Early Modern European Peace Treaties Online
Final Report

Dr.-Ing. Michael Piotrowski
Leibniz Institute of European History
Mainz, Germany
piotrowski@ieg-mainz.de

March 6, 2015

Abstract

Europäische Friedensverträge der Vormoderne online ("Early Modern European Peace Treaties Online") is a comprehensive collection of about 1,800 bilateral and multilateral European peace treaties from the period of 1450 to 1789, published as an open access resource by the Leibniz Institute of European History (IEG). The goal of the project funded by the Open Humanities Award was to publish the treaties metadata as Linked Open Data, and to evaluate the use of nanopublications as a representation format for humanities data.

This report describes the background of the project, the methods and tools used, the outcome, and future work.

1 Motivation

The use of databases in historical research is not new. Historical scholars have long used database management systems to store, organize, and query data they have gathered about sources, persons, places, or other items pertinent to their research questions. In many cases, these databases are never published, but even if they are made available, they tend to remain "solitary monoliths" unconnected to other data sources. There are a number of factors that are likely to contribute to situation; the following list is probably not exhaustive:

- The prevailing research culture in the humanities still awards hardly any recognition for outputs other than monographs and journal articles. In many historical research project, the primary output are thus printed publications, and no resources are allocated for preparing data for publication, let alone integrating the data produced in the project with other data sources.
- There is a general lack of (technical) coordination and a lack of standards for data and metadata (e.g., controlled vocabularies) in the humanities, which results in poor interoperability between different data sets.
- Many databases in historical research are not planned for in advance but start out as personal tools to address the specific needs of an individual scholar. Thus, in many cases "desktop" DBMS such as Filemaker or Microsoft Access are used, which tend not to scale well, and which cannot be used as a backend for a Web frontend.
- Even when a Web interface is available, the data remains isolated because typically no API is available to query the data in other ways than those offered by the human-oriented Web interface.

So, even though the individual databases are useful resources, their *full potential* is often not be realized. A case in point is the database "Europäische Friedensverträge der Vormoderne – online" ("Early Modern European Peace Treaties Online"), a comprehensive collection of about 1,800 bilateral and multilateral European peace treaties from the period of 1450 to 1789, published as an open-access resource by the Leibniz Institute of European History (IEG) in Mainz, Germany.[1] This database was created from 2005 to 2010 in a DFG-funded project with the same name.

Peace treaties between dynasties and states form an important part of our European cultural heritage. They are essential for research into early modern peacekeeping and diplomacy. "Europäische Friedensverträge der Vormoderne online" bundles manuscripts that are scattered over archives all over Europe, often hard to access, and partly undocumented. The manuscripts—in most cases the originals signed by the negotiators representing the involved powers—were digitized between 2005 and 2010 in a DFG-funded research project. All facsimiles are annotated with basic metadata, and some particularly important treaties are also available as full-text critical editions. This unique combination of digital facsimiles and critical editions has turned out to work as a well-received starting point for scholarly research in this area.

The collection data is currently stored in a relational database with a Web front-end and is one of the most popular digital offerings of the IEG. However, it has also has some shortcomings. The database is an open-access resource, but it is not machine-processable and reusable. It also lacks some important pieces of information, in particular the language or languages of the treaty texts, and the names of the undersigned negotiators. This data was collected in a later BMBF-funded project entitled "Übersetzungsleistungen von Diplomatie und Medien im vormodernen Friedensprozess. Europa 1450–1789"[2] ("Acts of translation by diplomacy and media in pre-modern peace processes. Europe 1450–1789"), which ran from June 2009 to May 2012. Researchers at the University of Augsburg gathered all the negotiators occuring in the treaties contained in the database, as well as the languages in which they are written. However, according to Penzholz and Schmidt-Rösler (2014), it was not possible to add this data to the database of treaties; instead, the scholars at the University of Augsburg created a *separate* Microsoft Access database.

The Access database is not publically available, but excerpts of the content are published as lists on a Web site[3]. Thus, even though they are based on the *same* collection of peace treaties, there exists no machine-processable link between these two databases.

Finally there is another, more conceptual problem. It is not specific to theses databases, but applies to most databases in historical research (and in many other humanities disciplines): Conventional databases are not designed to handle uncertain and contradictory data, and there is no easy way to associate certainty and provenance information with individual items. Databases in historical research thus create—usually unintentionally—an illusion of historical factuality, when, in many cases, the historical data is uncertain, scholars' interpretations of it significantly varies, and provenance information would be needed to assess its reliability.

2 Approach

Considering the issues outlined above, we conclude that conventional databases are not well suited to the requirements of historical data and research. The goal of the project funded by the Open Humanities Award was to demonstrate Linked Open Data (LOD) as a better alternative, by bringing Early Modern Peace Treaties Online to the "Linked Data cloud," allowing researchers not only to search and browse

[1] `http://www.ieg-friedensvertraege.de/`

[2] `http://www.uebersetzungsleistungen.de/`

[3] `https://www.uni-augsburg.de/de/institute/iek/projekte/historische-friedensforschung/Materialien/`

the collection but also to use and reuse the data in novel ways and to integrate it with other collections, including Europeana. By publishing our collection of European peace treaties as Linked Open Data we also wanted to make more content and data openly available for researchers to use, and make it possible to link it to other relevant information, e.g., persons and places via GND/VIAF.

In order to address the issues of uncertainty and provenance, we wanted to explore *nanopublications* as a novel approach. Nanopublications (Groth et al., 2010) were originally developed in the biomedical domain for integrating different ontologies in a common framework in order to describe scientific statements together with their context and their provenance, so that central scientific results can be unambiguously referenced and connected to their authors, and to support discovery and automatic aggregation and analysis. Nanopublications are encoded in RDF and use named graphs for grouping all information relevant for a scientific result in a single container; thus, they are compatible with the Linked Open Data approach.

Despite their highly interesting properties, the use of nanopublications in the humanities has so far only be attempted by Heßbrüggen-Walter (2013), who has used them to attribute philosophical statements—documented in their writings—to early modern philosophers. As this aspect of the project was highly experimental, we decided to proceed stepwise and first do a straightforward conversion of the existing database into RDF to make it available as Linked Open Data, and to then examine the use of nanopublications as a format for representing information about provenance and certainty in the future.

2.1 Converting Early Modern Peace Treaties Online to Linked Open Data

The process for converting the content of the existing database into LOD basically consisted of four steps:

1. **Analyzing the data.** No documentation was available for the existing database. It consists of 11 tables and numerous fields. Some of the fields have telling names, but not all of them. Another question was what the fields would actually contain. We found out that sometimes creative solutions were used. For example, the parties of a treaty are stored in a field declared as follows:

   ```
   ‘partners‘ varchar(255) NOT NULL DEFAULT ’’
   ```

   Thus, `partners` is a string field, but it does not contain the names of the parties to the treaty, but rather their IDs, e.g., `37,46,253` in string form.

   In order to determin the names of the partners, one has to first split the string, and then can look up the names in another table to find out that 37 is France, 46 is Genoa, and 253 is Naples–Sicily. This approach was used as a workaround for the problem of storing lists of variable length, which is quite tedious in a relational database. While this approach is better than hardcoding the names of the partners in every record, it moves a part of the semantics into the application, which has to know that some string fields actually contain lists of keys for a table.

   While this example is not particularly complicated, it illustrates that a thorough analysis of the database was necessary in order to accurately extract and convert the data it contains.

   During our analysis of the database, we also discovered that some potentially interesting information is only available as unstructured text, in particular references to contemporary prints and to secondary literature. We decided to skip these fields for the time being. A closer examination will be necessary to determine what additional information could realistically be extracted, i.e., with reasonable manual effort.

2. **Identifying and selecting pertinent ontologies.** We did not want to re-invent the wheel but rather build upon existing and proven ontologies for describing treaties.

3. **Modelling the information in RDF.** Once we knew how to conceptually model the information, we needed to define how to actually represent the information on a treaty in RDF.

4. **Generating the data.** Finally, we iterated over the database, extracted the information, combined it into RDF statements, and output them in a form suitable for importing them into a triple store.

At this point, we had converted the the structured metadata from the legacy database into RDF. As we expected, the conversion required a fair bit of interpretation and cleanup work, but all in all, it worked quite well.

As the basis for our data model we have, not surprisingly, used the DM2E model. Currently we have three main classes of entities, namely the *treaties*, the *treaty partners* (or signatories–but we prefer the term *partner* to avoid confusion with the negotiators, i.e., the persons who actually signed the treaties), and finally, the *locations* where the treaties were signed. We use `dm2e:Manuscript` as class for the treaties, `edm:Agent` as class for the partners, and `edm:Place` as class for the locations. Furthermore we use the following properties:

- `dc:title` for the treaty titles,
- `dc:date` for the treaty date,
- `edm:happenedAt` for linking to the location,
- `rdfs:label` for the names of partners and locations, and
- `skos:narrower` and `skos:broader` for modeling the hierarchy of partners.

The last point may need some explanation. Partners may be in a hierarchical relationship to each other to model that a power may be part of a larger entity. For example, Austria was a part of the Holy Roman Empire, whereas Milan, Mantova, and Sardinia were (at various points in time) parts of Austria. However, historical realities tend to be quite messy, so these relations are not necessarily "part-of" relations in the strict sense; for example, Austria also had territories outside the Empire. The hierarchy also contains "fictitious partners" as a help for searching; for example, introducing *Switzerland* or *Parts of the Empire* as "fictitious partners" makes it easier to search for treaties concerning certain regions of Europe. This pragmatic approach was taken over from the legacy database, as we think it makes sense, at least for the time being.

To link the treaties to the treaty partners we used the `dc:contributor` property. This usage stretches the meaning of "contributor" a bit, but we only use it as a provisional solution, as we will reconsider the modeling when moving to nanopublications.

If we consider a specific treaty, such as the *Provisional convention of subsidy* between Great Britain, the States General, and Austria, we have the following data:

Property	Value
Type	dm2e:Manuscript
Title (`dc:title`)	Provisorischer Subsidienvertrag (de)
Date (`dc:date`)	1746-08-31
Contributor (`dc:contributor`)	Austria, Great Britain, States General
Happened at (`edm:happenedAt`)	The Hague

This display is somewhat simplified for illustration. For reference, figure 1 shows the last page of the treaty; the last sentence before the seals and signatures gives the place and the date: "Fait à La Haye le trente un du Mois d'Aout de l'année mille Sept cent quarante Six."

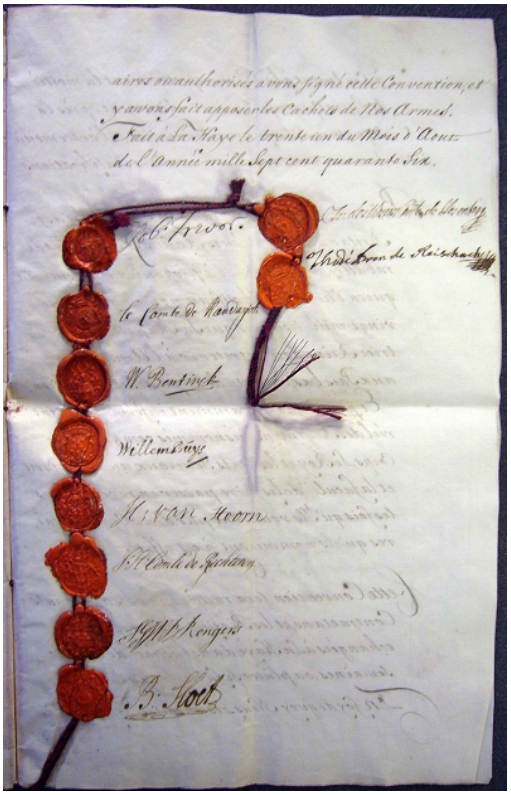

Figure 1: Provisional convention of subsidy between Great Britain, the States General, and Austria (Nationaal Archief, Den Haag, Staten-Generaal, nummer toegang 1.01.02, inventarisnummer 12597.187)

We loaded the data into Fuseki and set up a server at `http://data.ieg-friedensvertraege.de/`. Since one cannot really "see" Linked Open Data, we also set up Pubby, a Linked Data frontend for SPARQL endpoints, which gives the data a friendlier face. For example, the screenshot below shows how the information on the Friedenspräliminarien von Breslau (in English known as Treaty of Breslau) is presented in Pubby.

As noted above, there is not really much to "see" about the data, so there is not really much to show; what is exciting is the potential it has for automatic processing. As a low-key example, the homepage at `http://data.ieg-friedensvertraege.de/` currently shows a "live" list of all treaty partners, i.e., when one loads the Web page, a query is sent to the SPARQL endpoint to retrieve all entities of type `edm:Agent` (see figure 3). We intend to replace this list with a more interesting example such as a map showing the treaty locations, which could look like the mockup shown in figure 4.

However, in order to be able to draw such a map, the geographical coordinates of the treaty locations must obviously be known, whereas the original database only contains placenames. This requires that they are linked to a suitable data source. This step is necessary in any case in order to make the data not just open but also actually linked.

Since there are 478 locations and 201 partners, we used an automated process to look up the names used in the database in the GND. As expected, many of the names are ambigous, whereas others are not found at all. Here are some examples, illustrating the variation with respect to GND IDs found for some treaty locations:

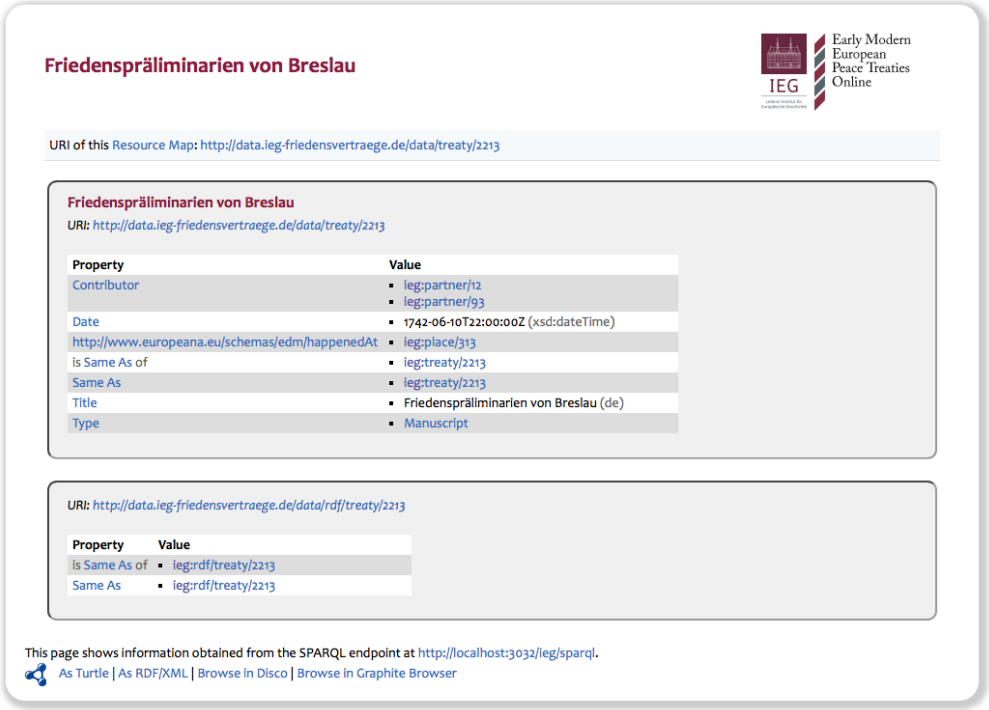

Figure 2: Screenshot of Pubby

ID	Name	GND
33	Altranstädt	4079738-7
4	Hubertusburg	5119515-X, 5119512-4
91	Malmö	4114951-8
6	Tyrnau	10172490-1, 500513-9, 10179031-4, 4555737-8, 7582117-5, 4696473-3, 1044374594, 4078490-3, 4696475-7

It is clear that such ambiguities cannot be resolved automatically, in particular, there is no guarantee that the correct entity is in fact among those found. We did not have resources in the project to perform manual resolution of location and partner links. However, it is one of the advantages of nanopublications that data can be qualified with provenance information, so that, for example, data added by automatic processes is given a lower certainty than data added by a human expert. The next section talks about nanopublications in some more detail.

2.2 Peace Treaties as Nanopublications

The second main goal of this project was to to explore the application of this approach to research in the humanities and to represent the key metadata about peace treaties (date, place, signatories, powers, type of treaty, etc.) as nanopublications. One important advantage of nanopublications is that they allow for associating provenance information with claims, which, in turn, also helps dealing with uncertain, unconfirmed, or conflicting claims.

Figure 5 shows an example of a nanopublication. It states—in the *assertion* part—that that treaty

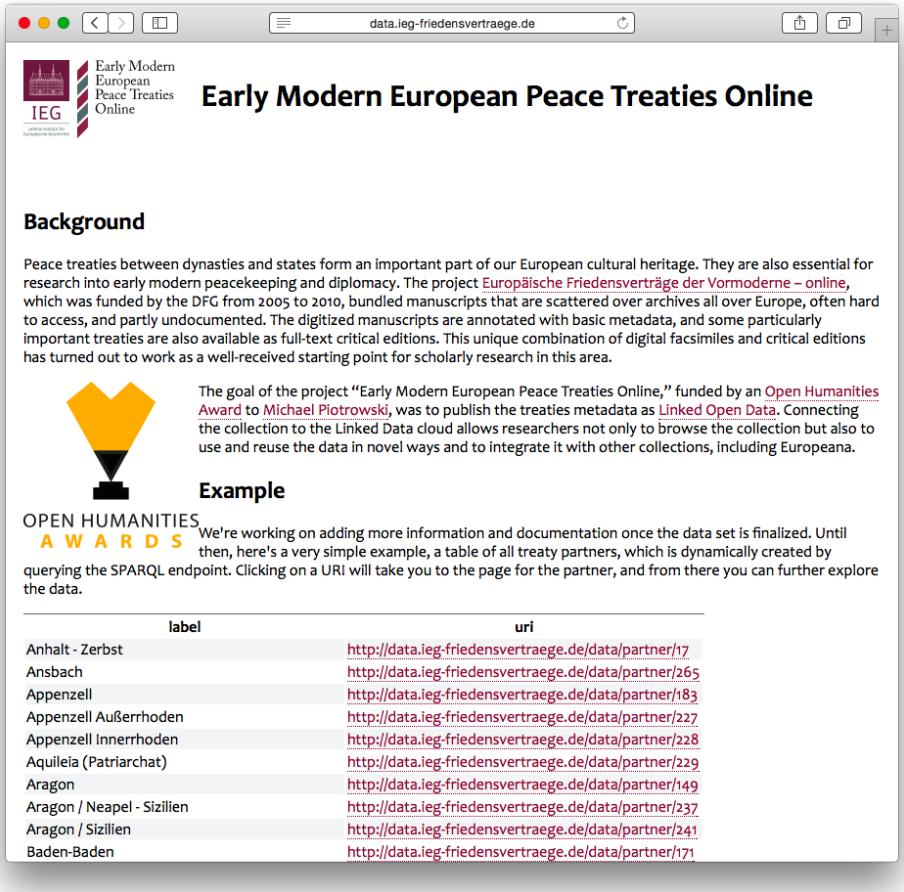

Figure 3: Screenshot of the `http://data.ieg-friedensvertraege.de/` home page

partner 12 ("Austria") is identical with the entity with the GND ID 10105216-9. The *provenance* part documents the provenance of this assertion; in this case, it was automaticall created by a script called *autolinks*, which automatically adds potential links to the GND. The *pubinfo* part, finally, contains metadata about the nanopublication as a whole.

As it is known that the assertion was generated automatically, it can be treated with the appropriate caution. In fact, while the GND ID given here does refer to an entity called "Austria," it is actually the British zone of occupied Austria, 1944–1955, which clearly cannot be the correct reference in the context of our early modern peace treaties.[4] This is, however, not a problem: First, as the provenance information is given, it is possible to filter data on the basis of this information; second, the nanopublications approach makes it possible to explicitly refute this assertion by another nanopublication.

The Nanobrowser by Kuhn et al. (2013) implements a user interface to nanopublications that specifically supports this type of interactions, so that a researcher can easily reject or support assertions. We are in close contact with the author of the Nanobrowser (which is open-source software) and are working on adapting it to our needs.

[4] A more likely reference would be either 4043271-3 "Austria" or 4075601-4 "Archduchy of Austria."

Figure 4: Mockup of a map display for treaty locations

```
@prefix nanopub: <http://www.nanopub.org/nschema#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix prov:    <http://www.w3.org/ns/prov#> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .
@prefix owl:     <http://www.w3.org/2002/07/owl#> .
@prefix :
 <http://data.ieg-friedensvertraege.de/nanopub/> .

:/1/head {
  :/1 a nanopub:Nanopublication ;
    nanopub:hasAssertion       :/1/assertion ;
    nanopub:hasProvenance      :/1/provenance ;
    nanopub:hasPublicationInfo :/1/pubinfo .
}

:/1/assertion {
  <http://data.ieg-friedensvertraege.de/data/partner/12>
  owl:sameAs <http://d-nb.info/gnd/10105216-9> .
}

:/1/provenance {
  :/1/assertion prov:wasGeneratedBy :pfeffer/autolinks/gnd/v0.1 .
}

:/1/pubinfo {
  :/1 dcterms:creator <http://d-nb.info/gnd/102572836X> ;
    dcterms:created "2015-01-28T11:32:30.758274Z"^^xsd:dateTime ;
    dcterms:rights <https://creativecommons.org/publicdomain/zero/1.0/> ;
    dcterms:rightsHolder <http://www.ieg-mainz.de/> .
}
```


Figure 5: Example nanopublication

Due to problems finding qualified personnel at the outset of the project, we lost about a month, and with the holiday season before the end of the project we lost more time. We thus could not complete the work on nanopublications during the allotted time. However, we have laid important foundations, and we continue to pursue this line of research even after the end of the project.

3 Conclusion and Outlook

In the project funded by the Open Humanities Award, we have converted the metadata of Early Modern Peace Treaties Online from a relational database into RDF and made it available as Linked Open Data. While we could not finish our work on nanopublications in the time frame of the project, we were able to lay the groundwork, and we are currently setting up the Nanobrowser and the tools for generating nanopublications from the RDF triples that we produced in the project.

For the work described here, I contracted with Prof. Magnus Pfeffer of the Stuttgart Media University (HDM). We are continuing the work on nanopublications as an unfunded research project. Prof. Dr. Kai Eckert, who previously worked in DM2E, and who has now also joined HDM, will now team up with us and contribute valuable experience from DM2E. We are currently working on a joint peer-reviewed paper, which will also cover the use of nanopublications.

References

Groth, Paul, Andrew Gibson, and Jan Velterop (2010). The anatomy of a nanopublication. *Information Services and Use*, 30(1–2):51–56. doi:10.3233/ISU-2010-0613.

Heßbrüggen-Walter, Stefan (2013). Tatsachen im semantischen Web: Nanopublikationen in den digitalen Geisteswissenschaften? In Peter Haber and Eva Pfanzelter, eds., *Historyblogosphere*. München: Oldenbourg. doi:10.1524/9783486755732.149.

Kuhn, Tobias, Paolo E. Barbano, Mate L. Nagy, and Michael Krauthammer (2013). Broadening the scope of nanopublications. In *Proceedings of the 10th Extended Semantic Web Conference (ESWC 2013)*, vol. 7882 of *Lecture Notes in Computer Science*, pages 487–501. Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-38288-8_33.

Penzholz, German and Andrea Schmidt-Rösler (2014). Die Sprachen des Friedens – eine statistische Annäherung. In Johannes Burkhardt, Kay Peter Jankrift, and Wolfgang E. J. Weber, eds., *Sprache. Macht. Frieden*, pages 311–322. Augsburg: Wißner. URL `http://www.uni-augsburg.de/institute/iek/scripts/Sprache_Macht_Frieden_Penzholz-Schmidt-Roesler.pdf`.

The research described in this report was funded by the DM2E project as part of the Open Humanities Awards.