



Linked Data Supported Content Analysis for Sociology

Tabea Tietz^{1,2}(✉) and Harald Sack^{1,2}(✉)

¹ FIZ Karlsruhe – Leibniz Institute for Information Infrastructure,
Karlsruhe, Germany

{[tabea.tietz](mailto:tabea.tietz@fiz-karlsruhe.de),[harald.sack](mailto:harald.sack@fiz-karlsruhe.de)}@fiz-karlsruhe.de

² Karlsruhe Institute for Technology, Institute AIFB, Karlsruhe, Germany

Abstract. Philology and hermeneutics as the analysis and interpretation of natural language text in written historical sources are the predecessors of modern content analysis and date back already to antiquity. In empirical social sciences, especially in sociology, content analysis provides valuable insights to social structures and cultural norms of the present and past. With the ever growing amount of text on the web to analyze, also numerous computer-assisted text analysis techniques and tools were developed in sociological research. However, existing methods often go without sufficient standardization. As a consequence, sociological text analysis is lacking transparency, reproducibility and data re-usability.

The goal of this paper is to show, how Linked Data principles and Entity Linking techniques can be used to structure, publish and analyze natural language text for sociological research to tackle these shortcomings. This is achieved on the use case of constitutional text documents of the Netherlands from 1884 to 2016 which represent an important contribution to the European cultural heritage. Finally, the generated data is made available and re-usable as Linked Data not only for sociologists, but also for all other researchers in the digital humanities domain interested in the development of constitutions in the Netherlands.

Keywords: Cultural heritage · Sociology · NLP · Linked Data · DBpedia

1 Introduction

Since the earliest existence of writing, text served as a means of human to human communication and is firmly established in human cultures [39]. The development of the Web and (for instance) the establishment of optical character recognition (OCR) and automated speech recognition (ASR) technologies increased the amount and diversity of natural language text available to humans and machines. Cultural heritage often is manifested in text and by now, numerous means to explore cultural heritage exist to make the data accessible and explorable to a broad audience, including interactive visualizations and recommendation systems. However, in order to understand cultural heritage scientifically, fields like digital humanities and social science exist. For sociologists,

© The Author(s) 2019

M. Acosta et al. (Eds.): SEMANTiCS 2019, LNCS 11702, pp. 34–49, 2019.

https://doi.org/10.1007/978-3-030-33220-4_3

this unthinkable expanse of information captured in the form of text provides an important entry to social reality [27]. Sociological content analysis therefore also represents a necessary gateway to understanding cultural heritage and the social reality cultural heritage data captures. The mentioned cultural heritage exploration tools created for a broad audience however, are often not sufficient for sociologists to perform a scientific content analysis with. Instead, tools are needed to process, store, model, annotate (*code*) and analyze the data in order to develop new theories or test existing theories. With the increasing amount of text to be analyzed, also more technologies have been created to fulfill these tasks. In sociology, computer-assisted content analysis started out with (from today's perspective) simple frequency and valence analyses during the 1950s [35] and grew to more sophisticated statistical Natural Language Processing (NLP) approaches which became increasingly accurate and efficient in a way that they supported to uncover linguistic structures as well as semantic associations [11]. By now, a broad range of interesting and promising methods of computer assisted data acquisition and analysis have established. However, [29] criticizes that especially in social scientific research, no standardized and systematic means of the analysis of complex text material has emerged. [27] emphasizes the necessity to establish universal standards for a sustainable computer assisted text mining in sociology. Another problem in sociology regards data sharing, which is to this day widely not standardized and often not practiced at all [17,44]. According to [4], this lack of transparency lowers the integrity and interpretability of the performed research and its results. Another widely discussed issue in sociology is the re-use of research data, especially qualitative data [30]. A study by [6] suggests that sociologists generally welcome re-using research data in sociology, but certain aspects which includes the difficulty of finding and accessing these data often prevents them to do so.

The Semantic Web provides “a common framework for the liberation of data” [1] by giving data an independent existence [13]. As the Linked Open Data Cloud¹ visualizes, numerous domains have already not only firmly established methods to utilize the possibilities provided by Linked Data, they have also found ways to take part in the development, providing new applications based on the general idea. However in the field of sociological content analysis, Linked Data has so far not played an important role despite the promising standards and principles it entails.

The goal of this paper is to leverage Linked Data and its principles for computer-assisted sociological content analysis. Furthermore, it is demonstrated how this field of research can benefit from the mentioned data liberation process. Thereby, open research problems in both, the Linked Data and social science communities are discussed which (if solved) may improve the process of content analysis in the future. A lesson learned here is that in order to better understand cultural heritage data and its meaning for the society it originated in, the Linked Data research community is challenged to support sociologists in improving their research process to be more transparent, reproducible and re-usable.

¹ <https://lod-cloud.net/>, last accessed: May 12, 2019.

This paper demonstrates and discusses intersection points between Linked Data and content analysis in sociology on the foundation of the use case of constitutional text documents of the Netherlands from 1884 to 2016. The use case is generalizable and integrates Linked Data in sociological text analysis on a real world research example and thereby utilizes and discusses knowledge engineering, Named Entity Linking (NEL), and querying. Building on the previous work achieved in [41], the presented paper takes the Linked Data perspective instead of the sole sociological view.

This paper is structured as follows. In the following Sect. 2, relevant previous works on the intersection between social science and Linked Data are presented. Section 3 presents the use case of constitutional text documents and on this foundation, sociological content analysis techniques in combination with Linked Data technologies are discussed in Sect. 4.3. Section 5 closes this paper.

2 Related Work

To the best of our knowledge, no previous work exists which discusses the intersection between content analysis in sociology and Linked Data in the hereby presented depth. [13] motivated this work mostly, because the author pointed out the possibilities and necessity of Semantic Web technologies in this sociological analysis process. [2] defines annotation requirements to be implemented in cultural heritage annotation projects. The results are based on case studies at the National Library of Latvia. While the results are insightful, they do not completely apply to the process of content analysis in sociology. [11] emphasizes the foundations and applications of text mining in sociology, however, without discussing Linked Data applications. The use case to reveal intersection points between sociologists and the Linked Data community involves Dutch constitution documents. These documents were converted from their original XML format into RDF. The constitute project² as presented by [10] aimed at creating a platform for professionals drafting constitutions, and thus requiring to read and compare constitutions of various countries with each other. The main differences to the work presented here are (1) that the data is modeled not for constitution drafting but for a scientific content analysis and (2) the documents by [10] represent the latest version of a constitution and not all historical editions as it is the case in this presented paper.

3 Use Case

To assess the feasibility and benefits of modeling, storing, annotating and querying documents for sociological content analysis based on Linked Data, a generalizable research example of constitutional documents was chosen. The original document corpus was created by [23]. It consists of 20 XML documents with each one version of the Dutch constitution from 1884 to 2016 in German language.

² <https://www.constituteproject.org/ontology/>, last visited: May 12, 2019.

The previous work achieved by the authors is as important as it was cumbersome, since no machine-readable and chronological dataset of European constitutions is publicly available on the Web. Even though an HTML representation of these constitutions in German language exists on the Web³, the information which changes appeared in which constitution edition is presented in an unstructured way. In sociological research, constitution texts enable to learn about state identities, definitions of affiliations (e.g. citizens, foreigners, heads of state) and their change over time [3, 12, 28]. Constitutions can be viewed as a mirror of society and as a self-description of the state in the context of global societies [14] and therefore represent an important contribution to cultural heritage. Sociological research questions involving constitution texts include the modeling of the relationship between the state and the citizen [23], the modeling of gender in a state [16, 20] and religious freedom [26, 40]. Constitutions follow a strict structure and hierarchy. Each document is divided into several main chapters which are furthermore divided into paragraphs, articles and sections. As often required in sociological content analysis, studying constitutional documents requires to research their structure, their content as well as their changes over time. Even though this use case covers only one domain of research for sociologists, it poses versatile research problems and is generalizable to a broad range of cultural heritage texts used for content analysis in sociology.

4 Linked Data Enabled Content Analysis for Sociology

In sociological research, data sharing and publishing is neither standardized nor is it widely practiced. Studies by [44] and [17] show that social science journals have just been starting to slowly adapt data sharing policies and most journals which enforce data publishing policies do so mostly in an incomplete and varied way. The problem gets more clear when having a look at the research process itself. In sociology, content analysis is generally performed in a process in which data is pre-processed (this can involve digitizing content as well as transforming the data into the needed format for analysis), followed by a coding process (i.e. categorizing the data in varying depth) and an analysis of the produced data to establish first hypotheses or test theories. However, as mentioned in Sect. 1, this process lacks standardized methods and reproducibility which jeopardizes the integrity of research results. This section addresses all three steps in this research process and shows how Linked Data can help to improve its reproducibility and transparency based on the use case scenario described in Sect. 3. Moreover, a number of insufficiencies are discussed which pose interesting long-term research questions for interdisciplinary research.

4.1 Modeling and Publishing Documents

The corpus introduced in Sect. 3 was originally created and made available as XML by [23]. While XML provides a number of benefits regarding the way data

³ <http://www.verfassungen.eu/>, last accessed: May 12, 2019.

can be encoded syntactically, the format also has many disadvantages in contrast to RDF, especially in terms of re-usability, data extension and linking to external resources [8]. [10] specifically point out the benefits of publishing constitutional documents as RDF rather than XML. The data from the presented use case were converted to Linked Data according to the best practices specified by the W3C [19]. The Constitute Project already developed an ontology for this domain of constitution documents, which was reused and adapted. The ontology treats all parts of a constitution in the same way, regardless of its structural element (e.g. article, section, paragraph). However, the information whether a piece of text belongs to a specific paragraph or a chapter is needed for querying in the context of a sociological analysis, therefore the ontology was adapted accordingly. Furthermore, the ontology by [10] models the year the respective constitution was created in, but often several constitution versions are created in the same year. The ontology was adapted accordingly for the presented use case. [23] created the constitutional XML documents which were utilized in the presented work. As a contribution for this paper, the data were modeled and published as Linked Open Data. As a result, anyone is now able to re-use the data, query the data using the standardized SPARQL query language as opposed to proprietary XML parsers, and to reference each single semantic unit of a document separately. An example snippet of the generated RDF data is depicted in Fig. 1. All generated RDF data are made available on Github⁴. For sociologists to model and publish their data as Linked Data for content analysis to become better reproducible and re-usable, this process seems straight forward. In order to find existing vocabularies for re-use, several tools exist, including Linked Open Vocabularies⁵ or Prefix.cc⁶. Furthermore, there are a number of tools and guides to support researchers in the development and reuse of ontologies, e.g. [18, 32, 37].

4.2 Semantic Annotation

As mentioned above, a major part of the analysis of textual content in sociology is referred to as *coding*. This process means to categorize texts for analysis in order to develop new theories or test existing ones. One issue in this process is that often closed source tools are used which store the resulting data in proprietary formats (e.g. MAXQDA⁷ or ATLAS.ti⁸). If neither the textual mentions the code is referring to nor the terms or categories used for coding (and their relationships) are made available immediately together with the concluding text drawn from the analysis, the research is not reproducible. One solution is to implement semantic annotation which makes use of ontologies, which explicitly structure knowledge and define relationships between concepts and individuals. On the example of the described use case, this section demonstrates and discusses semantic annotation for the content analysis process in sociology.

⁴ <https://github.com/tabbeatietz/semsoc>, last accessed: May 12, 2019.

⁵ <https://lov.linkeddata.es/dataset/lov/>, last visited: May 12, 2019.

⁶ <http://prefix.cc/>, last visited: May 12, 2019.

⁷ <https://www.maxqda.com/>, last accessed: May 12, 2019.

⁸ <https://atlasti.com/>, last accessed: May 12, 2019.

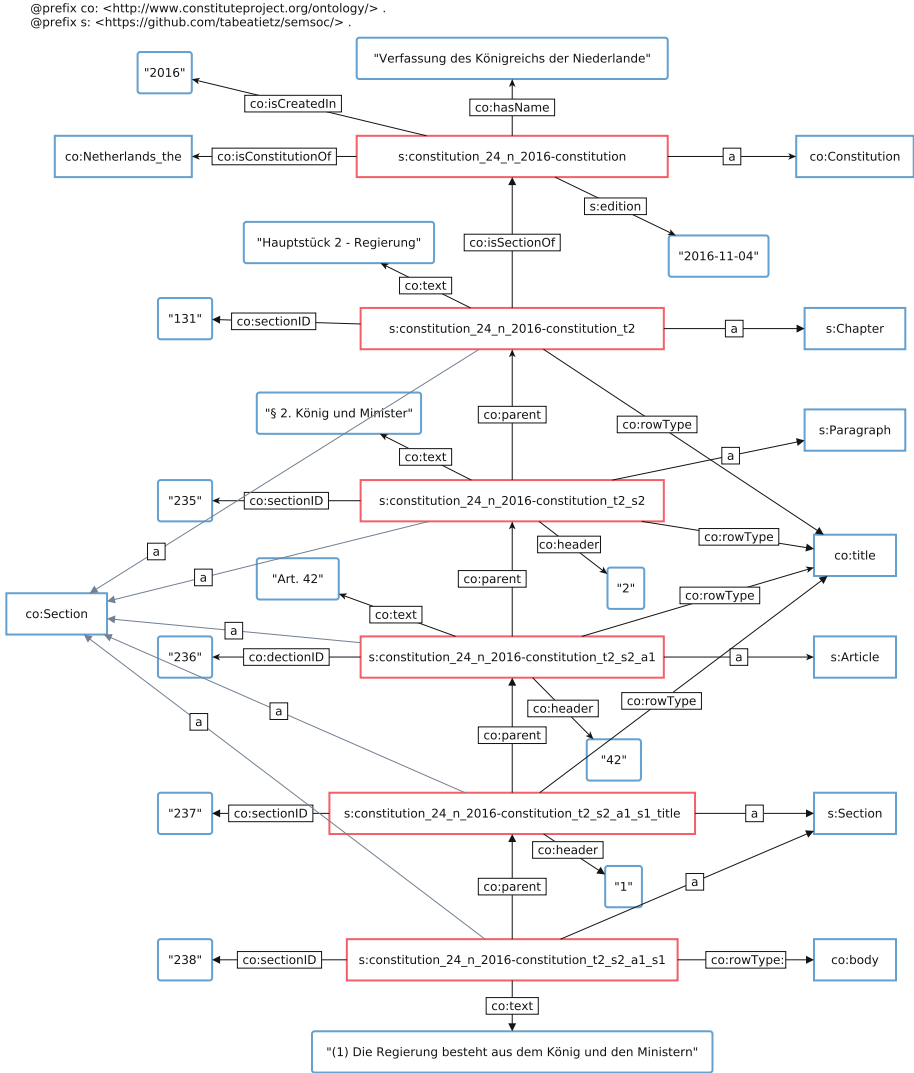


Fig. 1. Visualization of a subset of the generated RDF graph

Annotation System. Manual or semi-automatic annotation of text with entities from a large knowledge base like DBpedia requires an efficient user interface. The task of the user interface is to suggest possible entity candidates to the annotating user based on an input text. One of the major challenges is to present the entities in a way that users unfamiliar with Linked Data (lay-users) are able to make use of the interfaces. Lay-users typically have no further insight about what the content of a knowledge base is or how it is structured, which has to be considered when suggesting the entities the user should choose from [38].

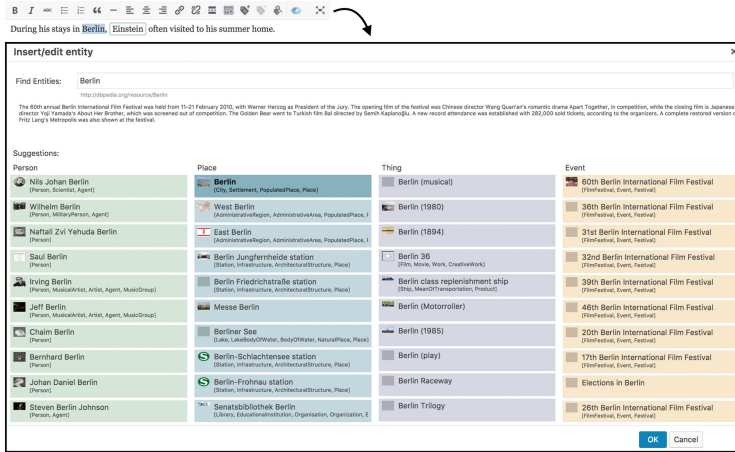


Fig. 2. *refer* Modal annotation interface

Some entity mentions yield to lists of thousands of candidates which a human cannot survey quickly to find the correct one. Therefore, *autosuggestion* utilities are applied to rank and organize the candidate lists according to e.g. string similarity with the entity mention, or general popularity of the entity [34].

There exist many semantic annotation systems, as e.g. [9], which enables semi-automated semantic text annotation in real-time. This feature seems promising but is not applicable for the presented use case. The *Pundit Annotator Pro* by [31] allows users to define their own properties and knowledge bases. However, the annotator is not available for free. Another alternative is the INCEPTION annotator by [22] which implements a variety of complex linguistic and semantic annotation functionalities. However, to semantically annotate parts of the constitution documents from the mentioned use case and to assess the sufficiency of the DBpedia knowledge base and annotation techniques for sociology, the *refer* annotation system was used [42]. *refer* consists of a set of powerful tools focusing on NEL. It aims at helping text authors to semi-automatically analyze textual content and semantically annotate it with DBpedia entities. In *refer*, automated NEL is complemented by manual semantic annotation supported by sophisticated autosuggestion of candidate entities. *refer* is chosen for this task, because it fulfills all annotation criteria mentioned by [21], is publicly available⁹, and configurable. Furthermore, a user study focusing on lay-users has shown that the *refer* annotation interface is easy to use and enables a sophisticated annotation process for lay users [42]. The user can choose between a manual and automated annotation process. For automated annotation, *refer* deploys KEA-NEL [43]. For manual (or semi-automated) annotations, the *refer* annotator includes two configurable interfaces for creating or correcting annotations: the Modal annotator, shown in Fig. 2 and an the Inline annotator. The interface

⁹ <https://www.refer.cx/>, last accessed: May 12, 2019.

leaves sufficient space for displaying relevant entities and additional information. Also, it provides a useful parallel view of all available categories. While this manual method seems (and is) cumbersome, it enables to evaluate the feasibility of DBpedia for constitution documents in depth.

Annotation Criteria. For the presented annotation task, several annotation criteria were defined, crucial for reproducibility. For sociological text analysis, it is generally assumed that rigid as well as non-rigid designators are important [25]. The rationale here is to generate as much knowledge as possible from the text to be able to analyze the data from multiple perspectives. Further entity annotation criteria regard entity specificity and completeness. It was defined to annotate textual mentions with semantic entities as specific and as complete as possible. A 'Not In List' (NIL) entity was created and included in the configurable annotation interface. Whenever the annotating user encountered an entity not available in the knowledge base, the NIL entity was used to assess the level of completeness of the annotations and the sufficiency of the knowledge base. When annotating historical text documents for scientific analysis, it is especially important to acknowledge the entities' temporal role. That means, if a text in a Dutch constitution document edition from the year 2016 mentions a term like *'der König'* (the King), the term was annotated with the DBpedia resource `dbr:Willem-Alexander_of_the_Netherlands`. This task of temporal role detection is part of current research in NLP. Advances in this field have been accomplished by [24], the topic is also tackled in a current research project led by the University of Zurich¹⁰. Even though the NLP and NEL technologies are constantly improving, this rather difficult task of disambiguation has not yet been solved in a way that it can be easily implemented in any domain. This aspect also affirmed the decision to proceed with a manual annotation process in this use case.

Result. Parts of three constitutional documents were semantically annotated with DBpedia entities according to the criteria and method discussed above. The RDFa output created with *refer* was converted into NIF2 to ensure interoperability between language resources and annotations [15]. Overall, 1.175 annotations were created in three constitution documents using 218 distinct DBpedia entities. This means that on average, each DBpedia entity was used around five times. Over all documents, 242 NIL annotations were used, which means that around 20% of all named entities in the documents were not in the knowledge base (or could not be found). All annotations and a list of NIL annotation surface forms is presented on Github¹¹.

¹⁰ <http://www.cl.uzh.ch/en/research/completed-research/hist-temporal-entities.html>, last accessed: May 12, 2019.

¹¹ <https://github.com/tabbeatietz/semloc>, last accessed: May 12, 2019.

Lessons Learned. Overall, it can be concluded that semantic annotations significantly improve the reproducibility of the research process, especially using ontologies like NIF2 or Open Annotation [36], because each conclusion drawn from the annotation (or coding) process can be proven directly in the annotation document up to character level. Data re-use is also ensured, especially if the annotation criteria are listed in the research process. The created annotations may be re-used in form of RDFa, useful for HTML pages, or NIF2 useful for querying and further adaptation. In general if the annotations are created thoroughly, they can furthermore function as a gold standard for computer scientists to improve and test NEL systems, especially with regard to the annotation criteria mentioned above. However, the process also revealed insufficiencies in terms of the underlying knowledge base, language problems and process automation. In the following, these shortcomings are listed and discussed with the goal to stress on their importance in future research work.

1. **Knowledge Graph:** Choosing DBpedia to annotate constitution documents seems reasonable, because the text corpus deals with constitutions, i.e. country specific information and facts about state leaders. These topics are generally well represented in Wikipedia. However, for 20% of all annotations NIL-entities were used. Therefore it can be concluded that solely using DBpedia is not enough for a profound annotation of these documents. One reason for this may be the systemic bias in Wikipedia [33]. It is easy to imagine that this problem does not only exist for constitution related documents but for a broad range of topics and domains. In general, solving this problem in the long term is crucial to enable sociologists to reliably use the knowledge base for their research process. In future work, also Wikidata should be tested as a knowledge base sufficient for the analysis, but to the best of our knowledge we could not find a user interface for annotating text with Wikidata items similar to *refer*. However, sociologists also need to partake in the process of creating knowledge graphs which fulfill their annotation needs. This way, the entire Linked Data community can benefit from this interdisciplinary approach as well.
2. **Language Issues:** Most NEL systems are created for English language text. This is a major problem when large non-English text corpora have to be analyzed. If non-English cultural heritage content is supposed to be analyzed and understood by sociologists and in any other domain, this is an important research task for the future. One prominent automated NEL system for German language text is DBpedia Spotlight [7]. However, initial experiments with the system revealed that the annotations did not meet the criteria mentioned above. Therefore it was eliminated from the research process.
3. **Historical Text:** The fact that the corpus in the use case includes documents dating back to 1884 further complicates the annotation process. For instance, it was important to map entities according to their temporal role. So far, there is no NEL system available which allows to annotate these temporal roles in German language with a decent quality. Apart from the temporal role disambiguation in this work, one challenge this corpus also provides is the changing style of language in the documents over time.

Even though these insufficiencies evolved during the annotation of constitutional texts, the problems are generalizable to a broad range of cultural heritage data. If these open research problems are resolved, social scientists seeking to understand large text corpora are able to use semantic annotation systems for their analysis process in a more automated manner.

4.3 Querying

When analyzing historical content in sociology, its changes over time as well as their causes and effects with regard to the society in which they appeared are crucial information to be studied. These changes may appear in the structure of a document as well as in the content itself. In this section it will be discussed on the example of the use case, how the previous data modelling and semantic annotation supports the analysis process. For this purpose all previously generated data was imported into the Blazegraph triple store¹² to be queried using SPARQL.

Time Based Analysis. Figure 3 visualizes an example of the analysis process which is enabled by previously modeling the data as Linked Data and querying. The different constitution editions are placed on a timeline along with information on structural changes and DBpedia context information. Constitutional documents follow a strict formal hierarchy. Each document is organized into several units, being the chapters, paragraphs, articles, and sections. A constitution’s

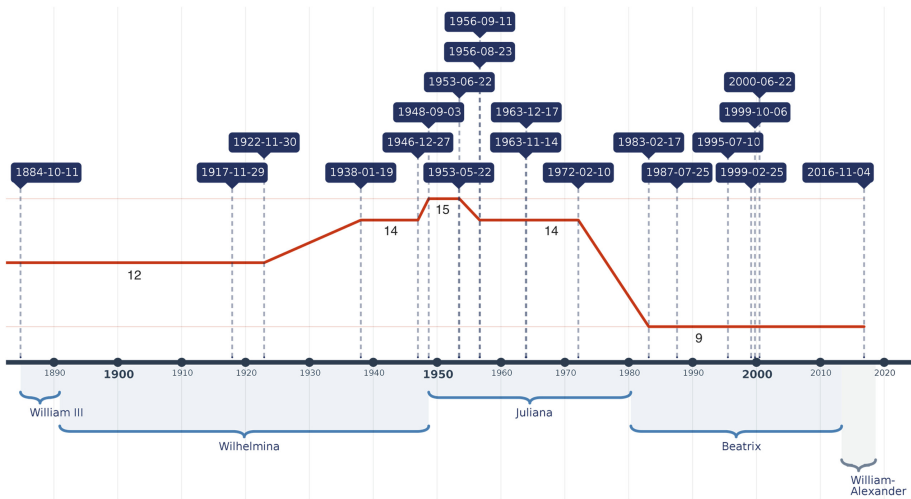


Fig. 3. Timeline of constitution editions, chapter numbers and context information (Color figure online)

¹² <https://www.blazegraph.com/>, last accessed: May 13, 2019.

chapter as the top level structural unit sets the entire framework of the constitution. Therefore querying and visualizing the changes of chapter numbers in constitutions (cf. red line) already reveal significant changes made in each document and allow the sociologist to focus on specific editions in the further analysis. Via federated querying, context information can be integrated into the process. In this case, information on the respective Dutch monarch was integrated via DBpedia, which may provide hints on the causes or effects on constitutional changes for further investigation.

Knowledge Graph Structure. Linked Data enabled sociological content analysis is especially useful when DBpedia entities are not only included into the analysis to widen the context, but also the underlying graph structure is utilized, as visualized in Fig. 4. In constitution texts, the monarch is named “King” at all times. Even if the monarch was a women (Queen). In sociology, the information of the monarch’s gender is vital [5]. With the temporal role annotations as described in Sect. 4.2, the respective constitution editions can be aggregated in

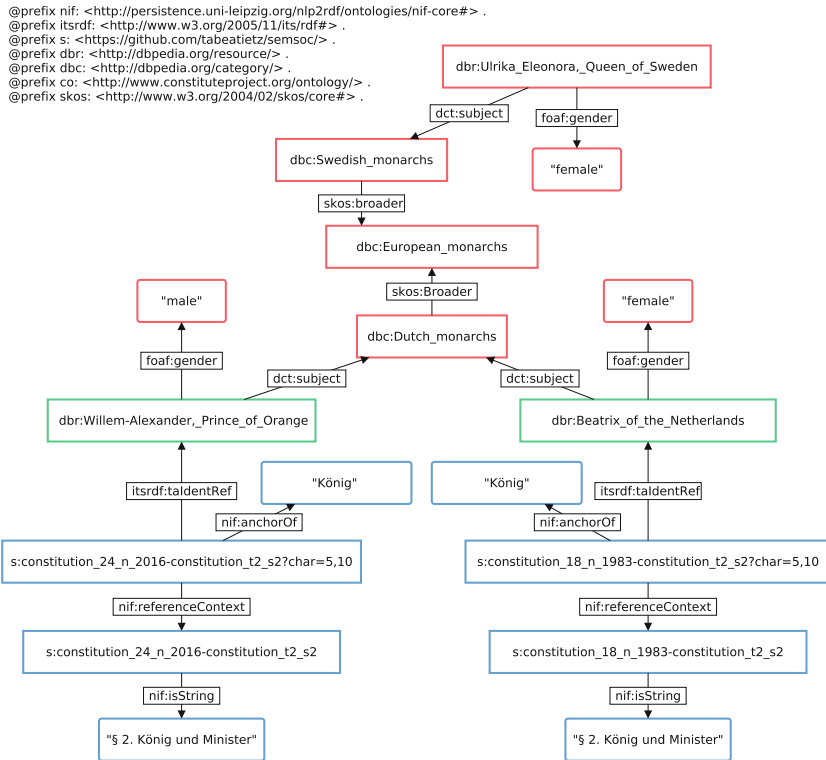


Fig. 4. In two separate constitution versions, “König” (King) was annotated with their respective DBpedia entity (Beatrix and Willem) which allows to exploit the graph structure of DBpedia.

a more meaningful manner. When only taking into account the Dutch constitution, this possibility seems rather unspectacular, but being able to aggregate all European constitutions according to the gender of the head of state emphasizes how useful Linked Data can be in this analysis process.

Visual Aids. The RDFa enrichment created with *refer* enables to visualize additional information about annotated entities directly within the context of the document which has proven to be useful for the research process. When the annotated text is published within Wordpress (as it is the case with *refer*), the annotations are immediately presented in the document’s HTML code. On mouseover, a so-called *infobox* as shown in Fig. 5 is displayed below the annotated text fragment. It contains basic information about the entity derived from DBpedia, e.g. a thumbnail and additional data from the entity RDF graph put in a table layout. When exploring an annotated document corpus of interest, sociologists can make use of these infobox visualizations to learn more about the data in front of them without having to leave the original context of the text. This can support a better understanding of the text, for instance if a certain term is unknown to them or, as shown in Fig. 5, they want to learn about the temporal roles of entities.

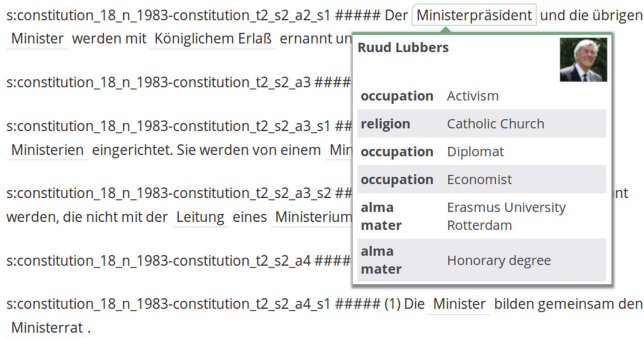


Fig. 5. Infobox visualization of former Prime Minister Ruud Lubbers

Discussion. Querying the documents for sociological content analysis with SPARQL revealed that the created data model and semantic annotations are immensely useful and allow to not only the aggregation of the data in the corpus on its own but also through the exploitation of DBpedia’s graph structure. Using SPARQL on a RDF dataset which is shared with the research community also enables to share each query which led to the respective results. To make these benefits available to a large number of sociologists, a task for interdisciplinary future work is to create effective interactive visualizations for content analysis. These visualizations can be timelines which also incorporate context information from an external knowledge graph as well as relationship visualizations.

This section demonstrated the benefits of applying Linked Data standards to the different tasks of content analysis in sociology. This involves data modeling and publishing, annotation and querying. Major open research problems include the extension and improvement of existing knowledge graphs, the improvement of NEL systems for non-English texts and the possibility to annotate entities with respect to their temporal roles. Furthermore, meaningful visualizations may be developed to enable a better scientific exploration for non-technical users.

5 Conclusion

Content analysis in sociology is a gateway to understanding cultural heritage data. While a number of methods evolved to contribute to this process of modeling, annotating and analyzing textual content, most methods lack sufficient standardization which results in a research process where the results are often not reproducible and the data cannot be reused. Linked Data may be one way to counter these problems. The goal of this paper was therefore to present and discuss intersection points between Linked Data and content analysis in sociology. On the use case of historical Dutch constitutional documents, it was shown how Linked Data can enhance the entire research process by modeling and distributing research data in RDF, by semantically annotating texts e.g. with DBpedia entities and by querying the documents using SPARQL. One contribution of this paper is to provide lessons learned from the process, which revealed important and interesting open problems to be solved in interdisciplinary research between Linked Data experts and sociologists. Finally, it became apparent that in order to better understand cultural heritage data and its meaning for society, the Linked Data research community is challenged to support sociologists in improving their research to be more transparent, reproducible and re-usable.

References

1. Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N., Weitzner, D.J., et al.: A framework for web science. *Found. Trends Web Sci.* **1**(1), 1–130 (2006)
2. Bojārs, U., Rašmane, A., Žogla, A.: The requirements for semantic annotation of cultural heritage content. In: *Proceedings of the 2nd Workshop on Humanities in the Semantic Web (WHiSe 2017)*. CEUR WS Proceedings. vol. 2014, pp. 69–79 (2017)
3. Boli-Bennett, J.: The ideology of expanding state authority in national constitutions, 1870–1970. *National Development and the World System*, pp. 212–237 (1979)
4. Büthe, T., Jacobs, A.M., Bleich, E., Pekkanen, R.J., Trachtenberg, M.: *Qualitative & Multi-method Research* (2008)
5. Crawford, K.: *Perilous Performances: Gender and Regency in Early Modern France*, vol. 145. Harvard University Press, Cambridge (2009)
6. Curty, R.G.: Factors influencing research data reuse in the social sciences: an exploratory study. *IJDC* **11**(1), 96–117 (2016)

7. Daiber, J., Jakob, M., Hokamp, C., Mendes, P.N.: Improving efficiency and accuracy in multilingual entity extraction. In: Proceedings of the 9th International Conference on Semantic Systems (I-Semantics), pp. 121–124 (2013)
8. Decker, S., Melnik, S., Van Harmelen, F., Fensel, D., Klein, M., Broekstra, J., Erdmann, M., Horrocks, I.: The semantic web: the roles of xml and rdf. *IEEE Internet Comput.* **4**(5), 63–73 (2000)
9. Eldesouky, B., Bakry, M., Maus, H., Dengel, A.: Seed, an end-user text composition tool for the semantic web. In: Groth, P., et al. (eds.) *ISWC 2016*. LNCS, vol. 9981, pp. 218–233. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46523-4_14
10. Elkins, Z., Ginsburg, T., Melton, J., Shaffer, R., Sequeda, J.F., Miranker, D.P.: Constitute: the world's constitutions to read, search, and compare. *Web Semant.: Sci. Serv. Agents World Wide Web* **27**, 10–18 (2014)
11. Evans, J.A., Aceves, P.: Machine translation: mining text for social theory. *Annu. Rev. Sociol.* **42**, 21–50 (2016)
12. Go, J.: A globalizing constitutionalism?: views from the postcolony, 1945–2000. *Int. Sociol.* **18**(1), 71–95 (2003)
13. Halford, S., Pope, C., Weal, M.: Digital futures? sociological challenges and opportunities in the emergent semantic web. *Sociology* **47**(1), 173–189 (2013)
14. Heintz, B., Schnabel, A.: Verfassungen als spiegel globaler normen? *58*, 685–716 (2006)
15. Hellmann, S., Lehmann, J., Auer, S., Brümmer, M.: Integrating NLP using linked data. In: Alani, H., et al. (eds.) *ISWC 2013*. LNCS, vol. 8219, pp. 98–113. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41338-4_7
16. Hergenhan, J.: Geschlechterdemokratie i. d. postrevolutionären Verfassung Tunesiens. *Femina Politica-Zeitschrift f feministische Politikwiss* **24**(1), 65–72 (2015)
17. Herndon, J., O'Reilly, R.: Data sharing policies in social sciences academic journals: evolving expectations of data sharing as a form of scholarly communication. *The Academic Data Librarian in Theory and Practice, Databrarianship* (2016)
18. Horridge, M., Knublauch, H., Rector, A., Stevens, R., Wroe, C.: A practical guide to building owl ontologies using the protégé-owl plugin and co-ode tools edition 1.0. University of Manchester (2004)
19. Hyland, B., Ateazing, G., Villazón-Terrazas, B.: Best Practices for Publishing Linked Data. W3C Recommendation, W3C (2014)
20. Kameri-Mbote, P.: Constitutions as pathways to gender equality in plural legal contexts. *Oslo Law Rev.* **5**(01), 21–41 (2018)
21. Khalili, A., Auer, S.: User interfaces for semantic authoring of textual content: a systematic literature review. *Web Semant.: Sci. Serv. Agents World Wide Web* **22**, 1–18 (2013)
22. Klie, J.C., Bugert, M., Boullosa, B., de Castilho, R.E., Gurevych, I.: The inception platform: machine-assisted and knowledge-oriented interactive annotation. In: Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations. pp. 5–9. Association for Computational Linguistics (2018)
23. Knoth, A., Stede, M., Hägert, E.: Dokumentenarbeit mit hierarchisch strukturierten texten: eine historisch vergleichende analyse von verfassungen. In: Vogeler, G. (ed.) *Kritik der digitalen Vernunft. Abstract zur Jahrestagung des Verbandes Digital Humanities im deutschsprachigen Raum*, pp. 196–203. University Köln (2018)

24. Koutraki, M., Bakhshandegan-Moghaddam, F., Sack, H.: Temporal role annotation for named entities. In: Proceedings of the 14th International Conference on Semantic Systems. (to be published) (2018)
25. Kripke, S.A.: Naming and necessity. *Semantics of Natural Language*, vol. 40, pp. 253–355. Springer, Dordrecht (1972). https://doi.org/10.1007/978-94-010-2557-7_9
26. Lagler, W.: Gott im grundgesetz? zur bedeutung des gottesbezugs in unserer verfassung und zum christlichen hintergrund der grund-und menschenrechte (2000)
27. Lemke, M., Wiedemann, G.: Einleitung text mining in den sozialwissenschaften. *Text Mining in den Sozialwissenschaften*, pp. 1–13. Springer, Wiesbaden (2016). https://doi.org/10.1007/978-3-658-07224-7_1
28. Lorenz, A.: How to measure constitutional rigidity: four concepts and two alternatives. *J. Theor. Politics* **17**(3), 339–361 (2005)
29. Mayring, P.: *Qualitative Inhaltsanalyse*, 12th edn. Beltz, Weinheim (2015)
30. Moore, N.: (re)using qualitative data? *Sociol. Res. Online* **12**(3), 1–13 (2007)
31. Morbidoni, C., Piccioli, A.: Curating a document collection via crowdsourcing with pundit 2.0. In: Gandon, F., Guéret, C., Villata, S., Breslin, J., Faron-Zucker, C., Zimmermann, A. (eds.) *ESWC 2015. LNCS*, vol. 9341, pp. 102–106. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-25639-9_20
32. Musen, M.A.: The protégé project: a look back and a look forward. *AI Matters* **1**(4), 4–12 (2015)
33. Oeberst, A., Cress, U., Back, M., Nestler, S.: Individual versus collaborative information processing: the case of biases in Wikipedia. In: Cress, U., Moskaliuk, J., Jeong, H. (eds.) *Mass Collaboration and Education. CCLS*, vol. 16, pp. 165–185. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-13536-6_9
34. Osterhoff, J., Waitelonis, J., Sack, H.: Widen the peepholes! entity-based auto-suggestion as a rich and yet immediate starting point for exploratory search. In: *Proceedings of 2nd Workshop Interaction and Visualization in the Web of Data (IVDW)*. Gesellschaft für Informatik (2012)
35. Popping, R.: *Computer-Assisted Text Analysis*. Sage, Newcastle Upon Tyne (2000)
36. Sanderson, R., Ciccicarese, P., Young, B.: *Web Annotation Ontology* (2016). <https://www.w3.org/ns/oa#>. Last accessed 19 July 2018
37. Schandl, T., Blumauer, A.: PoolParty: SKOS thesaurus management utilizing linked data. In: Aroyo, L., et al. (eds.) *ESWC 2010. LNCS*, vol. 6089, pp. 421–425. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13489-0_36
38. Shneiderman, B., Plaisant, C., Cohen, M.S., Jacobs, S., Elmqvist, N., Diakopoulos, N.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Prentice Hall, Pearson (2016)
39. Silberman, N.A.: *The Oxford Companion to Archaeology*. 1. *Ache-Hoho*, vol. 1. Oxford University Press, Oxford (2012)
40. Starck, C.: Staat und religion. *Juristenzeitung* pp. 1–9 (2000)
41. Tietz, T.: *The Application of Semantic Web Technologies to Content Analysis in Sociology*. Master’s thesis (2018)
42. Tietz, T., Jäger, J., Waitelonis, J., Sack, H.: Semantic annotation and information visualization for blogposts with refer. In: *Workshop on Visualization and Interaction for Ontologies and Linked Data*, co-located with ISWC, pp. 28–40 (2016)

43. Waitelonis, J., Sack, H.: Named entity linking in #tweets with kea. In: Proceedings of 6th Workshop on 'Making Sense of Microposts', Named Entity Recognition and Linking (NEEL) Challenge in conjunction with 25th International WWW Conference CEUR-WS (2016)
44. Zenk-Möltgen, W., Lepthien, G.: Data sharing in sociology journals. *Online Inf. Rev.* **38**(6), 709–722 (2014)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

