Semantic National Biography of Finland

Eero Hyvönen[1,2], Petri Leskinen[1], Minna Tamper[1,2],
Jouni Tuominen[1,2], and Kirsi Keravuori[3]

[1] Semantic Computing Research Group (SeCo), Aalto University, Finland and
[2] HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland
http://seco.cs.aalto.fi, http://heldig.fi
firstname.lastname@aalto.fi
[3] Finnish Literature Society (SKS)
firstname.lastname@finlit.fi

Abstract. This paper presents the vision of publishing and utilizing textual biographies as Linked (Open) Data on the Semantic Web. As a case study, we publish the live stories of the National Biography of Finland, created by the Finnish Literature Society, as semantic, i.e., machine "understandable" metadata in a SPARQL endpoint using the Linked Data Finland (LDF.fi) service. On top of the data service various Digital Humanities applications are built. The applications include searching and studying individual personal histories as well as historical research of groups of persons using methods of prosopography. The biographical data is enriched by extracting events from unstructured and semi-structured texts, and by linking entities internally and to external data sources. A faceted semantic search engine is provided for filtering groups of people from the data for prosopographical research. An extension of the event-based CIDOC CRM ontology is used as the underlying data model, where lives are seen as chains of interlinked events populated from the data of the biographies and additional data sources, such as museum collections, library databases, and archives.

1 The Vision: Biographies as Linked Data

Biographical dictionaries, a historical genre dating back to antiquity, are scholarly resources used by the public and by the academic community alike. Most national biographical dictionaries follow the traditional form of combining a lengthy non-structured text, often written with authorial individuality and personal insight, with a structure supplement of basic biographical facts, such as family, education, works, and so on. Biographies are an invaluable information source for researchers across the disciplines with an interest in the past. [21]

Biographical on-line collections may contain tens of thousands of short biographies of historical persons of national importance whose contents are interlinked by historical events, places, acquaintances, family relations, times, objects, traditions, etc. The Oxford Dictionary of National Biography [7], with more than 60 000 lives, was first published on-line in 2004, and since then major biographical dictionaries have opened their editions on the Web. Other on-line national biographical collections include USA's American National Biography [1], Germany's Neue Deutsche Biographie [5], France's

Nouvelle Biographie générale [6], Biography Portal of the Netherlands [2], BiographyNet [3], and Dictionary of Swedish National Biography [4]. While the biographical dictionaries had an unconcealed nationalist agenda well into the 20th century, the contemporary on-line national biographies have all made an effort to include groups previously ignored by national history. Pioneering women in all fields are included, as well as many marginal and minority groups.

The National Biography of Finland started out as an on-line publication in 1997, a decade before its publication as a 10-volume book series was completed. A selection of its 6 000 lives was published on-line in Swedish as Biografiskt lexikon för Finland. In addition to the national Biography of Finland, the Biographical Centre of the Finnish Literature Society has published several other peer-reviewed biographical collections on-line, such as the Finnish Business Leaders, the Finnish Clergy (1552–1920), and the Finnish Generals and Admirals in the Russian Armed Forces (1809–1917).

Even if lots of biographical information is available online for humans to read and interpret, the information is seldom available as machine readable data for 1) Digital Humanities research and 2) to be used in Cultural Heritage (CH) portals, such as Europeana[4] and Digital Public Library of America[5], or in CH applications for the public. Furthermore, the information is distributed in different national data silos using heterogeneous formats and is written in different languages. This makes aggregation and reuse of biographical data challenging.

A biographical data source can be used to address various research questions from perspectives of different disciplines and nations, such as:

1. What kind of persons and institutions are actually included in the various national biographical dictionaries?
2. How do countries portray their "heroes"; is there a place for people that belong to minorities or opposition groups? Are women portrayed in a different way than men? What about race and ethnicity? What qualifications are used for selecting the portrayed individuals?
3. Which disciplines are the scholars representing? Are there national differences? What disciplines are considered to be of international value?
4. What kind of persons are included in the corpora at different times? What topics are considered to be breakthroughs on the long term?
5. What can we learn about the historical groups and institutions — social, religious, political, etc. — by analyzing the biographical details of their members? What is the nature of the networks that existed among them?

To address questions like these, biographies as data are needed, in many cases linked across nations and languages. This can only be done in multi-disciplinary collaboration between humanists, computer scientists, and linguists. There is a need for methods to transform semi-structured biography entries and unstructured texts into structured forms. We need methods to represent knowledge in an interoperable way across language barriers, and tooling for data analysis, visualization, and knowledge discovery.

---

[4] http://www.europeana.eu/portal/en
[5] http://dp.la

This paper focuses on Finnish biographies of persons of national importance, selected and edited by the scholarly editorial board of the National Biography of Finland and four additional collections: Business Leaders, the Finnish Clergy (1554–1721) and (1800–1920), and the Finnish Generals and Admirals in the Russian Armed Forces (1809–1917). In our earlier work [16] on the Semantic National Biography of Finland (SNBF), we addressed the research question: How can the reading experience of biographies be enhanced using web technologies? As a solution approach, the idea of data linking and a spatiotemporal visualization based on an interconnected timeline and a geographical map view were presented. To continue and complement this work, this paper presents:

1. An extended aggregated collection of biographies from different databases.
2. A new datamodel Bio CRM for representing the biographical data.
3. A new knowledge extraction pipeline for mining entity references and events from unstructured and semi-structured texts.
4. A faceted search engine for searching the biographies/persons.
5. New data analysis and visualization tools for research on groups of persons.

We first present the datasets and data model underlying the new version of SNBF. Then the process of transforming biographies into data is discussed. As end-user perspectives to the data, a faceted search and browsing application is presented, with additional data analysis and visualization tools for biographical research based on filtered datasets. In conclusion, contributions of our work are summarized, related work discussed, and directions for further research are outlined.

2   Data Model and Datasets

To enrich and link biographical data with related datasets, the data must be made semantically interoperable, either by data alignments (using, e.g., Dublin Core and the dumb down principle) or by data transformations into a harmonized form [14]. Since biographies are based on life events we selected the data harmonization approach and the event-centric CIDOC CRM[6] [10] ISO standard as the ontological basis in our case study. To adapt CIDOC CRM for biographical data it was first extended to a model we call "Bio CRM". This model was then populated by instance data from different biographical databases.

Bio CRM[7] [33] is a semantic data model for harmonizing and interlinking heterogeneous biographical information from different data sources. It is a domain specific extension of CIDOC CRM, effectively providing compatibility with other cultural heritage information, too. A natural choice for modeling life stories is the event-based approach where a person's life is seen as a sequence of spatiotemporal, possibly interlinked events from birth to death. The data model includes structures for basic data of people, personal relations, professions, and events with participants in different roles.

---

[6] http://www.cidoc-crm.org
[7] http://ldf.fi/schema/bioc/

Bio CRM makes a distinction between enduring unary roles of actors, their enduring binary relationships, and perduring events, where the participants can take different roles modeled as a role concept hierarchy. Bio CRM provides the general data model for biographical datasets. The individual datasets may concern different cultures, time periods, or are collected by different researchers that may introduce extensions for defining additional event and role types. For representing the roles of actors, we chose an VIVO/BFO-inspired[8] [32], intuitive, and simple approach where specific roles are instantiated from the role classes, and connected to actors with the property "inheres_in".

The Bio CRM model can be used as a basis for semantic data validation and enrichment by reasoning. The data model aims to support principal prosopographical query types, and is designed to be intuitive in terms of knowledge representation and writing SPARQL queries in flexible ways. Use cases for data represented using Bio CRM include prosopographical information retrieval, network analysis, knowledge discovery, and dynamic analysis. The development of Bio CRM was started in the EU COST project Reassembling the Republic of Letters[9] and was first piloted in the case of enriching and publishing the printed register of over 10 000 alumni of the Finnish Norssi high school as Linked Data [24].

Datasets The National Biography of the Finland[10] consists of biographies of notable Finnish people throughout history. The biographies describe the lives and achievements of these historical figures, containing vast amounts of references to notable Finnish and foreign figures, including internal links to other biographies of the National Biography of Finland. In addition, the text contains references to historical events, notable works (such as paintings, books, music, and acting), places (such as place of birth and death), organizations, and dates.

The data consists of several person registry databases listed in Table 1. These source datasets ware made available as CSV tables, which were converted into RDF format[11], the foundation of the Semantic Web standard stack. In addition to actors, the resulting data includes (at the moment) 13 144 biographies, 51 937 family relations, 4953 places, 3101 occupational titles, and 2938 companies extracted from the source data.

Dataset name	# of People
National Biographies	6478
Business Leaders	2235
Finnish Generals and Admirals 1809–1917	481
Finnish Clergy 1554–1721	2716
Finnish Clergy 1800–1920	1234

Table 1: The datasets provided by the Finnish Literature Society.

---

[8] http://vivoweb.org

[9] http://www.republicofletters.net

[10] http://kansallisbiografia.fi

[11] http://www.w3.org/RDF/

To earn the 5th star in the Linked Data 5-star model[12], the data was linked not only internally but also enriched with `owl:sameAs` links to the external data sources of Table 2. This facilitates data aggregation of a person described in several data sources.

Data Source	# of Links	Description
Wikipedia	5760	http://fi.wikipedia.org
Wikidata	5749	http://www.wikidata.org
BLF	972	Biografiskt Lexikon för Finland
BookSampo	715	Finnish fiction literature on the Semantic Web service
WarSampo	243	Second World War LOD service and portal
ULAN	171	Union List of Artist Names Online
VIAF	2272	Virtual International Authority Files
Geni	4935	Family research and family tree data
Homepages	43	Personal web sites
Parliament of Finland	628	Web pages of Parliament of Finland

Table 2: External data sources linked to the Semantic National Biography.

3   Content Creation: Entity and Event Extraction

This section discusses the process of entity linking and knowledge extraction from semi-structured and unstructured biographies.

Semi-structured Data Extraction A simple custom event extractor was created for transforming biographies into the Bio CRM model represented in RDF. The extractor analyzes the major parts of a biography: a textual story followed by systematically titled sections listing major achievements of the person, such as "works", "awards", and "memberships" as snippets. A snippet represents an event and typically contains mentions of years and places. For example, the biography of architect Eliel Saarinen tells "WORKS: ...; Suomen Kansallismuseo (National Museum of Finland, 1902–1911;..." indicating an artistic creation event. Also known family relations were extracted from the textual descriptions, and for each mentioned relative also a resource was added into the person ontology of the system.

The system has its own place ontology, and the ARPA linker [25] was used for finding and linking place names mentioned in the event snippets. Places in Finland were extracted from the Finnish Gazetteer of Historical Places and Maps (Hipla) databases and data service[13] [19,17]. Foreign placenames were linked using the Google Maps APIs[14]. For example, the locations of medieval universities in Europe, towns of the Hanseatic

---

[12] http://5stardata.info/en/

[13] http://hipla.fi

[14] http://developers.google.com/maps/

League, Finnish mansions, churches, and other well-known buildings in Finland were added to the place ontology by using the Google services.

In order to create temporal links for the events, textual expressions of dates and their intervals were analyzed and recognized using a set of regular expressions. The actor of the event snipped was easily determined as the subject person of the biography.

The result of processing a biography was a list of spatio-temporal events with short titles (snippet texts) related to the corresponding person. Altogether, first version of the extracted knowledge graph of the Semantic National Biography of Finland had $37\,730$ births, $25\,552$ deaths, $102\,300$ other biographical events, and $52\,000$ family relations. At the moment, the extractor uses only the snippets for event creation—more generic event extraction from the free biography narrative remains a topic of further research.

From a data linking viewpoint, the birthday and full name of the persons were known at this point, which could be used to enrich the data from several external datasets listed in Table 2. Links were created to Wikipedia, Wikidata, Biografiskt lexikon för Finland BLF[15], BookSampo[16] Linked Data, WarSampo[17] portal, ULAN[18] authority register of The J. Paul Getty Trust, VIAF[19], and the genealogical data service Geni[20]. Furthermore, some special links, like personal web pages or a person entry at the web sites of the Finnish Parliament, where extracted from corresponding Wikidata resources. At the moment no additional information is extracted from the external databases, but the plan is utilize them in this way in the future, too.

For entity linking to external databases offering a SPARQL endpoint, the tool SPARQL ARPA[21] was used. In cases where the database provides a REST API, like Wikipedia or Geni.com, a special Python script was created and used. A database specific script was used also in the case of BLF, where the data was available as a CSV formatted table.

A Pipeline for Text Analysis In order to identify entities and events from unstructured texts, a tool is being constructed to extract knowledge from them. The application uses multiple different linguistic tools to do morphological analysis, part of speech tagging, and lemmatization. In addition, the tool's purpose is to transform all the data into NLP Interchange Format (NIF)[22] and to enrich it with linguistic information gathered from the linguistic tools. This linguistic information can be then used in decision making in named entity recognition [27,11] by providing context to the entities, so that they can be linked more correctly and efficiently into the corresponding ontologies and datasets. In addition, the tool's results can be changed into a format where it is possible to disambiguate entities also manually if the results are not satisfactory to the user.

Our pipeline model is illustrated in the Fig. 1. In this model, the application can retrieve the texts from a CSV file or from a SPARQL endpoint. In order to read a CSV file, the application needs to know the columns for the text and possible document

---

[15] http://www.sls.fi/sv/projekt/blf-biografiskt-lexikon-finland

[16] http://www.kirjasampo.fi

[17] http://sotasampo.fi/en/

[18] http://www.getty.edu/research/tools/vocabularies/ulan/

[19] http://www.viaf.org

[20] http://www.geni.com

[21] http://seco.cs.aalto.fi/projects/dcert/

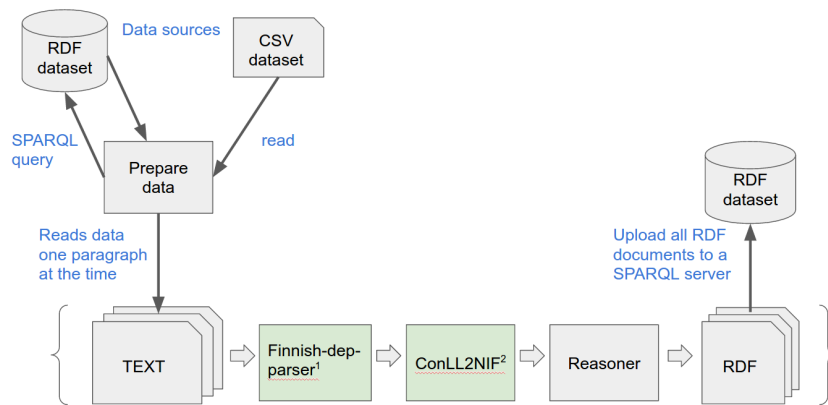[22] http://persistence.uni-leipzig.org/nlp2rdf/specification/api.html

Fig. 1: Model of the pipeline application.

identifiers. SPARQL endpoint usage requires that the text is in a SPARQL endpoint and that it is split into ordered text paragraphs. To query the text, the application needs to be given a query and an endpoint to retrieve the text.

After the application has acquired the text, it transforms (in the Prepare data module) the document structure (i.e. the document, its paragraphs and titles) into NIF format. Each document is represented as an instance of Structure class that has a property that refers to the document identifier. The document identifier can be read from the CSV file or by querying the SPARQL endpoint along with the text. In addition, the text is divided into paragraphs and titles that are divided into multiple different text files for the pipeline to process one paragraph at a time. The division into paragraphs is done because each sentence and word needs to be connected to a particular paragraph later on in the process.

The process that the paragraphs go through is represented in the figure within the curly brackets. It starts by taking a single text paragraph, transforms it into CoNLL format using the Finnish dependency parser[23]. After this, the CoNLL file is transformed into NIF format by the CoNLL2NIF [9] application. This application constructs from each word and sentence their corresponding classes (Sentence and Word) and adds the CoNLL information for each word instance. The last part of the process is a reasoner that infers missing information for each instance. Currently, the reasoner infers only information about the order of the words and sentences from the existing information. After processing a paragraph of text, the application serializes paragraph data into RDF format.

Once the application has processed all the texts, there is a possibility for a user to upload all the RDF files into a specific SPARQL server automatically. The output of this application is the given text document in NIF format. This data can be later on used to identify named entities and mentions of events in the texts. For future work we are

---

[23] http://turkunlp.github.io/Finnish-dep-parser/

planning to implement an application that can use this data to identify named entities and events, based on semantically interlinked texts and contexts. We envision that this will help semantic disambiguation of the entities as well as extraction of complex events significantly. An application of this is, for example, to use a contextual reader [26] for giving the readers of the biographies more contextual information about the entities and events mentioned in the texts.

4   Faceted Search Engine and Prosopography
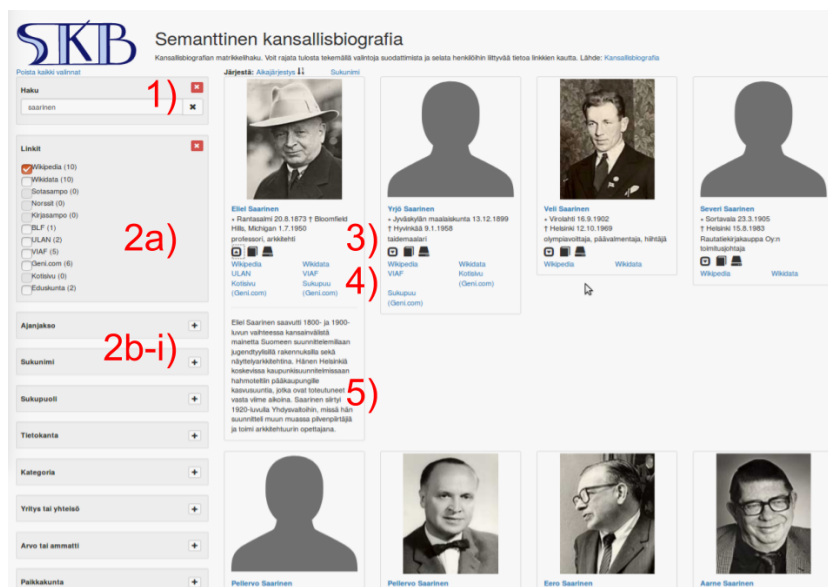
Fig. 2: Faceted search in Semantic National Biography

Based on the RDF data, a faceted search and browsing application[24] depicted in Fig. 2 was created using the SPARQL Faceter tool [22] and AngularJS[25] framefork. On the left, the first column contains a free text search 1) and facets 2a-i) for searching or filtering person entries included in external databases 2a), by time period 2b), by family name 2c), by gender 2d), etc. For example, by selecting WarSampo or Wikipedia on the external dataset facet 2a), people with a history in the WarSampo Second World War history portal or people having a Wikipedia page can be filtered, and corresponding personal homepages on these external services can be found. In the result list each person is presented by an image (if there is one), her/his lifespan 3), and his/her linkage

---

to external databases 4). The last column 5) contains original introduction text from the biographical description. Specially interesting from a data linking perspective is the facet and column for links to other external data sources. Using the links, the reading experience of an end user can be extended substantially beyond the biography text in SNBF.

After clicking on a person's image or name fields, a personal "homepage" depicted in Fig. 3 is opened. This page consists of the person's basic information 2), links to external databases 3), his relatives 4), and full biographical descriptions 5) – a biographical description can be much longer than what is shown in the figure. At the top of the page 1), the user can switch between the person page with textual descriptions or a timeline page, depicted in Fig. 4. On the timeline page, a list of events relating to the person is shown on the left 1), events with known locations are shown on the map 2), and below there is a timeline 3) showing the timespans of each event. The timeline has four horizontal lines for showing family events, career events, achievements, and mentions of honour. When an event is hovered on the list or timeline by the mouse cursor, the corresponding marker on the map is highlighted.

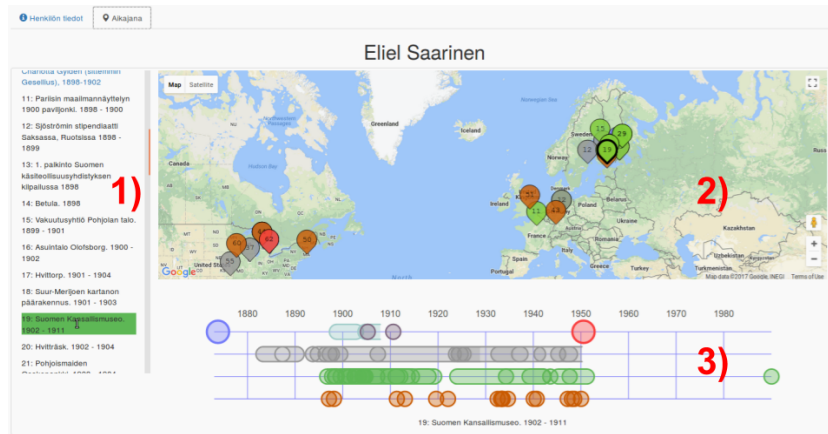Fig. 3: A person's page in Semantic National Biography

Fig. 4: A person's timeline in Semantic National Biography

The faceted search engine provides the end user with a means for filtering and studying subgroups of historical people in the data service for prosopographical research. The criterion for filtering the group can be specified flexibly by using the facets. For example, one can study people having a Wikipedia page, born in the same area during a time period, having the same education or profession, etc.

A simple tool for such analysis is business graphics, but also other methods and tools, such as network analysis or knowledge discovery could be applied here to support Digital Humanities research. To start with, pie charts and histograms, based on Google graphics, will be added to the system in a similar manner as with our earlier project on Norssi high school alumni on the Semantic Web [18]. To study the behavior of groups and other phenomena, a map and timeline application similar to the person pages will be implemented.

5   Discussion, Related Work, and Future Research

Our case study suggests that biography publication is a promising application case for Linked Data. The event-based modeling approach was deemed useful and handy, after learning basics of the fairly complex CIDOC CRM model.

For the current RDF version of the biographies, only their semistructured parts have been considered and linked by using the ARPA tool and custom linkers. The snippet events could be extracted and aligned with related places, times, and actors fairly accurately using simple string-based techniques without deeper semantic disambiguation. However, the precision and recall of event extraction and entity linking have not been evaluated formally. It is obvious that problems grow with larger datasets and when analyzing free texts. These issues remain topics of future research.

Biographies have been studied by genealogists (e.g., (Event) GEDCOM[26]), CH organizations (e.g., the Getty ULAN[27]), and semantic web researchers (e.g., BIO ontology[28]). Semantic web event models include, e.g., Event Ontology [30], LODE ontology[29], SEM [12], and Event-Model-F[30] [31]. A history ontology with map visualizations is presented in [28], and an ontology of historical events in [15]. Visualization using historical timelines is discussed, e.g., in [20], and event extraction in [13].

Previous works of applying Linked Data technologies to biographical data include, e.g., [23], Biography.net[31] [29], and our own earlier work [16]. The conference proceedings [8] includes several papers on bringing biographical data online, on analyzing biographies with computational methods, on group portraits and networks, and on visualizations. Complementing these works, the study of this paper focuses on extracting linked data from semi-structured biographies. Our work also emphasizes the idea of enriching the texts with external links to other biographical datasets. As for applications, faceted search and browsing of biographical data for prosopographical studies was considered as well as spatiotemporal visualization of life stories.

Our work continues, e.g., on developing new models of biographical data for prosopographical research, on semantic disambiguation, on finalizing and evaluating the data extraction and linking process (precision and recall), and on extending the demonstrator with new tools for solving the DH research questions discussed in section 1. In a project such as this it is of critical importance to assess the outcome also from the vantage point of historical and biographical scholarship and to evaluate relevance of the new knowledge that these methods enable us to find and analyze.

References

1. American National Biography (2017), http://www.anb.org/aboutanb.html
2. Biography Portal of the Netherlands (2017), http://www.biografischportaal.nl/en
3. BiographyNet (2017), http://www.biographynet.nl
4. Dictionary of Swedish National Biography (2017), https://sok.riksarkivet.se/Sbl/Start.aspx?lang=en
5. Neue Deutsche Biographie (2017), http://www.ndb.badw-muenchen.de/ndb_aufgaben_e.htm
6. Nouvelle Biographie générale (2017), https://fr.wikipedia.org/wiki/Nouvelle_Biographie_g%C3%A9n%C3%A9rale
7. Oxford Dictionary of National Biography (2017), http://global.oup.com/oxforddnb/info/

---

[26] http://en.wikipedia.org/wiki/GEDCOM

[27] http://www.getty.edu/research/tools/vocabularies/ulan/

[28] http://vocab.org/bio/0.1/.html

[29] http://linkedevents.org/ontology/

[30] http://www.uni-koblenz-landau.de/koblenz/fb4/AGStaab/Research/ontologies/events

[31] http://www.biographynet.nl

[32] http://seco.cs.aalto.fi/projects/severi

[33] https://openscience.fi

8. ter Braake, S., Anstke Fokkens, R.S., Declerck, T., Wandl-Vogt, E. (eds.): BD2015 Biographical Data in a Digital World 2015. CEUR Workshop Proceedings (2015), http://ceur-ws.org/Vol-1272/

9. Chiarcos, C., Fäth, C.: CoNLL-RDF: Linked corpora done in an NLP-friendly way. In: Language, Data, and Knowledge - First International Conference LDK. pp. 74–88 (2017), https://doi.org/10.1007/978-3-319-59888-8_6

10. Doerr, M.: The CIDOC CRM—an ontological approach to semantic interoperability of metadata. AI Magazine 24(3), 75–92 (2003), https://doi.org/10.1609/aimag.v24i3.1720

11. Hachey, B., Radford, W., Nothman, J., Honnibal, M., Curran, J.R.: Evaluating entity linking with Wikipedia. Artificial Intelligence 194, 130–150 (January 2013), http://dx.doi.org/10.1016/j.artint.2012.04.005

12. van Hage, W.R., Malaisé, V., Segers, R., Hollink, L., Schreiber, G.: Design and use of the simple event model (SEM). Web Semantics: Science, Services and Agents on the World Wide Web 9(2), 128–136 (2011)

13. Hogenboom, F., Frasincar, F., Kaymak, U., de Jong, F.: An overview of event extraction from text. In: DeRiVE 2011, Detection, Representation, and Exploitation of Events in the Semantic Web (2011), http://ceur-ws.org/Vol-779/

14. Hyvönen, E.: Publishing and using cultural heritage linked data on the semantic web. Morgan & Claypool, Palo Alto, CA (2012)

15. Hyvönen, E., Alm, O., Kuittinen, H.: Using an ontology of historical events in semantic portals for cultural heritage. In: Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007) (2007), http://www.cs.vu.nl/ laroyo/CH-SW.html

16. Hyvönen, E., Alonen, M., Ikkala, E., Mäkelä, E.: Life stories as event-based linked data: Case semantic national biography. In: Proceedings of ISWC 2014 Posters & Demonstrations Track. CEUR Workshop Proceedings (October 2014), http://ceur-ws.org/Vol-1272/

17. Hyvönen, E., Ikkala, E., Tuominen, J.: Linked data brokering service for historical places and maps. In: Proceedings of the 1st Workshop on Humanities in the Semantic Web (WHiSe). pp. 39–52. CEUR Workshop Proc. (2016), http://ceur-ws.org/Vol-1608/#paper-06, vol 1608

18. Hyvönen, E., Leskinen, P., Heino, E., Tuominen, J., Sirola, L.: Reassembling and enriching the life stories in printed biographical registers: Norssi high school alumni on the semantic web. In: Language, Technology and Knowledge. pp. 113–119. Springer–Verlag (2017)

19. Ikkala, E., Tuominen, J., Hyvönen, E.: Contextualizing historical places in a gazetteer by using historical maps and linked data. In: Proceedings of Digital Humanities 2016, short papers. pp. 573–577 (2016)

20. Jensen, M.: Vizualising complex semantic timelines. NewsBlip Research Papers, Report NBTR2003-001 (2003), http://www.newsblip.com/tr/

21. Keith:, T.: Changing conceptions of National Biography. Cambridge University Press (2004)

22. Koho, M., Heino, E., Hyvönen, E.: SPARQL Faceter—Client-side Faceted Search Based on SPARQL. In: Troncy, R., Verborgh, R., Nixon, L., Kurz, T., Schlegel, K., Vander Sande, M. (eds.) Joint Proc. of the 4th International Workshop on Linked Media and the 3rd Developers Hackshop. CEUR Workshop Proceedings (2016), http://ceur-ws.org/Vol-1615/semdevPaper5.pdf

23. Larson, R.: Bringing lives to light: Biography in context (2010), http://metadata.berkeley.edu/Biography_Final_Report.pdf, Final Project Report, University of Berkeley

24. Leskinen, P., Tuominen, J., Heino, E., Hyvönen, E.: An ontology and data infrastructure for publishing and using biographical linked data. In: Proceedings of the Workshop on Humanities in the Semantic Web (WHiSe II). CEUR Workshop Proceedings (October 2017)

25. Mäkelä, E.: 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 (May 2014)
26. Mäkelä, E., Lindquist, T., Hyvönen, E.: CORE – a contextual reader based on linked data. In: Proceedings of Digital Humanities 2016, long papers. pp. 267–269 (July 2016)
27. Mendes, P.N., Jakob, M., García-Silva, A., Bizer, C.: DBpedia Spotlight: shedding light on the web of documents. In: Proceedings of the 7th international conference on semantic systems. pp. 1–8. ACM (2011)
28. Nagypal, G., Deswarte, R., Oosthoek, J.: Applying the semantic web: The VICODI experience in creating visual contextualization for history. Lit Linguist Computing 20(3), 327–349 (2005), http://doi.org/10.1093/llc/fqi037
29. Ockeloen, N., Fokkens, A., ter Braake, S., Vossen, P., De Boer, V., Schreiber, G., Legêne, S.: BiographyNet: Managing provenance at multiple levels and from different perspectives. In: Proceedings of the 3rd International Conference on Linked Science (LISC'13). pp. 59–71. CEUR-WS.org (2013), http://ceur-ws.org/Vol-1116/paper7.pdf
30. Raimond, Y., Abdallah, S.: The event ontology (2007), http://motools.sourceforge.net/event/event.html
31. Scherp, A., Saathoff, C., Franz, T.: Event-Model-F (2010), http://www.uni-koblenz-landau.de/koblenz/fb4/AGStaab/Research/ontologies/events
32. Smith, B., Almeida, M., Bona, J., Brochhausen, M., Ceusters, W., Courtot, M., Dipert, R., Goldfain, A., Grenon, P., Hastings, J., Hogan, W., Jacuzzo, L., Johansson, I., Mungall, C., Natale, D., Neuhaus, F., Overton, J., Petosa, A., Rovetto, R., Ruttenberg, A., Ressler, M., Rudniki, R., Seppälä, S., Schulz, S., Zheng, J.: Basic formal ontology 2.0 – specification and user's guide (2015), https://github.com/BFO-ontology/BFO/raw/master/docs/bfo2-reference/BFO2-Reference.pdf, June 26
33. Tuominen, J., Hyvönen, E., Leskinen, P.: Bio CRM: A data model for representing biographical data for prosopographical research. In: Biographical Data in a Digital World (BD2017) (November 2017), https://doi.org/10.5281/zenodo.1040712