire statistique de la Belgique et du Congo Belge*, volume 46. Brussels, 1922.
- [4] Belgium. Ministère de l'Intérieur et de l'Hygiène. *Annuaire statistique de la Belgique et du Congo Belge*, volume 46. Brussels, 1922.
- [5] Martin Doerr. The CIDOC CRM – an ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3):75–92, 2003.
- [6] Patrick Lefevre ed. *Belgique et la Première Guerre mondiale, bibliographie: (Brussels: Musée Royal de l'Armée, 1987-2001)*. Musée royal de l'Armée, 1987-2001.
- [7] Jim Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, dec 2005.
- [8] John Horne and Alan Kramer. *German atrocities 1914: a history of denial*. New Haven: Yale University Press, 2001.
- [9] Eero Hyvönen, Jouni Tuominen, Miika Alonen, and Eetu Mäkelä. Linked data finland: A 7-star model and platform for publishing and re-using linked datasets. May 2014. Proceedings of ESWC 2014 Poster Track, Springer-Verlag.
- [10] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 2014.
- [11] Thea Lindquist, Michael Dulock, Juha Törnroos, Eero Hyvönen, and Eetu Mäkelä. Using linked open data to enhance subject access in online primary sources. *Cataloging & Classification Quarterly*, 51(8):913–928, 2013.
- [12] Thea Lindquist and Holley Long. How can educational technology facilitate student engagement with online primary sources?: A user needs assessment. *Library Hi Tech*, 29(2):224–241, 2011.
- [13] Brendan Luyt and Daniel Tan. Improving wikipedia's credibility: References and citations in a sample of history articles. *Journal of the American Society for Information Science and Technology*, 61(4):715–722, 2010.
- [14] Eetu Mäkelä. Combining a rest lexical analysis web service with sparql for mashup semantic annotation from text. In *Proceedings of the ESWC 2014 demonstration track*, Springer-Verlag, 2014.
- [15] Eetu Mäkelä and Eero Hyvönen. Sparql saha, a configurable linked data editor and browser as a service. In *Proceedings of the ESWC 2014 demonstration track*, Springer-Verlag, 2014.
- [16] Great Britain. Committee of Imperial Defence. *Principal events, 1914-1918*. History of the Great War based on official documents. 1922.
- [17] Lucy Holman Rector. Comparison of wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, 36(1):7–22, 2008.
- [18] Georg Tessin. *Deutsche Verbände und Truppen*. Osnabrück, 1974.
- [19] Julius Volz, Christian Bizer, Martin Gaedke, and Georgi Kobilarov. Discovering and maintaining links on the web of data. In Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, editors, *International Semantic Web Conference*, volume 5823 of *Lecture Notes in Computer Science*, pages 650–665. Springer, 2009.
- [20] Neil L. Waters. Why you can't cite wikipedia in my class. *Commun. ACM*, 50(9):15–17, September 2007.