



# Linked Data supported Content Analysis for Sociology

SEMANTICS 2019

**Tabea Tietz (@tabea\_t) and  
Harald Sack (@lysander07)**

September 11, 2019

# Introduction

...azy and <sup>as</sup> I saw my dreams looking  
ling hills, covered with small  
flowers, cut off abruptly, as the  
sea. I am in awe. Lost  
to the distance... snow caps  
to the sky, beautifully rolling but so  
formed clouds. Vaguely appearing  
ondors in ~~the~~ night.  
still out, and is beginning to set.  
now on the Oromo Volcano turns over  
the sky on the opposite horizon  
walk back through the Plaza,  
his meet be the hang-out area for  
young people at night. At the same



# Introduction

Digital humanists, historians, social scientists seek to understand cultural heritage

**Sociology** → **Content analysis:**

Necessary gateway to understanding cultural heritage and the social reality captured in the data



# Introduction



“I've also directed the Pentagon to begin the process of creating the sixth branch of our military, the Space Force -- building the space force.”



# Introduction



“I've also directed the Pentagon to begin the process of creating the sixth branch of our military, the Space Force -- building the space force.”

military

organization

country

United States

space



# Introduction



“I've also directed the Pentagon to begin the process of creating the sixth branch of our military, the Space Force -- building the space force.”

military  
organization  
country  
United States  
space



# Introduction



“I've also directed the Pentagon to begin the process of creating the sixth branch of our military, the Space Force -- building the space force.”

military

organization

country

United States

space



# Introduction

## Qualitative Content Analysis

The screenshot displays the MAXQDA software interface. The top menu bar includes Home, Import, Codes, Variables, Analysis, Mixed Methods, Visual Tools, Reports, Stats, and MAXQDA. The main window is divided into several panes:

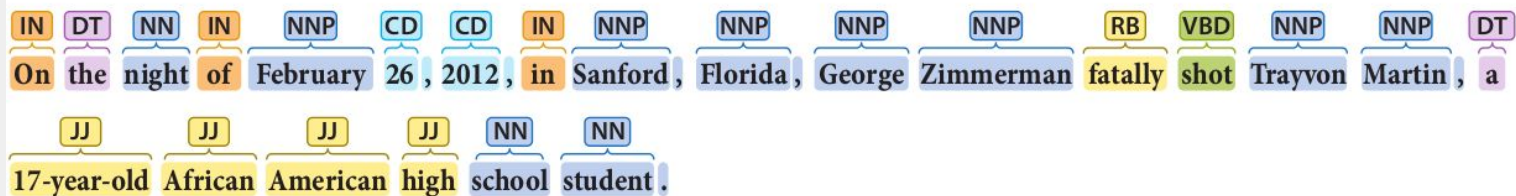
- Document System:** A tree view on the left showing a project structure with folders for 'Seminar notes', 'Interview transcripts', 'Focus group discussion', 'Videos', 'Images', 'Webpages', 'Twitter data', and 'Literature'. The 'Interview transcripts' folder is expanded, showing a list of documents.
- Code System:** A tree view on the bottom left showing a hierarchical structure of codes. The 'Interview Guide Topics' code is selected, and its sub-codes are visible.
- Document Browser:** The central pane shows a document titled 'Interview Guide Topics/Recreation/music'. It displays a text document with several paragraphs. The first paragraph is titled 'SECTION 1 - Main Topics' and contains two sections: 'What is your career now or career you are working toward?' and 'How satisfied are you with your home life?'. The text is highlighted in yellow, indicating it has been coded. A diagram on the left side of the document browser shows a network of nodes and connections, with nodes labeled 'Career', 'Home Life', and 'Parents'. The 'Career' node is connected to 'Home Life' and 'Parents', and 'Home Life' is connected to 'Parents'.
- Retrieved Segments:** A pane on the bottom right showing a list of segments retrieved from the document. The segments are listed with their corresponding codes and a small preview of the text.



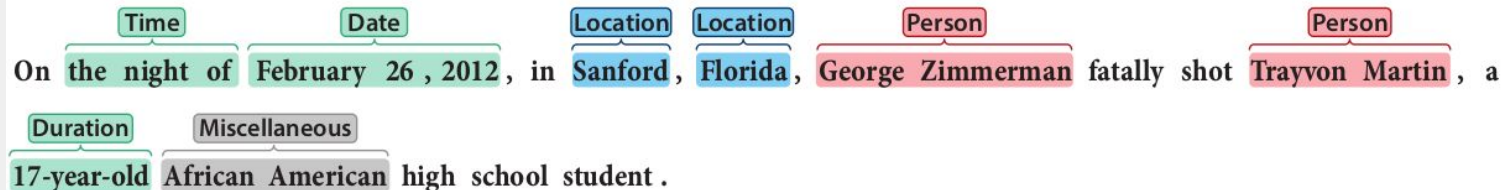
# Introduction

## Quantitative Content Analysis

### Part-of-speech tags



### Named entities tags



# Introduction

- Proprietary formats
- No systematic and standardized means of analysis and result communication

Issues with transparency, integrity of research results, re-use of data, interoperability

- Many content exploration tools are not useful for social scientific research questions



# Introduction

- Proprietary formats
- No systematic and standardized means of analysis and result communication

Issues with transparency, interoperability, re-use of data, results, re-use of data, re-use of data, re-use of data

**Linked Data based applications?**

- Many current exploration tools are not useful for social scientific research questions



# Introduction

Leverage Linked Data for computer-assisted content analysis in Sociology  
on a use case

Discover open tasks in both - the semantic web community and sociology  
- to enable a transparent, reproducible and re-usable research process

# Agenda

1. Introduction and Problem Outline
2. Related Work
3. Use Case with Lessons Learned
  - a. Data
  - b. Publishing
  - c. Analysis
4. Conclusion and Future Work

# Related Work

## **Halford et al.**

Digital futures? sociological challenges and opportunities in the emergent semantic web. *Sociology* 47 (1), 173–189 (2013)

## **Evans et al.**

Machine translation: Mining text for social theory. *Annual Review of Sociology*. 42, 21–50 (2016)

## **Bojars et al.**

The requirements for semantic annotation of cultural heritage content. In: *Proc. of the 2nd Workshop on Humanities in the Semantic Web (WHiSe 2017)*. CEUR WS Proceedings. vol. 2014, pp. 69–79 (2017)

## **Hyvönen, E. et al.**

BiographySampo–Publishing and Enriching Biographies on the Semantic Web for Digital Humanities Research. In *European Semantic Web Conference* (pp. 574-589). Springer, Cham. (2019)

# Use Case

- Text of the Constitution of the Netherlands
- In total 20 editions in the years from 1884 to 2016
- In German language



# Use Case

## Constitutions

- Are a body of fundamental principles or established precedents according to which a state or other organization is acknowledged to be governed.
- Follow a strict structure and hierarchy

Changes in individual chapters, sections or articles may appear several times a year or only every few years.

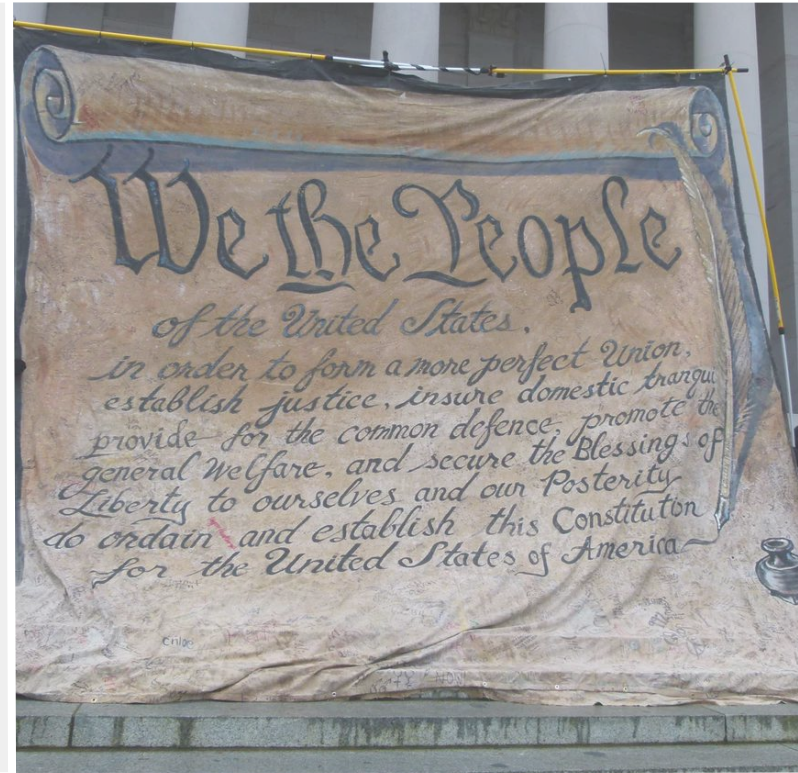
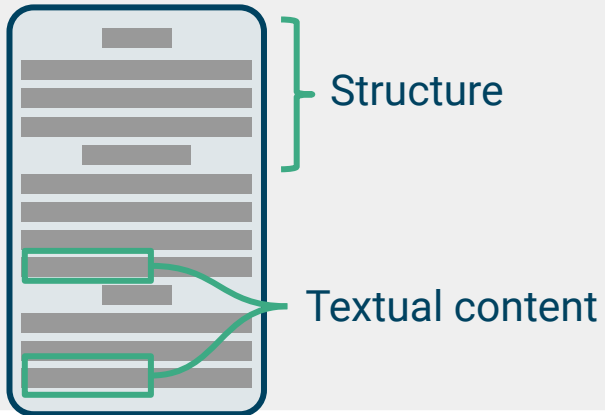




# Use Case

## Constitutions in Sociology:

- Self-descriptions of the state → mirror of society
- Research state identities, definitions of affiliations and their changes over time
- Structure + textual content



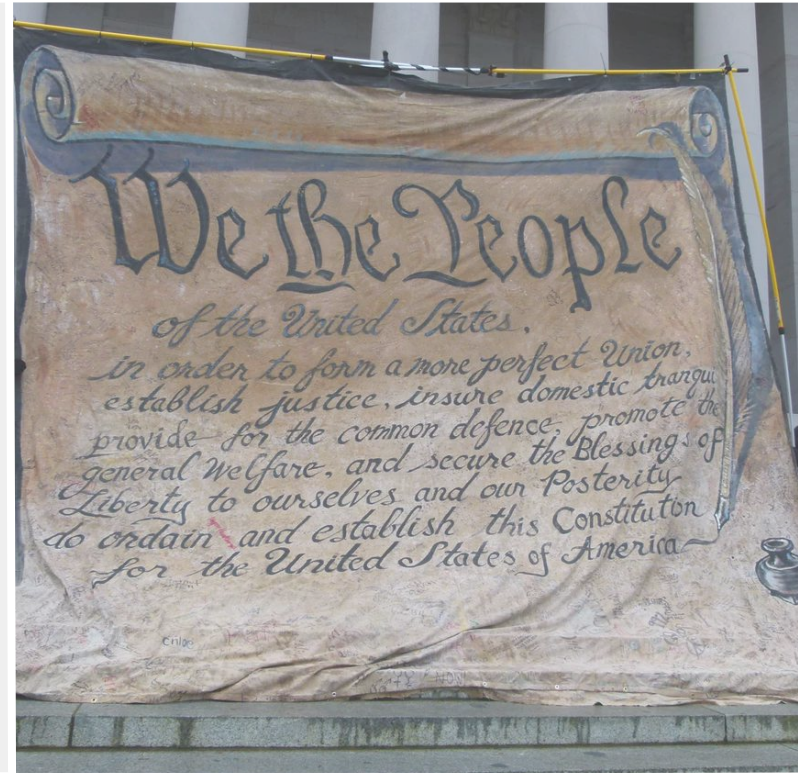
# Use Case

## Constitutions in Sociology:

- Self-descriptions of the state → mirror of society
- Research state identities, definitions of affiliations and their changes over time
- Structure + textual content

### Typical RQs:

- How are the concepts of state citizens and residents designed [1]
- How are sex/gender relations constructed? [2,3]
- How is religious freedom secured in states? [4]



## Original corpus of XML documents created manually from HTML [1]

|                |  |
|----------------|--|
| 1884.xml       | <b>Art. 8.</b> Das Recht auf Bildung von Vereinen wird anerkannt. Dieses Recht kann im Interesse der öffentlichen Ordnung durch Gesetz eingeschränkt werden.   |
| 1917.xml       | <b>Art. 9.</b> (1) Das Recht zur Versammlung und Demonstration wird anerkannt, unbeschadet der Verantwortung jedes einzelnen vor dem Gesetz.   |
| 1922.xml       | (2) Zum Schutze der Gesundheit, im Interesse des Verkehrs und zur Beseitigung oder Abwehr von Störungen können gesetzliche Vorschriften erlassen werden.   |
| 1938.xml       | <i>siehe auch Zusatzartikel V.</i>   |
| 1946.xml       | <b>Art. 10.</b> (1) Jeder hat, unbeschadet der Einschränkungen durch Gesetz oder kraft eines Gesetzes, das Recht auf Wahrung seiner Privatsphäre.  |
| 1948.xml       | (2) Der Schutz der Privatsphäre wird im Zusammenhang mit der Speicherung und Weitergabe persönlicher Daten durch Gesetz geregelt.  |
| 1953.xml       | (3) Der Anspruch von Personen auf Einblick in die über sie gesammelten Daten und deren Verwendung sowie auf Berichtigung solcher Daten wird durch Gesetz geregelt.   |
| 1953_22.6.xml  | <i>siehe auch Zusatzartikel VI.</i>  |
| 1956_23.8.xml  | <b>Art. 11.</b> Jeder hat, unbeschadet der Einschränkungen durch Gesetz oder kraft eines Gesetzes, das Recht auf körperliche Unversehrtheit.   |
| 1956_11.9.xml  | <i>siehe auch Zusatzartikel VII.</i>   |
| 1963_14.11.xml | <b>Art. 12.</b> (1) <i>Das Betreten einer Wohnung gegen den Willen des Bewohners ist nur den durch Gesetz oder kraft eines Gesetzes bezeichneten Personen in den durch Gesetz oder kraft Gesetzes bezeichneten Fällen erlaubt.</i>   |
| 1963_17.12.xml | (2) <i>Für das Betreten einer Wohnung gemäß Absatz 1 ist die vorherige Legitimation und die Mitteilung des Zwecks des Betretens der Wohnung erforderlich. Der Bewohner erhält einen schriftlichen Bericht über das Betreten der Wohnung.</i>   |
| 1972.xml       | Durch Gesetz vom 7. Februar 2002 erhielt der Artikel 12 folgende Fassung:  |
| 1983.xml       | <b>"Artikel 12.</b> (1) Das Betreten einer Wohnung ohne Zustimmung des Bewohners ist nur den durch Gesetz oder kraft Gesetzes bezeichneten Personen in den durch Gesetz oder kraft Gesetzes bezeichneten Fällen erlaubt.   |
| 1987.xml       | (2) Für das Betreten einer Wohnung gemäß Absatz 1 ist die vorherige Legitimation und die Mitteilung des Zwecks des Betretens der Wohnung erforderlich, unbeschadet der im Gesetz vorgesehenen Ausnahmen .  |
| 1995.xml       | (3) Der Bewohner erhält schnellstmöglich eine schriftliche Benachrichtigung über das Betreten der Wohnung. Wenn das Betreten der Wohnung im Interesse der nationalen Sicherheit oder der Strafverfolgung erfolgt ist, kann nach durch Gesetz festzustellenden Regeln die Benachrichtigung zurückgestellt werden. In den durch Gesetz zu bezeichnenden Fällen kann die Benachrichtigung unterbleiben, wenn sie dem Interesse der nationalen Sicherheit dauerhaft zuwiderläuft . |
| 1999_25.2.xml  |  |
| 1999_6.10.xml  |  |
| 2000.xml       |  |
| 2016.xml       |  |

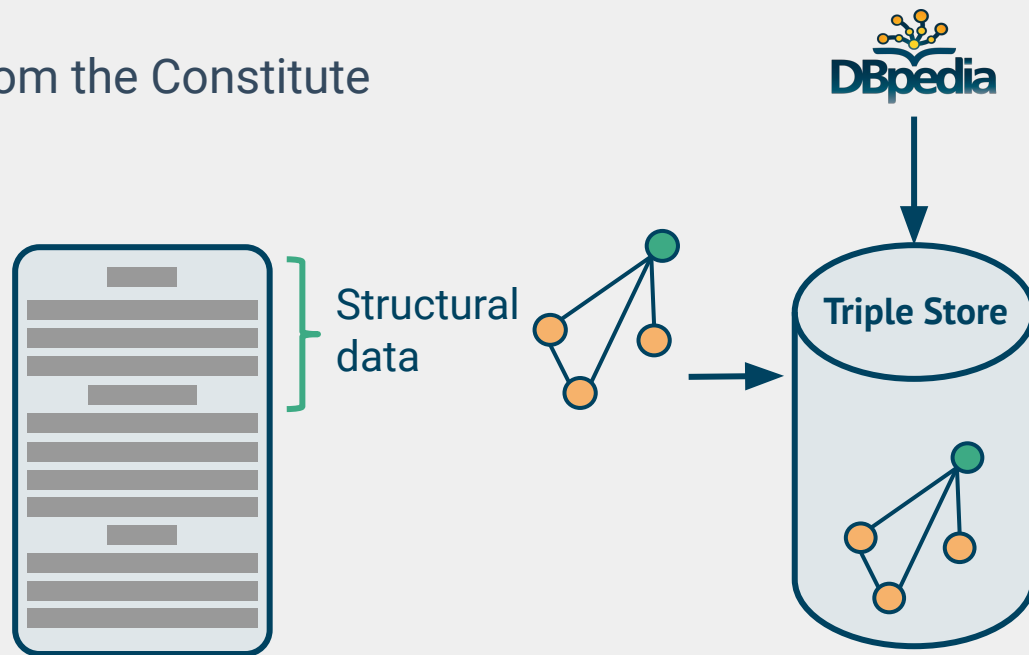
# Use Case

Converted XML to RDF

Re-using and adapting the Ontology from the Constitute Project [6]

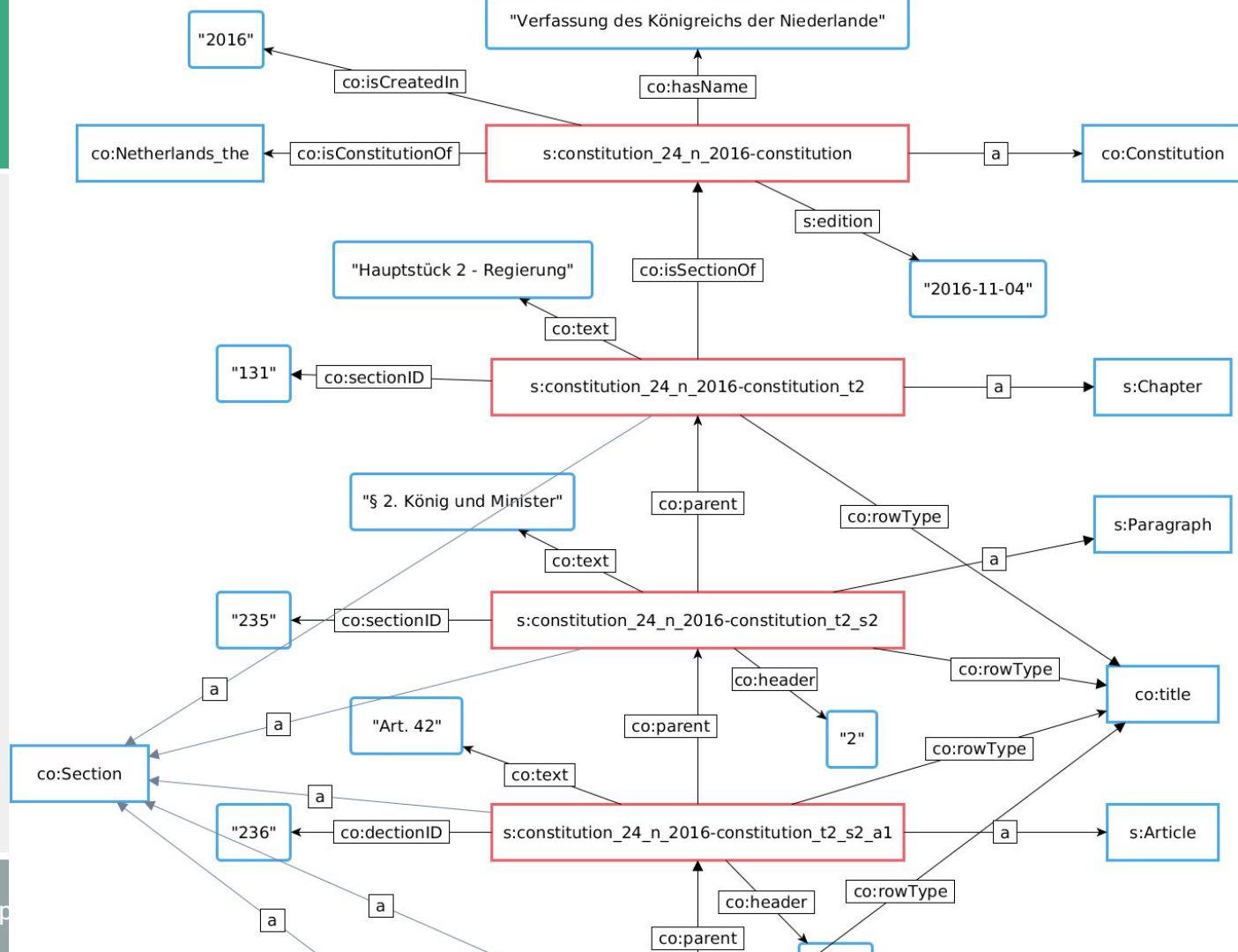
RDF Dump Files available for re-use [7]

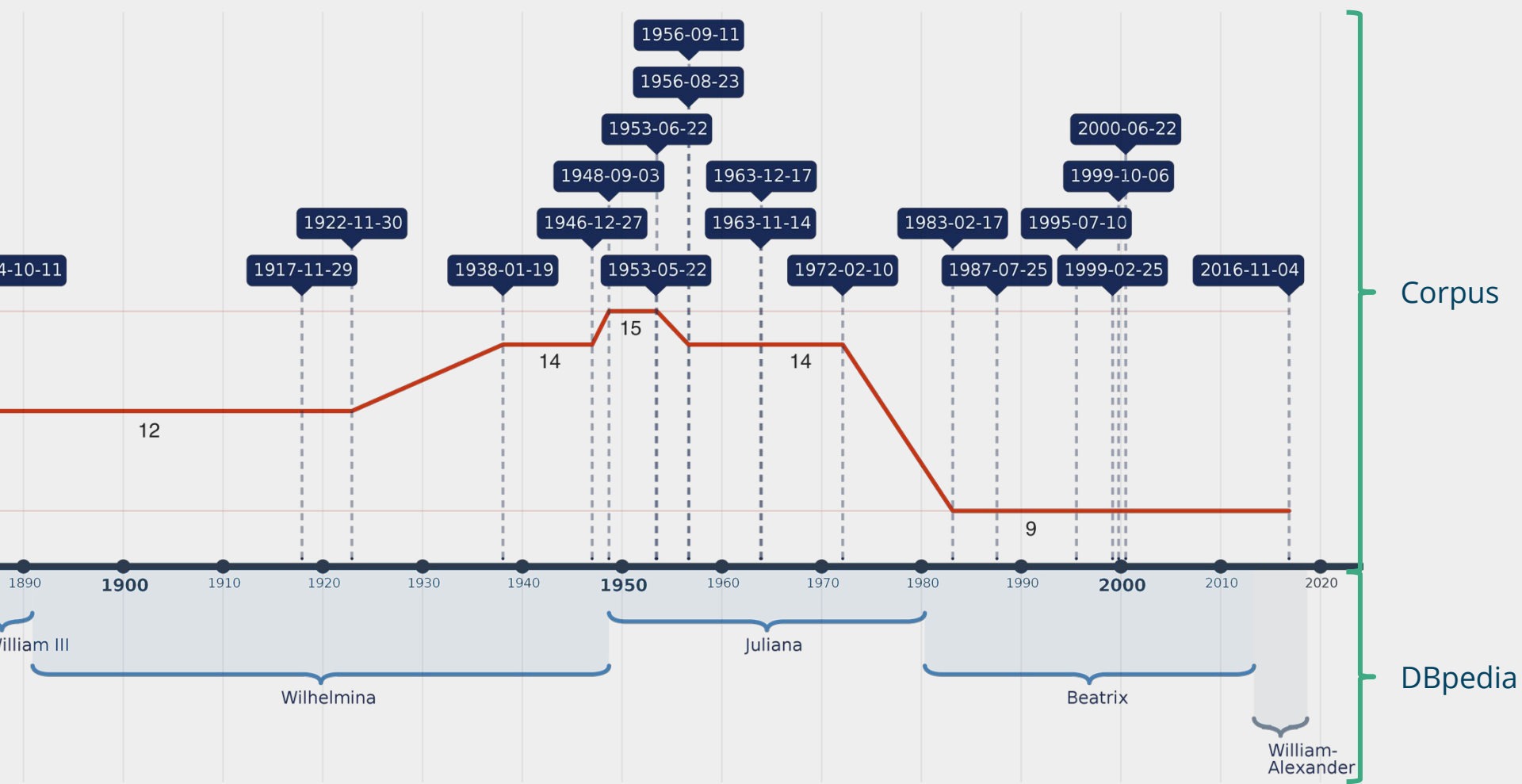
Enriched with DBpedia data



# Use Case

## Re-use and adaptation of Constitute Ontology





# Use Case

## **Analysis of the textual content:**

Constitution documents were annotated with DBpedia entities (NEL) according to a number of criteria:

### **Completeness:**

As complete as possible → added NIL “entity” for entities that could not be found

### **Respecting Temporal roles/Rigidity**

Edition from 1999: “King” → dbr:Beatrix\_of\_the\_Netherlands

# Use Case

## Annotation tool: *refer* [8]

During his stays in **Berlin**, **Einstein** often visited to his summer home.

**Insert/edit entity**

Find Entities: Berlin  
<http://dbpedia.org/resource/Berlin>

The 60th annual Berlin International Film Festival was held from 11-21 February 2010, with Werner Herzog as President of the Jury. The opening film of the festival was Chinese director Wang Quan'an's romantic drama *Apartment 404*, in competition, while the closing film is Japanese director Yoji Yamada's *About Her Brother*, which was screened out of competition. The Golden Bear went to Turkish film *Bal* directed by Semih Kapranoglu. A new record attendance was established with 282,000 sold tickets, according to the organizers. A complete restored version of Fritz Lang's *Metropolis* was also shown at the festival.

| Person   | Place   | Thing   | Event   |
|--|---|---|---|
| Nilo Johan Berlin (Person, Scientist, Agent)                     | Berlin (City, Settlement, PopulatedPlace, Place)  | Berlin (musical)  | 60th Berlin International Film Festival (FilmFestival, Event, Festival) |
| Wilhelm Berlin (Person, MilitaryPerson, Agent)                   | West Berlin (AdministrativeRegion, AdministrativeArea, PopulatedPlace, Place)           | Berlin (1980)   | 36th Berlin International Film Festival (FilmFestival, Event, Festival) |
| Naftali Zvi Yehuda Berlin (Person)                               | East Berlin (AdministrativeRegion, AdministrativeArea, PopulatedPlace, Place)           | Berlin (1894)   | 31st Berlin International Film Festival (FilmFestival, Event, Festival) |
| Saul Berlin (Person)   | Berlin Jungfernheide station (Station, Infrastructure, ArchitecturalStructure, Place)   | Berlin 36 (Film, Movie, Work, CreativeWork)                           | 32nd Berlin International Film Festival (FilmFestival, Event, Festival) |
| Irving Berlin (Person, MusicalArtist, Artist, Agent, MusicGroup) | Berlin Friedrichstraße station (Station, Infrastructure, ArchitecturalStructure, Place) | Berlin class replenishment ship (Ship, MeanOfTransportation, Product) | 39th Berlin International Film Festival (FilmFestival, Event, Festival) |
| Jeff Berlin (Person, MusicalArtist, Artist, Agent, MusicGroup)   | Messe Berlin  | Berlin (Motorroller)  | 46th Berlin International Film Festival (FilmFestival, Event, Festival) |
| Chaim Berlin (Person)  | Berliner See (Lake, LakeBodyOfWater, BodyOfWater, NaturalPlace, Place)                  | Berlin (1985)   | 20th Berlin International Film Festival (FilmFestival, Event, Festival) |
| Bernhard Berlin (Person)   | Berlin-Schlachtensee station (Station, Infrastructure, ArchitecturalStructure, Place)   | Berlin (play)   | 17th Berlin International Film Festival (FilmFestival, Event, Festival) |
| Johan Daniel Berlin (Person)                                     | Berlin-Frohnau station (Station, Infrastructure, ArchitecturalStructure, Place)         | Berlin Raceway  | Elections in Berlin   |
| Steven Berlin Johnson (Person, Agent)                            | Senatsbibliothek Berlin (Library, EducationalInstitution, Organisation, Organization, E | Berlin Trilogy  | 26th Berlin International Film Festival (FilmFestival, Event, Festival) |

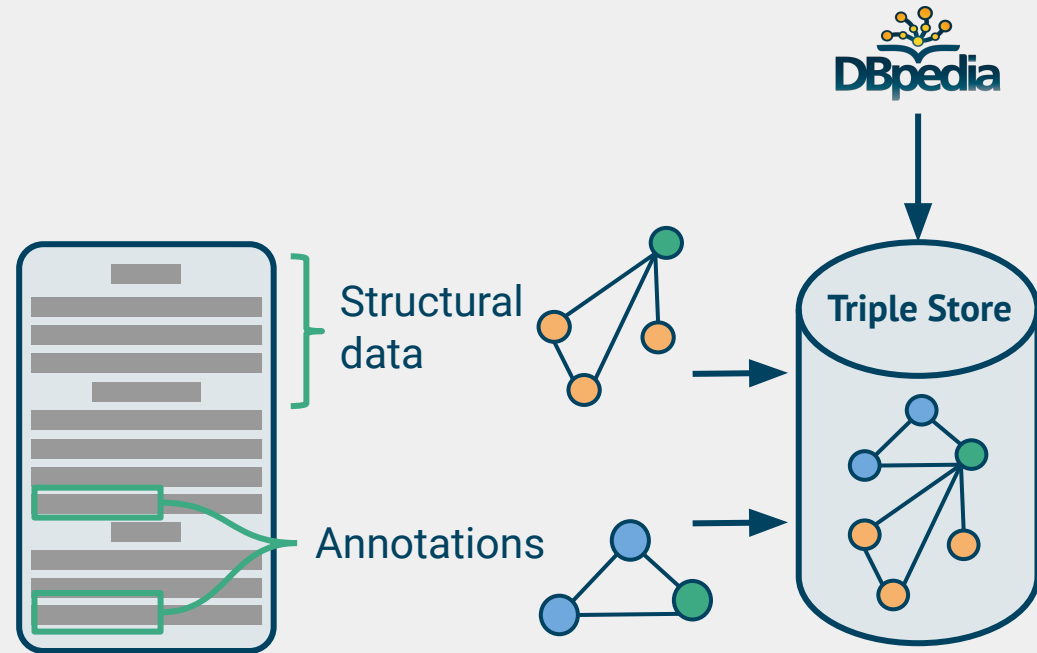
OK Cancel



# Use Case

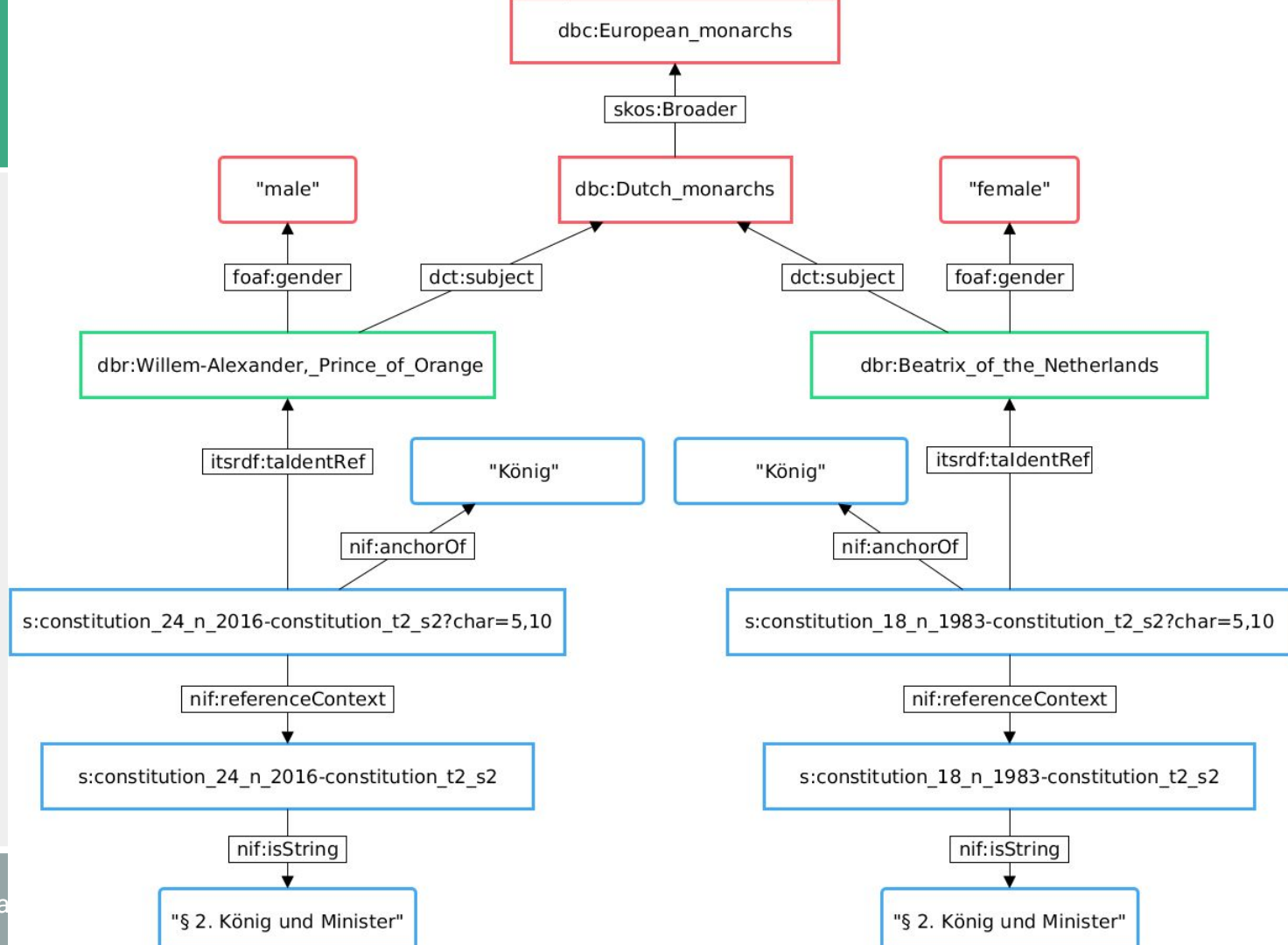
Annotations are encoded in NLP Interchange Format **NIF2** [9] and imported into the triple store

Querying now possible based on the annotations in the text and the structural context.

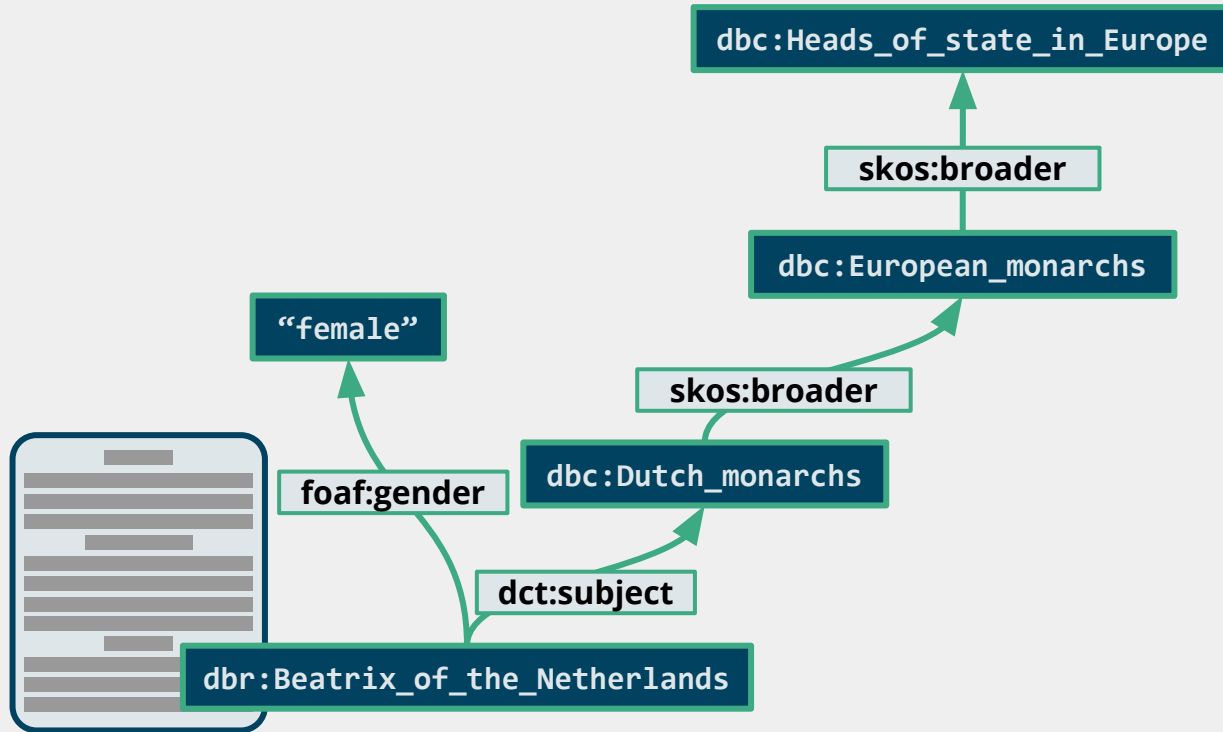


# Use Case

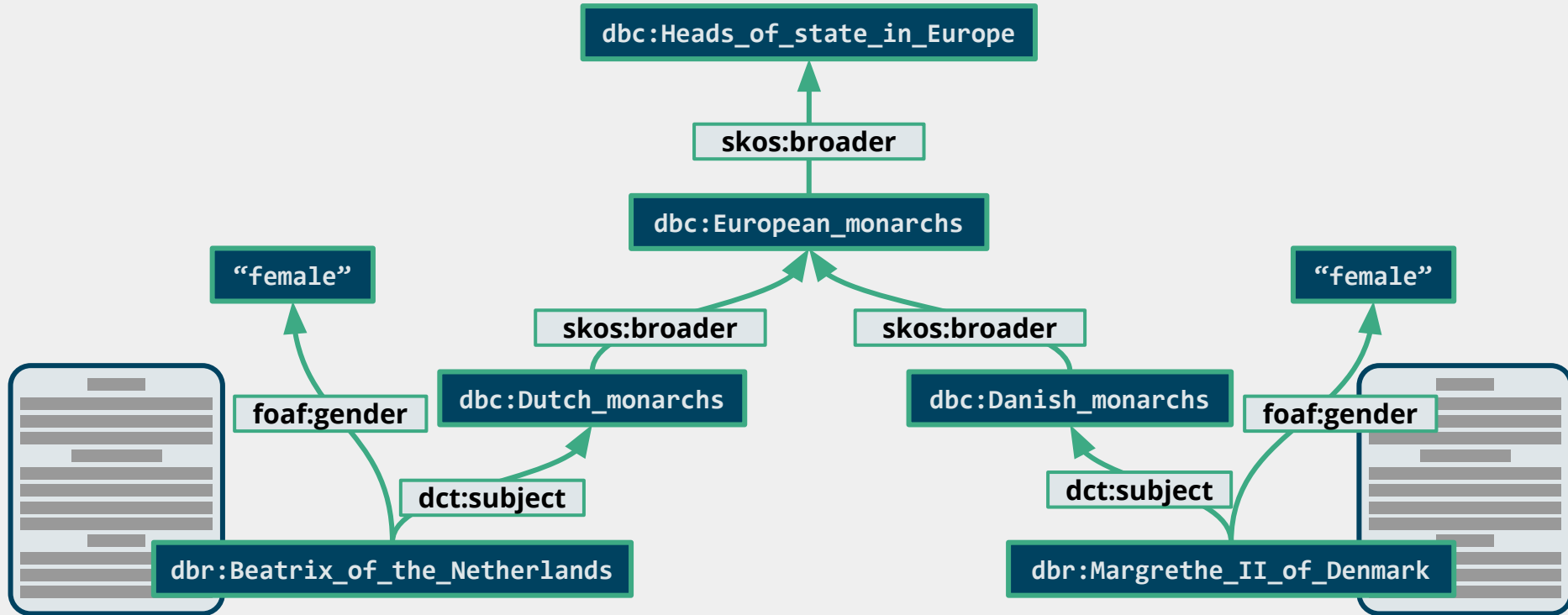
## NLP Interchange Format (NIF2) annotations



# Use Case



# Use Case



# Use Case

dbc:Heads\_of\_state\_in\_Europe

skos:broader

Using the French constitution, [10] examined how gender and regency of early-modern France interacted with each other.

A large-scale analysis of constitutions would allow patterns to be discovered. Were there similar occurrences in other European monarchies that confirm this observation?

dct:subject

dbr:Beatrix\_of\_the\_Netherlands

dct:subject

dbr:Margrethe\_II\_of\_Denmark

# Use Case

- In sociology, Linked Data principles and applications can be used to develop new hypotheses or test existing ones in content analysis
- Modeling the data as Linked Data allows easy exploration
- Dataset can be easily extended and re-used by others
- External KGs can help to analyze textual data in a specific context
- The standards and principles of the Semantic Web (URIs, NIF2) reproducible research

# Use Case - Lessons Learned

## Knowledge Graphs

20% of all annotations → NIL

→ NIL entity list available

- Knowledge graphs need to be more reliable and consistent to be used in a profound content analysis in social science
- Get domain experts to create their own ontologies

|                      |    |  |   |
|----------------------|----|--|---|
| Königswürde          | 18 | Mitglied dieser Organe                   | 2 |
| Amt                  | 18 | Einschränkungen                          | 2 |
| Amtes                | 10 | Mitglieder allgemeiner Vertretungsorgane | 2 |
| Stimmen              | 8  | schriftlich Gesuche                      | 2 |
| Königlichem Erlaß    | 8  | Veranstaltungen                          | 2 |
| Kammern              | 7  | Sitten                                   | 2 |
| Auftrag              | 6  | Störungen                                | 2 |
| Regenten             | 6  | Ministerrat                              | 2 |
| Regent               | 6  | Vorsitzenden der Versammlung             | 2 |
| Qualität             | 4  | Vorlage                                  | 2 |
| Bedingungen          | 4  | ungeborenes Kind                         | 2 |
| Saatsrat             | 4  | Mitglied                                 | 2 |
| Ernennung            | 3  | Mitgliedern                              | 2 |
| Betreten             | 3  | Personensteuer                           | 2 |
| Ausnahmen            | 3  | Übertragungs                             | 2 |
| dritten Grade        | 2  | Schenkungssteuer                         | 2 |
| Amtsübernahme        | 2  | Benachrichtigung                         | 2 |
| Einheitlichkeit      | 2  | Münzrecht                                | 1 |
| Königlichem Erlasse  | 2  | Fremde Orden                             | 1 |
| Königlichen Erlasse  | 2  | Erlaubnis                                | 1 |
| Reinigungseid        | 2  | Untertanen                               | 1 |
| Reinigungserklärung  | 2  | obersten Gerichtshofes                   | 1 |
| Reinigungsgelöbniß   | 2  | Kapitalverbrecher                        | 1 |
| Besitzungen          | 2  | Niederschlagung                          | 1 |
| Weltteilen           | 2  | Dispensationen                           | 1 |
| öffentlichen Ordnung | 2  | Landstreitmacht                          | 1 |
| Eingesessenen        | 2  | kolonialen Finanzbehörden                | 1 |
| Schwester            | 2  | Kollegien                                | 1 |
| Bewohners            | 2  | Bestimmungen                             | 1 |
| Betretens            | 2  | Derzelter Kreis                          | 1 |
| Bewohner             | 2  | Bürgerschaftsrecht                       | 1 |
| Mitteilung           | 2  | Körperschaften                           | 1 |
| Fernmeldegeheimnis   | 2  | Zuständigkeiten                          | 1 |
| Rechts               | 2  | Einwohnern                               | 1 |
| Verwaltungssachen    | 2  | Karl Georg August                        | 1 |
| Mitbestimmung        | 2  | Repräsentationsrecht                     | 1 |
| Existenzsicherheit   | 2  | Präsentationsrecht                       | 1 |
| Lebensunterhalt      | 2  | Brüder                                   | 1 |
| Schutz               | 2  | männlichen absteigenden Linie            | 1 |
| Umwelt               | 2  | männliche absteigende Linie              | 1 |
| Förderung            | 2  | ältere weibliche absteigende Linie       | 1 |
| Wohnraum             | 2  | auswärtigen Angelegenheiten              | 1 |

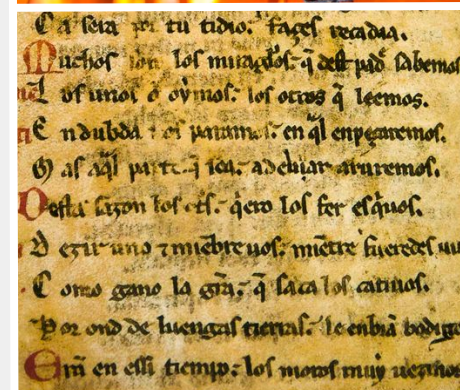
# Use Case - Lessons Learned

## Language Issues

- Most NEL systems are created for English language text
- Makes the annotation process for other languages more difficult and labor intensive

## Historical Text

- The style of language changes from 1884 to 2016 (“Königswürde”)
- Sophisticated temporal role annotations not possible at the moment, but would help a lot





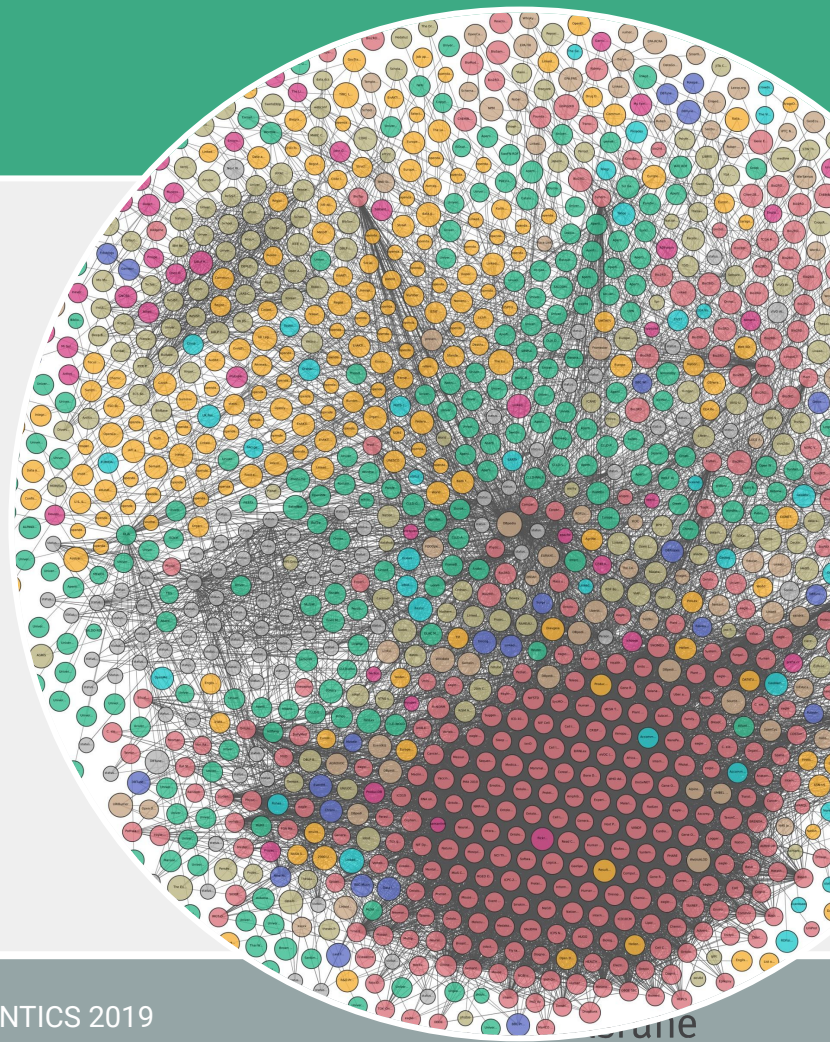
# Conclusion

- Content analysis in sociology is important to understand cultural heritage
- Research is mostly done using proprietary formats and/or often without systematic and standardized means of analysis and result communication
- Linked Data based applications can help to tackle that
- Future work:
  - Improvement of KGs
  - NEL techniques → language, style
  - Visual means of exploration → RQs in social science

# Conclusion

Help social scientists to become part of the Linked Open Data Cloud!

- valuable datasets
- annotated data
- interesting application use cases for a wide range of technologies
- enable more reproducible and transparent research process



# Bibliography

- [1] **Knoth, Alexander, Manfred Stede, und Erik Hägert** (2018). "Dokumentenarbeit mit hierarchisch strukturierten Texten: Eine historisch vergleichende Analyse von Verfassungen." In: Kritik der digitalen Vernunft. Abstracts zur Jahrestagung des Verbandes Digital Humanities im deutschsprachigen Raum, 26.02.-02.03. 2018 an der Universität zu Köln, veranstaltet vom Cologne Center for eHumanities (CCEH). Ed. von Georg Vogeler. Universität zu Köln, pp. 196 –203.
- [2] **Hergenhan, Jutta** (2015). "Geschlechterdemokratie in der postrevolutionären Verfassung Tunesiens". *Femina Politica -Zeitschrift für feministische Politikwissenschaft*,
- [3] **Kameri-Mbote, Patricia** (2018). "Constitutions As Pathways to Gender Equality in Plural Legal Contexts". In: *Oslo Law Review*
- [4] **Starck, Christian** (2000). "Staat und Religion". In: *JuristenZeitung*, 55. Jahrg., Nr. 1 Mohr.
- [5] **Lagler, Wilfried** (2000). Gott im Grundgesetz? Zur Bedeutung des Gottesbezugs in unserer Verfassung und zum christlichen Hintergrund der Grund-und Menschenrechte.
- [6] **Elkins, Zachary, Tom Ginsburg, James Melton, Robert Shaffer, Juan F Sequeda, und Daniel P Miranker** (2014). "Constitute: The World's Constitutions to Read, Search, and Compare." In: *Web Semantics: Science, Services and Agents on the World Wide Web* 27, pp. 10–18.
- [7] <https://github.com/tabeatietz/semsoc>
- [8] **Tietz, Tabea, Joscha Jäger, Jörg Waitelonis, and Harald Sack** (2016). "Semantic Annotation and Information Visualization for Blogposts with refer." In: *Workshop on Visualization and Interaction for Ontologies and Linked Data, co-located with the 15th International Semantic Web Conference*, pp. 28–40.
- [9] **Hellmann, Sebastian** (2013). NIF 2.0 Core Ontology. *Ontology Description*. AKSW, University Leipzig
- [10] **Crawford, Katherine** (2009). *Perilous Performances: Gender and Regency in Early Modern France*. Vol. 145. Harvard University Press.

## Images:

<https://unsplash.com/photos/00nHr1Lpq6w>  
<https://unsplash.com/photos/QXv1T5c359o>  
<https://unsplash.com/photos/Rd01U0tPmQl>  
<https://unsplash.com/photos/g7hB9VnBNV4>  
<https://unsplash.com/photos/RReEAgtB7P4>  
<https://unsplash.com/photos/zeH-ljawHtg>  
<https://unsplash.com/photos/1qkyck-UL3g>

# THANK YOU

## TEASER



“The Return of German DBpedia”  
Show Case Session 11:15

#### Contact

Tabea Tietz

tabea.tietz@fiz-karlsruhe.de  
@tabea\_t

© FIZ Karlsruhe 2018

Leibniz-Institut für Informationsinfrastruktur GmbH

[www.fiz-karlsruhe.de](http://www.fiz-karlsruhe.de)

Diese Unterlagen sind ausschließlich zu Präsentationszwecken bestimmt.  
Das Copyright liegt bei FIZ Karlsruhe.  
Die Weitergabe und Verwendung ganz oder in Teilen bedarf der  
ausdrücklichen Zustimmung durch  
FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH.

© FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur GmbH 2017

 **FIZ Karlsruhe**

Leibniz Institute for Information Infrastructure

 **KIT**

Karlsruher Institut für Technologie