

Europäische Friedensverträge der Vormoderne in Daten (FriVer+)
Project Documentation

Version 2
16.01.2024

By Dr Jaap Geraerts

Content

Introduction and background

Europäische Friedensverträge der Vormoderne online

- 1) Structure of the SQL database

Europäische Friedensverträge der Vormoderne in Daten

- 2) TEI XML Schema
- 3) Extraction and export
- 4) Cleaning and enriching the data
- 5) Transforming the data to XML
- 6) Creating and amending individual XML files
- 7) Accessing the data

Lessons learned

- 8) Recommendations for future DH projects at IEG Mainz

Introduction and background

This documentation describes the various activities and editorial decisions conducted and made in the context of the *Europäische Friedensverträge der Vormoderne in Daten* (FriVer+) project. This project was generously funded by the [NFDI Text+](#) and ran in 2023 at the [Leibniz Institute of European History](#) (IEG) in Mainz. It details the various steps that were taken regarding the extraction and export of relevant data from the SQL database that was created as part of the legacy project [Europäische Friedensverträge der Vormoderne online](#) (more about which below), the creation of a TEI XML schema for FriVer+, and the subsequent transformation of the exported data into valid XML in accordance with this schema. It, too, specifies the various editorial decisions that were taken in this process as well as the ways in which the original data was enhanced. Lastly, it provides a short overview where the *Europäische Friedensverträge der Vormoderne online* (henceforth EFVO) and FriVer+ data can be found.

In the end, this documentation is intended for internal and external use. On the one hand, by outlining the various (editorial) decisions that were taken during the project it hopes to be of aid for the design and execution of future DH projects at the IEG Mainz. On the other hand, by providing an overview of the structure of the FriVer+ data and the editorial decisions that underpin it, it aims to facilitate the use and reuse of this data by other scholars. Ultimately, the hope is that the FriVer+ project and its data can make a meaningful contribution to the study of peace settlements in the early modern world.

Virtually all FriVer+ data is derived from the legacy project EFVO. This project, which was funded by the Deutsche Forschungsgemeinschaft (DFG), ran from 2005 until 2010 at the IEG. It gathered (meta)data about more than 1800 peace treaties that originated in the period 1450-1789 and, in addition, it included 40 scholarly editions and 121 transliterations of the text of specific treaties. All this information was captured and stored in a SQL database (see section one) that powered a website with a search interface through which this data could be accessed. The fact that it was not that easy to get to the raw data – the text of editions had to be copy-pasted from the website into a text document, to mention but one example – was a main impetus for the FriVer+ project.

After the EFVO project had come to an end, some data was added on an ad-hoc basis, e.g. the language in which a treaty was written. Unfortunately, these and other decisions taken during the project were not documented, reflecting a customary practice in earlier DH projects that luckily seems to have been superseded. The lack of proper documentation constituted an

obstacle at the start of the FriVer+ project and necessitated a substantial effort aimed at understanding the data generated by the EFVO project. This documentation, although primarily aimed at explaining the FriVer+ project, provides information about the earlier legacy project as well.

1. Structure of the SQL database

The structure of the database created by the EFVO project, a schematic overview of which can be found in Appendix A, basically comprises two separate databases that do not “talk” to each other. In the tables that start with the prefix “IEG_publications_” the metadata about the publications listed in the “[Publikationsportal](#)” of the EFVO website are given. The text of these publications is displayed on this website as well and are stored in the SQL database as HTML. For example, some of the data pertaining to an article by Andrea Weindl, available [here](#), is stored in the following manner:

```
<p>[head1:Einleitung:head1][note:Überarbeitete Version des Artikels <em>The
Asiento de Negros</em> and International Law, in: Journal of the History of
International Law 10 (2008), S.229-257; mit freundlicher Genehmigung der
Herausgeber.:note]<br />
&nbsp;</p>
```

The metadata pertaining to these publications, such as the historical period(s) they cover, the language in which they are written, their genres, etc, are stored in several tables. All the information in these table feeds into the central table, IEG_publications.

Although not identical, the structure of the “second” database, pertaining to the early modern peace treaties themselves, mimics the structure of the “publications database.” A series of tables comprises information regarding the archives where the treaties can be found, the locations where they were signed, the languages in which they were written as well as the “partners,” i.e. the signatories of the treaties, and so forth. In other tables, however, more extensive information is to be found. For example, in the table IEG_treaty_pages, the file names of the digital surrogates of the pages of individual peace treaties are given as well as the page number and the text of transliterations associated with that digital image. The text is marked up in HTML, as the following example of a part of a transliteration shows:

```
<br /> <br /> <br /> <br /> <br /> <br /> <table summary="" class="htmltableborders"
style="border-style: none; display: table; float: right; width: 400px; background-image:
none;" border="0" cellpadding="1" cellspacing="1" frame="void"
rules="none"><tbody> <tr> <td><font color="#990000"><span class="trans14">
Sa Majest&#233; le Roi de Prusse et Sa Majest&#233; <br /> le Roi de Pologne
Electeur de Saxe
```

In the main table, the IEG_treaty table, information such as the archive, language, location and parties associated with individual treaties are gathered, based on the primary keys established in the abovementioned tables that record this information. Metadata pertaining to existing anthologies of particular treaties as well as their digital surrogates are captured in the table IEG_treaty_anthologies and IEG_treaty_anthologies_pages. Of great importance is IEG_treaty_comments table, which gathers the contextual information regarding individual peace treaties, such as a description of a treaty's contents, a selection of available literature, the historical contextual in which it originated, and so forth. The text of the editions of individual treaties is stored in this table as well. All the text is marked up in HTML so that it can be swiftly displayed online. As mentioned, the schematic overview in Appendix A shows how the tables are related to one another.

Uncertainties and editorial decisions

The SQL database also comprises 310 treaties of which the status is set to “inactive.” What this exactly means is unclear and the lack of proper documentation makes it impossible to verify. However, according to the *communis opinio*, work on these treaties had not been finished. Because it is impossible to check where the work had started and ended, we decided not to include these treaties in the data provided by the FriVer+ project. Likewise, one treaty is said to be a “beta-version” and, due to persistent uncertainty about what this qualification precisely entails, it has not been included either. Otherwise, all the other relevant data from the SQL database has been extracted and exported. The details of this process can be found in section three.

2. TEI XML Schema

After having obtained a clear overview of the (structure of the) data contained in the SQL database, we started thinking about the XML schema that could provide a new structure for the transformed data. Already at the very inception of the project, it was decided that we would work with the guidelines of the [Text Encoding Initiative](#) (TEI), simply because TEI is widely adopted and using it would therefore further enhance the accessibility and interoperability of the FriVer+ data.

One of the major advantages of the TEI guidelines, its comprehensiveness, at the same time is one of its drawbacks. For even a [TEI Lite](#) schema, a heavily trimmed-down version of the TEI schema, still comprises hundreds of lines. We had the advantage though, that we did not need to create an intuitive and straight-forward schema that should enable other people to create transcriptions manually, but rather one that could handle our transformed data. As a result, the existence of a large number of superfluous elements and attributes was unproblematic for our purposes.

In the end, we created and slightly modified a TEI schema by making use of [Roma](#), an ODD editor,¹ which allows one to select and customize the various building blocks of a TEI schema. The schema used by *Europäische Religionsfrieden Digital* (EuReD), an allied project which deals with data very similar to our own, offered a powerful example of a TEI schema for similar data. Ultimately, also with the aim to foster the exchange of data between the FriVer+ and EuReD projects, our schema mimics theirs to a large extent.

In order to accommodate our data, a few additions as well as minor departures from the original TEI schema were introduced:

- the <repository> element is now a child element of <MsDesc>
- the “role attribute” of <person> has a new value, “signatory” (in order to make a distinction between the people or parties signing a treaty and those mentioned in the edition of a treaty’s text)
- the “type” attribute of the <div> element has new values (Ausfertigungen, Inhalt, etc) that reflect the different types of contextual information about peace treaties

¹ ODD stands for “One Document Does-it-all” and is a “TEI XML format used to express schema fragments, prose documentation, and reference documentation for any XML markup scheme as a single document” (<https://tei-c.org/guidelines/customization/getting-started-with-p5-odds/>).

- a distinction can be made between the standardized date, captured in the “when” attribute of the <title> element and the original date on the treaty, specified in the “when-custom” attribute

The [resulting schema](#) is able to incorporate all the FriVer+ data. All the relevant information pertaining to the FriVer+ project as well as the particular treaty is stored in the <teiHeader>. For instance, the treaty’s ID, its date, and its standardized name, can all be found in the child elements of the <titleSmt> element. Likewise, the name and the (internal) ID of the repository where a copy of the treaty can be found and, in some cases, links to digital surrogates of primary source material, are captured in the <sourceDesc> element.

StandOff element

The <standOff> element contains specific information relating to a treaty, including the names of its signatories and, in case of editions, the people mentioned in the edition’s text, all of which is captured in the <listPerson> element. The signatories of the peace treaties as well as the people mentioned in the text of editions are captured in this list and the distinction between them is made in the role element (the value of which is either “signatory” or nothing). In addition, the different IDs of the people tagged reflect whether these were tagged during the EFVO project or as part of the FriVer+ project (more about which below). All the people that are tagged, are linked to online resources such as the GND Explorer or Wikidata. For example:

<listPerson>

```
<person xml:id="pp0017">
  <persName ref="https://explore.gnd.network/gnd/116754974">
    <name>Ewald Friedrich Graf von Hertzberg</name>
  </persName>
</person>
```

</listPerson>

In the text of an edition, this person is tagged as follows:

```
<rs ref="#pp0017">Ewald Frédéric de Hertzberg</rs>
```

It should be noted that next to people, the listPerson comprises countries or other political bodies (or people representing such entities). The EFVO project did not make an (ontological) distinction between the signatories of peace treaties and neither have we.

In addition, the <listPlace> element comprises the location where a treaty was signed. Just as with the people, the places that have been tagged are linked to online resources. Lastly, if there are digital surrogates of a peace treaty available, then this information is stored in the <graphic> element, a child element of the <facsimile> element (see below for an example).

Front element

The <front> element, a child element of the <text> element, contains contextual information that is associated with a treaty. This information, which can include a summary of a treaty's content or a (short) description of the historical context in which a treaty originated, was provided by the scholars working on the EFVO project. By making use of the <div> element in combination with the "type" attribute, a distinction is made between the various sorts of information about a treaty that is provided. For example:

```
<div type="Drucke">  
<div type="Inhalt">  
<div type="Kommentar">
```

This division corresponds to the [various headers on the page of an individual treaty on the EFVO website](#).

Body element

The <body> element, another child element of the <text> element, is where the text of editions or transliteration is located. Again, the "type" attribute of the <div> element indicates whether the text is part of an edition or transliteration. If a connection between a digital surrogate and (transcribed) text is made in the EFVO project, these links are retained and made operable in the <pb> element.

```
<pb facs="#FacsPnr1"/>
```

This refers to a digital surrogate that is specified in the <facsimile> element that forms part of the <standoff> element, as mentioned above.

```
<facsimile>  
  <surface xml:id="FacsPnr1">  
    <graphic url="https://www.ieg-  
friedensvertraege.de/dbimages/img_4383352bcfbb7"/>  
</facsimile>
```

It should be noted that some transliterations, although containing numbered page breaks, are not linked to digital images. In these cases, the XML file does not contain the <facsimile> element nor its child elements. Lastly, although the FriVer+ data contains 40 editions that are often accompanied with digital surrogates of the treaty, unlike transliterations, the text has not been divided in pages and linked to the corresponding digital surrogates.

Back element

Last but not least, the <back> element consists of the <listBibl> element where biographical details of studies related to a particular peace treaty are given. Moreover, the <biblScope> element can contain links to (the URL of) a digital anthology (or anthologies) of a peace treaty. For example:

```
<listBibl>
  <bibl xml:id="a80">
    <title level="a">Theatrum Europaeum (1697-09-20)</title>
    <title level="m">Theatrum Europaeum, Band 15 (1696-1700) Aus:
Dokumentenarchiv der Universitätsbibliothek Augsburg (Drucke vor 1900 /
Historische Geschichtswerke).</title>
    <biblScope>
      <graphic xml:id="A1Pnr1"
url="https://www.ieg-friedensvertraege.de/dbimages/img_49e6f563a7bd5"/>
    </biblScope>
  </bibl>
</listBibl>
```

The value in the xml:id attribute refers to the number of the row in the treaty_anthologies table in the SQL database (since the anthologies were not given a unique ID in the EFVO project).

External DTD

A number of special characters are defined in an external DTD to which each of the transcriptions generated by the FriVer+ project refers. The DTD can be accessed [here](#).

3. Extraction and export

Ideally, all the data could have been exported in the correct format at once, but this proved to be difficult for various reasons. First of all, in some cases a simple export did not work because the sheer quantity of data which caused, whatever the exact reason might have been, that portions of the data were not included in the export files. Secondly, when exporting the relevant data from the SQL database to CSV files, errors occurred due to the fact that the database comprises large amounts of text marked up in HTML. Thirdly, because of the ways in which data was stored in the tables in the SQL database, it needed to be transformed and stored in temporary tables before it could be exported.

Exporting the meta-data pertaining to each individual treaty was relatively straight-forward and was done via a simple export script. (N.B. All the scripts created by the FriVer+ project can be accessed [here](#).) As mentioned, we decided to include only “active” treaties – the treaties that are finalised and accessible via the EFVO website – because of the persistent uncertainty about the exact status of the “inactive” treaties. The metadata was exported to a CSV file where the data about each treaty was stored in a single row across several columns.

It was also decided that the data in the table “cascading_partners,” in which IDs of signatories of peace treaties were linked to other signatories that were connected somehow (neither the scant documentation nor the database provides the exact content of this connection), was not to be exported and transformed. This decision was taken mainly because the connection between signatories was not always clear and because this connection changed over time, as countries were conquered, changed rulers, and so forth – historical developments not captured in the SQL database. Therefore, since the data in this table was confusing more than it was enlightening, it was decided to exclude it.

Storing all the data pertaining to each treaty in a single row would make it easier to transform it into XML and this required us to transform the data in the SQL database and store it in temporary tables. For unlike the metadata in the SQL database, the contextual information of each peace treaty tended to be stored in different rows (see Figs. 1 and 2).

id	ts	treaty	scan	scan_rescaled	transliteration	transliteration_on_scan	page
30	2005-11-26 20:15:32	3	dbimages/img_4388b45038d57	dbimages/img_4388b4503d44a			9
29	2005-11-26 20:14:41	3	dbimages/img_4388b41dcad98	dbimages/img_4388b41dcfb88			8
28	2005-11-26 20:13:56	3	dbimages/img_4388b3f10edcb	dbimages/img_4388b3f113be8			7
27	2005-11-26 20:12:58	3	dbimages/img_4388b3b629b78	dbimages/img_4388b3b635ec6			6
14	2005-11-22 16:11:43	3	dbimages/img_4383352bcbfb7	dbimages/img_4383352bd49d7	Sa Majesté le Roi de Prusse et Sa Majest...	 ...	1
15	2005-11-22 20:24:27	3	dbimages/img_43837066b4e08	dbimages/img_43837066b9c88	lesquels après s´être diueme...	 <...	2
16	2005-11-22 20:29:06	3	dbimages/img_4383717ec5f77	dbimages/img_4383717ecad98	moindre atteinte. Sa Majesté le Roi de Pr...	 <table summary...	3
17	2005-11-22 20:30:07	3	dbimages/img_438371b790419	dbimages/img_438371b795238	compris la Haute et Basse Lusace. Si les ordres ...	 <table summary...	4
18	2005-11-22 20:31:04	3	dbimages/img_438371f4c3868	dbimages/img_438371f4c8688	Article III Sa Majesté le Roi de Pr...	 <table summary...	5
40	2005-11-28 15:01:50	3	dbimages/img_438b0dca8c688	dbimages/img_438b0dcacfc1f			10
41	2005-11-28 15:02:40	3	dbimages/img_438b0dfc84c7c	dbimages/img_438b0dfc88ee7			11
42	2005-11-28 15:03:24	3	dbimages/img_438b0e2918a08	dbimages/img_438b0e2918d28			12

Fig. 1 Detail of the treaty_pages table

id	ts	treaty	language	title	comment	sort
3	2020-11-25 12:32:54	3	de	Ausfertigungen	<p>Graphic: AGADWarschau, Zbiór Dokumentów Pergaminowych Nr. 1439. </p> <p>Unterhändlerinstrument: GSABerli...	3515
4	2005-11-22 21:49:06	3	de	Drucke	CTS 42, S. 361(363)-(378)386, Wend, Codex 3, S. 380-395 (frz.). Faber, Neue Europäische Staats-Cantzley 9, S. 417 (f...	3516
61	2005-11-29 13:18:45	3	de	Literatur	Beaulieu-Marconnay, Der Hubertusbürger Frieden. Schmidt, Die Meißner Vorverhandlungen zum Hubertusbürger Frieden, i...	3573
93	2005-12-02 00:47:46	3	de	Edition	Sa Majesté le Roi de <acronym title="Friedrich II., genannt "der Große" (1712-1786), seit 1740 König in (seit 1772 von) Preußen...	3513
376	2006-01-18 11:03:32	3	de	Inhalt	<p>Art. 1 Über den Frieden. Freundschaftsbeziehung. Art. 2 Beendigung der Feindseligkeiten. Herstellung der Waffen...</p>	3574
377	2006-01-18 11:04:10	3	de	Kontext	1740/42-1745/48: Österreichischer Erbfolgekrieg und zeitgleich Erster Schlesischer Krieg (1740/42). Der Vorfriede von Breslau (...</p>	3575
378	2006-01-18 11:04:51	3	de	Kommentar	Mit dem Einmarsch in Sachsen am 29. August 1756 reagierte Friedrich II. auf Angriffspläne Rußlands und <acronym title="Frederick the Great" (1712-1786)...	3576

Fig.2 Detail of the treaty_comments table

Through a two-step process, various [pivot tables](#) were created in the SQL database, basically storing the information related to one treaty that was spread across several rows as values in different columns in a single row. After the creation of the pivot tables, the data could then be exported to a CSV file. This data, together with the other data that was exported to CSV files, was reassembled in a large Excel file, a master file where the data was cleaned and enriched (see section 4, below). This approach could be applied to information about the digital surrogates and the anthology pages that is captured in the SQL database.

However, as already mentioned, some data in the SQL database defied easy exportation simply because exporting large quantities of text marked up in HTML to CSV files turned out to be a rather unhappy combination. By making use of the same workflow described above, the data in the `treaty_comments` table was first put in a pivot table (see Figs. 3 and 4) and thereafter directly exported to a large XML file.

[illegible]

Fig. 3 Pivoted table comprising the data in the original treaty_comments table

treaty	html_area	Ausfertigungen	Drucke	Edition
3	Archiwum Główny Akt Dawnych Warszaw, Zbió...	<p>Graphik: AGADWarschau, Zbiór Dokumentó...	CTS 42, S. 361(363)-(378)386. Wendk, Codex ...	Sa Majesté le Roi de <acronym title="Friedrich I...
8		AAE Paris: M.A.E. Traités; </spa...	CTS, Vol. 21, S. (371), 3...	NUL
11		[Transkriptionsgrundlage:] HHStAWien AUR 161...	Gooss, 437-456. Monumenta Hungariae Histori...	NUL
12		[Transkriptionsgrundlage:] HHStA Wien: A. Rati...	Roderich Gooss, Österreichische Staatsvertrá...	NUL
13		<p>[Transkriptionsgrundlage:] HHStAWien AU...	<p>Gooss, S. 484-493. Dumont V/2 S. 358-359....	NUL
14		<p>[Transkriptionsgrundlage:] HHStAWien Aur...	Gooss, S. 545-554. Dumont V/2, S. 407-408 (A...	NUL
15		<p>HHStAWien: A. Aur 1624 IV 22: Rev...	<p> Gooss, Roderich, Österreichische Staatsve...	NUL
16		<p>Unterhändlerinstrument: HHStAWien [kein ...	Roderich Gooss, Österreichische Staatsvertrá...	NUL
17	Archivportal N...	<p>Unterhändlervertrag: Nordrhein-We...	<p>Akten und Urkunden zur Außenpolitik Christ...	NUL
18	Archivportal N...	<p>Nordrhein-Westfälisches Staatsarchiv Müns...	<p>Akten und Urkunden zur Außenpolitik Christ...	NUL
19		Nordrhein-Westfälisches Staatsarchiv Münster ...	NUL	NUL
21		<p>BHSAMünchen Bayern Urkunden 1086/1: H...	NUL	NUL
22		<p>Nordrhein-Westfälisches Staatsarchi...	Dumont, VII/2, S. 347-351.	NUL
23	Archivportal N...	<p>Nordrhein-Westfälisches Staatsarchiv Müns...	Max Braubach (vgl. Literatur), S. 169-172. Secr...	NUL
24		Nordrhein-Westfälisches Staatsarchiv M&...	Braubach [vgl. Literatur], S. 177-180. Resolutie...	NUL
25	Stockholm, Riks...	Schwedisches Reichsarchiv, Stockholm: Original...	NUL	NUL

Fig. 4 Detail of the pivoted table comprising the data in the original treaty_comments table

In a similar fashion, all the information pertaining to transliterations was exported to another large XML file. Luckily, software such as [Liquid Studio](#) allows one to open and work with such large files (whereas other XML editors are prone to crash when working with files of this size). At this stage of the process, then, the extracted data was stored in three sources: 1) an Excel file (the master file) in which most of the treaties' metadata was stored; 2) a large XML file containing the contextual information, the <note> element, a child element of <repository>, the <msDesc> element and the information contained in it, and the text of the editions (i.e. the information stored in the treaty_comments table); 3) the transliterations (i.e. the data stored in the treaty_pages table).

4. Cleaning and enriching the data

The decision to gather and reassemble the exported data in a master file (Excel) was taken because the data needed to be cleaned and enriched before it could be transformed into valid XML. The need to enhance the quality of the data was informed by both a couple of editorial decisions taken by the FriVer+ project and the realization that the original EFVO data could be refined and polished further. To start with the former, we decided to try to capture all the people mentioned in every edition in the EFVO corpus (rather than just the parties that signed a treaty). This was possible largely due to the diligent work of the members of the EFVO team, who had provided information in footnotes about most of the people mentioned in the text. This greatly sped up the process of tracking, tracing, and tagging these people, who were “defined” in the `listPerson` element and tagged in the text, as explained in section 2, above. Where possible, links to the records of these people in online resources such as the [GND Explorer](#) and [Wikidata](#) were included. Similar links were also added to the people (the signatories) and the geographical locations that were tagged in the EFVO project. In order to distinguish them from the signatories of peace treaties that were tagged by the EFVO team, the IDs of the persons tagged in the FriVer+ project start with “pp” (e.g. pp0017) instead of with one “p” only (e.g. p0001). In the end, more than 300 people were added by the FriVer+ project, and they are listed in [this CSV file](#).

Other interventions arose from the need to be more precise than the EFVO project has been. For example, in the original data, one of the signatories of many a peace treaty was the entity “Kaiser” (id=61), a reference to the Holy Roman Emperor. A fair number of Holy Roman Emperors were grouped under this entity, without specifying which one had signed a particular treaty. Based on the date at which a treaty was signed, the FriVer+ data makes a distinction between fourteen emperors. To offer another example, in the SQL database Spain and the Spanish Netherlands were joined in one ID. In the FriVer+ data, a distinction is made between these countries (which could be done by making use of the different IDs for these countries that already were existent in the EFVO data, namely IDs 34 and 104).

The way in which the language of a treaty is captured, was changed as well. Whereas the SQL database comprised a simple ID consisting of a number followed by the name of a language (e.g. “latein”), this was changed in the master file and hence in the resulting XML transcriptions, where this information has been given in accordance with the TEI guidelines, namely:

```
<language ident="la" xml:id="lan6">Latein</language>
```

Here the xml:id reflects the IDs in the SQL database (while the prefix “lan” has been added as XML requires IDs to start with a letter). In general, all the IDs in the SQL database have received a prefix in order to turn them into valid IDs in XML. The letter or letters signify the kind of ID we are dealing with (so “lan” for “language,” “loc” for “location,” “t” for “treaty,” and so forth). Other changes to the data were related to making links to digital images operable, simply by including the prefix “https://www.ieg-friedensvertraege.de/”, resulting in, for example: https://www.ieg-friedensvertraege.de/dbimages/img_4a7c27c15fb66. Moreover, in the SQL database, the one-to-many relation between a peace treaty and its signatories was established in one cell (which contained the IDs of the signatories separated by a comma). This was amended in the “master file”: the information about each individual signatory (or “partner” as they were called in the original database) is stored in a one row and divided among several columns. This makes the data more amenable to transformation into XML.

After the necessary changes and additions were made to the data in the “master file,” changes were made to the data in the two large CSV files as well. These amendments mainly consisted of renaming elements and, occasionally, attributes, as well as renaming or deleting HTML tags that could not be retained in the XML transcriptions. In sum, the work on the large XML files was undertaken with the aim to speed up the process of assembling the individual XML transcriptions, the topic of the next section. In some of these XML transcriptions, the original HTML has been put in a comment field, just to remind our users of the underlying transformation of the data. For example, in the file 8.xml one can find:

<!--JG: in the original HTML, the aforementioned names were captured in a table.

```

<table class=" htmtableborders" style="WIDTH: 100%" cellpadding="1" border="0"> <tbody>
  <tr>
    <td style="WIDTH: 50%">&nbsp;NA Harlay Bonneuil.
  </td>
    <td style="WIDTH: 50%">&nbsp;A. Heinsius
  </td>
  </tr>
  <tr>
    <td style="WIDTH: 50%">&nbsp;Verjus De Crecy.
  </td>
    <td style="WIDTH: 50%">&nbsp;De Weede
  </td>
  </tr>
  <tr>
    <td style="WIDTH: 50%">&nbsp;De Callieres.
  </td>
    <td style="WIDTH: 50%">&nbsp;W.V. Harren
  </td>
  </tr></tbody>
</table>
-->

```

5. Transforming the data to XML

It took a bit of time to decide on how to transform the data. One could, for instance, use proprietary software such as Altova Mapforce in order to create (complex) transformation scenarios. However, this meant having to familiarise myself with new software packages, something which can be deceptively time consuming. Moreover, I knew that, because of the fact that the exported data was spread across three files, namely the Excel “master file” and the two large XML files, that I had to reassemble the data manually. Lastly, every XML transcription had to be checked because of quality control. In the end, therefore, it made more sense, at least to me, to settle for a solution that was relatively quick, straight forward, and bullet proof.

Hence I decided to use an [online and free convertor](#) in order to transform the data in the “master file” into valid XML. The online convertor at Freeformatter.com allows one to provide an XML template in which can be detailed and specified where which information should come. The template specifies the structure of the output file (the elements and attributes) and refers to the number of a column in the CSV file (the import file), the source where the convertor derives its data from. Here are, for example, the first lines from the one of the output templates (two output templates were created and used, based on some differences in the data such as the number of facsimile images that were associated with particular treaties).

```
<titleStmt>
  <title level="s">European peace treaties of the pre-modern era in data / Europäische
    Friedensverträge der Vormoderne in Daten (FriVer+)</title>
  <title when="##3##" xml:id="##1##" when-custom="##5##" level="a"
ref="##445##"
    >##6##</title>
</titleStmt>
```

After the transformation, the resulting output looks like this:

```
<titleStmt>
<title level= "s">European peace treaties of the pre-modern era in data / Europäische Friedensverträge der
Vormoderne in Daten (FriVer+)</title>
<title when="1654-07-30" xml:id="t1104" when-custom="1654 VII 30" level="a" ref="https://www.ieg-
friedensvertraege.de/treaty/t-1104-1-de.html">Beitritt zum Friedensvertrag von Westminster (1654 IV
15)</title>
</titleStmt>
```


6. Creating and amending individual XML files

For each individual treaty, the output generated by the convertor was copied into an XML document. Thereafter, the assembly process went into the next stage: the data pertaining to a specific treaty stored in the two large XML files was copied and added to the same XML document. It should be noted that these two files only contain data relating to a select number of treaties such as editions and transliterations. In most cases, the EFVO project only recorded the metadata of treaties, information that was stored in the “master file.” In other words, for most of the treaties the process was extremely straight-forward: transform the data (this was done in batches of several treaties at the time), copy it to an XML document, check the validity of the XML, and move on to the next treaty. For sure, this work is mechanic, repetitive, and not a little tedious, but within the limited (temporal) scope of the FriVer+ project, it made most sense to proceed in this manner.

In general, after the data transformation and the creation of the individual transcriptions, the work was done for the vast majority of the treaties. Only editions and transliterations required more additional work. The main reason was the discrepancy between the annotations in the HTML markup and the way in which annotations are created according to the TEI guidelines. Consider, for example, the text of one of the editions in our corpus:

```
<p>Convention entre S[on]. A[ltesse]. R[oy]ale Monseigneur le Duc de <acronym  
title="Wilhelm August, Herzog von Cumberland (1721-1765), zweiter Sohn des engl. Königs  
Georg II., General.">Cumberland</acronym> et <acronym title="Wenck: S.E.Mr.">S[on].  
A[ltesse]. M[onsieur]</acronym>. Le <acronym title="Danske Traktater: le Maréchal, Duc.  
>Duc</acronym> de <acronym title="Textvariante: Wenck: conclue à Closter-Zeven.  
Sachanmerkung: Louis François Armand de Vignerot Plessis, Herzog von Richelieu (1696-  
1788), Marschall. ">Richelieu</acronym> </p>
```

The HTML tag <acronym> was used to single out (some of) the people mentioned in the text as well as to point at the differences between the various editions of a given treaty. In some cases, the latter use was signified by the word “Textvariante,” but this was not done consistently across all the editions in the EFVO corpus.

```
<acronym title="Textvariante: Wenck: la Luhe. Sachanmerkung: Flußlauf der Luhe.">l'Aue  
</acronym>
```

The HTML markup needed to be changed in various ways in order to turn it into valid XML in accordance with the TEI guidelines. Unfortunately, the <acronym> tag could not simply be renamed as <note>, the element used in the TEI schema. For the note tag appears after the word or words it comments on, whereas the <acronym> tag envelops this word/these words, as the examples above show. Hence this needed to be done manually, which was a fair amount of work, given that some editions have hundreds of notes. We decided to make use of the “type” attribute of the <note> element in order to distinguish between notes that provide additional information (type = “annotation”) and between notes that signify textual differences (type = “crit_app”). In this way, the users of our data can differentiate between the types of notes our editions and transliterations comprise.

Other necessary changes consisted of the ways in which people mentioned in the edition’s text were tagged (see p. 7). This needed to be done manually as well and this was fairly time consuming as well, in particular because the identity of the people mentioned had to be established, the corresponding record in the GND Explorer or Wikidata had to be retrieved and saved, and so forth. In the end, the data in the XML transcription looks as follows:

```
<p>Convention entre S[on]. A[ltesse]. R[oy]ale Monseigneur le <rs ref="#pp0235">Duc
de Cumberland</rs><note n="1" type="annotation">Wilhelm August, Herzog von
Cumberland (1721-1765), zweiter Sohn des engl. Königs Georg II.,
General.</note> et S[on]. A[ltesse]. M[onsieur].<note n="2"
type="annotation">Wenck: S.E.Mr.</note> le <rs ref="#pp0236">
Duc<note n="3" type="crit_app">Danske Traktater: le Maréchal, Duc.</note> de
Richelieu</rs><note n="4" type="crit_app">Textvariante: Wenck: conclue à
Closter-Zeven. Sachanmerkung: Louis François Armand de Vignerot Plessis,
Herzog von Richelieu (1696-1788), Marschall.</note>
</p>
```

More infrequently, the <acronym> tag was used to indicate uncertainty:

```
<acronym title="Unsichere Lesart. CTS; Danske Traktater: d'Ost. Wenck: de l'Oste. Oste,
Flußlauf, der Zeven und Bremervörde verbindet und in die Elbe mündet.
">&lt;d'loste&gt;</acronym>
```

<acronym title="Unsichere Lesart."><quaeque></acronym>

Because the edition's text itself denotes this uncertainty, we could simply make use of the same <note> tag, which appears after the word or words the editor was not entirely sure about (rather than enveloping them).

In stark contrast to the editions, the notes in transliterations were indicated in this way:

respectifs à ¹[des dits S[ieu]rs Etats
Generaux]² de la part
des Etats de
Hollande et de Westfrise,

 <hr size="2" width="100%" />

The actual notes would follow later, after a section of the text and often in a separate paragraph:

<p>1 Nachträglich eingefügt über durchgestrichenem Text.
2 Nachträglich eingefügt,
links am Rand</p>

In order to maximize coherence across the XML transcriptions generated by the FriVer+ project, we opted to stick to the use of the <note> element. The transformed data looks as follows:

respectifs à<note n="1" type="crit_app">Nachträglich eingefügt über durchgestrichenem
Text.</note> l'Assemblée [des dits S[ieu]rs Etats Generaux]<note n="2"
type="crit_app">Nachträglich eingefügt, links am Rand.</note> de la part <lb/>des Etats de
Hollande et de Westfrise, <lb/> <lb/>

7. Accessing the data

The FriVer+ data is stored and can be accessed in different repositories. First of all, the data created by the EFVO project continues to be available on its [original website](#). Second, all the transcriptions generated by the FriVer+ project are hosted on the servers of the [University and State Library Darmstadt](#). All the transcriptions can be accessed and viewed online [here](#). The treaties are ordered chronologically and can be accessed via the drop-down menu under the header “Inhalt” in the left side of the screen (Fig. 5).

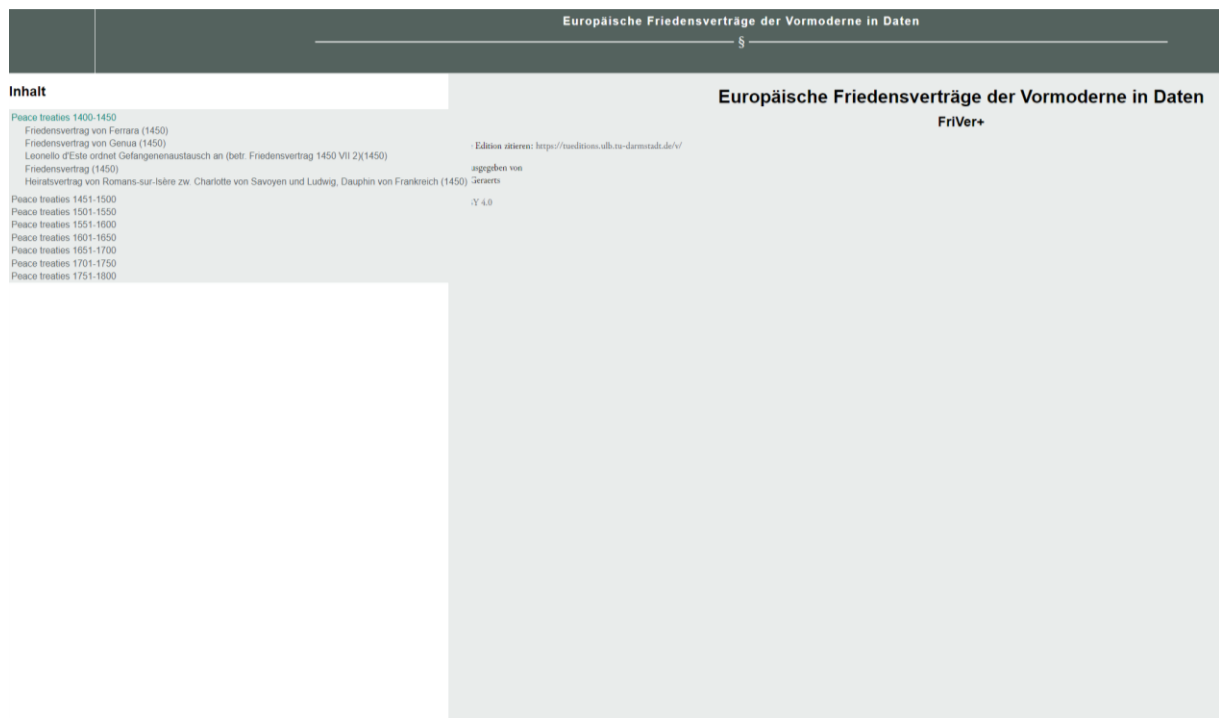


Fig. 5 Overview of the starting page of the FriVer+ project on the TU Darmstadt infrastructure.

When selecting a particular treaty, the browser shows the text of the transcription and, if existent, the digital surrogates of the treaty's text (Fig. 6).

[illegible]

The “navigation” button in the top left corner can be used to open other treaties in the corpus, whereas the “suche” button allows on to search the text of the transcriptions that are part of a certain project or resource (that can be selected from a drop-down button above the search bar) (Fig. 7). Alternatively, one can search the text of the transcriptions by entering a search term in the designated area in the upper right corner (Fig. 8). Apart from querying the transcriptions, it is possible to scroll through the images displayed on the right side of the screen and to manipulate them (zooming in and out, and so forth). Lastly, by clicking on the file name of a transcription in the lower left corner of the screen (Fig. 9), one can download it.

Fig. 7 Search function of the TU Darmstadt framework

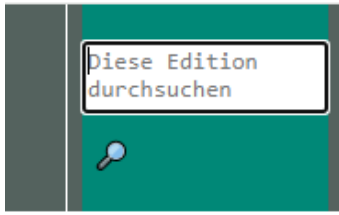


Fig. 8 Detail of the website's header at the TU Darmstadt framework

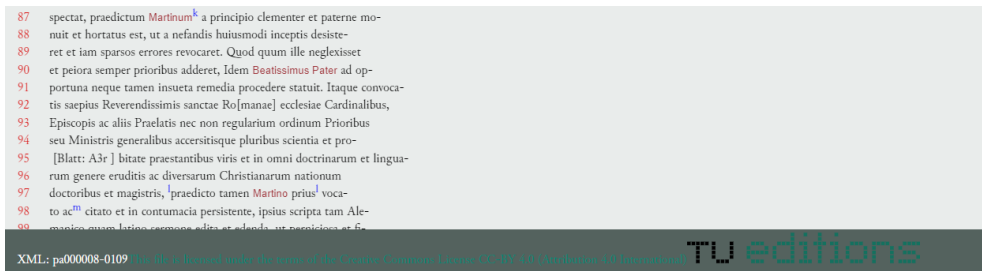


Fig. 9. Detail of the website's footer at the TU Darmstadt framework

Third, all the documents created by and during the FriVer+ project, including the documentation, the scripts, and the XML transcriptions, can be freely accessed and downloaded from the projects GitHub repository, which can be found [here](#). The scripts that have been used to amass and export the relevant data from the SQL database, as explained in section three, above. The documentation folder also contains the CSV file with the new parties that have been added to the original EFVO data.

An overview of the 1834 transcriptions as well as URLs to the different repositories where they are hosted can be found [here](#).

8. Recommendations for future DH projects at IEG Mainz

This section comprises a (brief) reflection on the lessons learned during the FriVer+ project. Lessons which may have a wider significance than this project alone, given that it is likely that the IEG Mainz, with its DH Lab firmly established now, will execute similar projects in the future. Because FriVer+ grew out of a legacy project, EFVO, this reflection comprises an evaluation of (aspects of) this earlier project. If, at times, the tone of this evaluation is somewhat critical, this should not be seen as a negative assessment of the EFVO. Times change, in particular in the digital humanities. Requirements which nowadays are part and parcel of digital humanities projects, were not so in the past. However, with a firm eye on the future, we can learn from the execution of past DH projects so as to enhance the design and implementation of future ones.

Lesson 1: Documentation is key

The importance of robust documentation, detailing the various aspects of DH projects, ranging from the editorial decisions taken to the software used, and everything in between, seems all too obvious now. However, in the fairly recent past, this aspect of DH projects was often neglected. Solid documentation is fundamental, in academia at large and certainly at an institution like the IEG, where the vast majority of members of staff are employed on temporary projects and recurrent brain drains as “knowledge leaves the building” are occurring. Unfortunately, EFVO is a perfect example of earlier DH projects in that the documentation it provides is scant, to say the least. Apart from an overview of the project that contains a reflection on the need of digital editions, available [here](#), there is not much more. However, although the entire project team has left the IEG in the meantime, luckily the library management at the IEG was involved with EFVO and managed to provide us with important information about this project. Hence we learned that information about the language in which the peace treaties are written was only added to the SQL database at a later stage and that this work was incomplete as at some point the funds to employ a HIWI for this work dried up.

The lack of robust documentation made itself felt in different ways. First, it is unclear how the corpus of 1800 peace treaties was assembled and on the basis of which selection criteria. Second, there is no information about the editorial guidelines underpinning the EFVO data. Which (meta)data about these peace treaties were included and which were not? Third, information about the ways in which the data is modelled and the reasons behind this model is absent as well. Fourth, which research questions guided this project and how did these influence

the selection of the corpus as well as the data model? Phrased differently, the (conceptual) links between the historical and digital components of the EFVO project are shrouded in darkness. In the end, because of the lack of robust documentation, we needed to induce EFVO's intellectual framework based on the existing database and the data contained in it. As mentioned before, ideally the existence of detailed documentation allows one to understand how and why a dataset has been created, which makes it much easier for scholars to judge whether and how the data can be used for their own specific purposes. Seen from this perspective, solid documentation has the potential to enhance a project's scholarly impact.

Lesson 2: Users and use cases

Fundamental questions underpinning most DH projects are which data to capture and how to go about this. It rarely happens that scholars with various disciplinary backgrounds have an equal interest in the entirety of a project's data. For example, historical linguists might be interested in the transcribed text in the editions and transliterations that are part of the EFVO/FriVer+ corpus, whereas for historians the treaties' metadata as well as the places and persons mentioned (and tagged) in them might be of greater importance. Although it is impossible to envisage all possible use cases and tailor the data to the specific needs of each and every potential user, in order to create a data set that is useful to a large and diverse audience, one needs to think about the creation of a data model that allows the data to be queried and approached in different ways. Rather than creating a resource that is designed to answer very specific research questions only, one should try to create something that is more versatile and addresses several scholarly desiderata or needs. It goes without saying that one's own research questions will be prioritized, but at the same time one should make the effort to accommodate the needs of other scholars (if time and money allows, of course).

In the case of the FriVer+ project, this was less relevant as we mainly transformed data that had already been generated by EFVO, yet at least the creation of structured data in accordance with the TEI schema and the FAIR principles means that more scholars can and hopefully will use it. Moreover, whereas previously the data could only be accessed through querying a SQL-database, now everyone has free and unrestricted access to our (raw) data. For advanced users in particular, this access is essential. In addition, therefore, to taking the different disciplinary backgrounds of potential users of the data into account, it is helpful to make a distinction between the competences of the end users and the different needs or desires this might generate. Are we creating a resource that is predominantly for humanities scholars, digital humanists, or

the wider public? Asking such questions is helpful when designing a project and determining its aims, outcome, and scope.

Lesson 3: Be pragmatic

The creation of data according to the TEI schema and the FAIR-principles was one of the main aims of the FriVer+ project. Indeed, the use of the former is an important component of achieving the latter, given that the widespread use of the TEI schema fosters the easy exchange of data. Still, in particular with (even smaller) legacy projects, it is important to be pragmatic. TEI can do a lot, but not everything. Moreover, the customization of the TEI schema takes time and sometimes trying to shoehorn (historical) data in this schema is not necessarily the most elegant solution. Hence TEI should not be uncritically adopted for every project. Instead, based on the available time and money, a careful decision should be made whether the use of TEI is worth it in relation to the other envisaged aims and goals. Having said that, there are expectations that DH projects use TEI, so when deciding not to do so, be ready to defend this decision in public and private.

The decision to convert the data by making use of an online transformer instead of (proprietary) software was also guided by a healthy dose of pragmatism. To be sure, more elegant solutions were available and it would have been nice to acquire some new skills, yet within the context of a parttime and one-year project, finding something that works trumps elegance. Having said that, the size of the dataset allowed us to proceed in the way we did in FriVer+. With a larger dataset, the hours of manual labour required would have been too much and more sophisticated methods would have been necessary.

Lesson 4: Find allied projects

Although small legacy projects such as EFVO continue to be useful for scholars, because of its size it is easier overlooked than larger projects on similar topics. This can be remedied partly by linking one's own data to other datasets, like we have done in the FriVer+ project. In this way, a project becomes part of an ecosystem of datasets, thereby enhancing its discoverability. Another option is to cement alliances with other related projects, like FriVer+ did with EuRed. This has the benefit that projects can reinforce each other and benefit mutually through the exchange of ideas, knowledge, and skills. In the case of FriVer+, its data now is hosted by one of EuReD's partner institutions, TU Darmstadt. Again, this fosters the visibility and discoverability of FriVer+, while EuReD can make use of the FriVer+ data, a win-win situation for both parties involved. Moreover, given the limited IT infrastructure at the IEG, seeking an

institutional partner with such facilities is even more desirable. Naturally, all of this hinges on the existence of allied projects as well as the willingness of other institutions to cooperate and share resources.

Appendix A: Structure of the EFVO database

