Journal of the Text Encoding Initiative
Rolling Issue | 2022

Modeling a Born-Digital Factoid Prosopography using the TEI and Linked Data

Daniel L. Schwartz, Nathan P. Gibson and Katayoun Torabi

Electronic version
URL: https://journals.openedition.org/jtei/3979
DOI: 10.4000/jtei.3979
ISSN: 2162-5603

Publisher
TEI Consortium

Electronic reference
Daniel L. Schwartz, Nathan P. Gibson and Katayoun Torabi, "Modeling a Born-Digital Factoid Prosopography using the TEI and Linked Data", *Journal of the Text Encoding Initiative* [Online], Rolling Issue, Online since 21 March 2022, connection on 22 March 2022. URL: http://journals.openedition.org/jtei/3979 ; DOI: https://doi.org/10.4000/jtei.3979

Modeling a Born-Digital Factoid Prosopography using the TEI and Linked Data

Daniel L. Schwartz, Nathan P. Gibson, and Katayoun Torabi

ABSTRACT

Although the TEI has traditionally been used for encoding text, its combination of structured and semi-structured data has made it a compelling choice for born-digital, linked-data resources as well. Our intent here is to demonstrate the advantages it offers for digital prosopographies along with a model that can be used for them. Syriac Persons, Events, and Relations (SPEAR) is a born-digital prosopography project in the field of Syriac studies. Where traditional prosopographies focused on prose descriptions of individual persons of significance, SPEAR follows recent developments in research methodologies that instead produce prosopographical factoids. Factoids are structured data about persons drawn from the analysis of historical texts. Most factoid prosopographies use relational databases to model data. Instead, SPEAR uses a customized TEI schema to model factoids that can be queried and visualized in an XML database as well as

serialized in HTML for human viewers and in RDF for data sharing. The TEI's provisions for structured and semi-structured data make it ideal for encoding data from heterogeneous historical source material. Moreover, its linking capabilities connect SPEAR data to related data sets. By modeling prosopographical factoids, and not the source texts themselves, SPEAR offers an example of how a born-digital, data-oriented approach to using the TEI can circumvent some of the challenges posed by the tree structure of XML. It also disrupts traditional understandings of *data* and *stand-off markup* through combining linked open data approaches with the use of the TEI.

INDEX

Keywords: factoids, Linked Open Data, prosopography, stand-off markup, Syriac studies

ACKNOWLEDGEMENTS

Nathan P. Gibson's contributions to this article have been funded by the German Federal Ministry of Education and Research in the framework of the Communities of Knowledge project.

1. Introduction: SPEAR and the Syriaca.org Data Graph

1 The TEI offers multiple ways for humanists to render and engage with text as data. Although the TEI was primarily designed to encode text, it also allows the encoding of data that lie at a greater distance from the text itself. Many text encoding projects use these portions of the TEI to encode stand-off markup containing data about aspects of their encoded texts: entities like persons or places, linguistic material, or taxonomies.

2 Syriaca.org is a linked open data (LOD) project that places this use of the TEI at the center of its research agenda. It uses the TEI to encode authority files on persons, places, primary source texts, manuscripts, scholarly bibliography, and a taxonomy of concept keywords, all of which are related to the cultural heritage of the Syriac communities. These communities share the use of the Syriac language, a form of Aramaic which flourished in Greater Syria and Mesopotamia throughout the late antique and medieval periods and continues to be a part of religious and cultural life in parts of the Middle East and in diasporas in India, the U.S., and Europe. As Syriac has coexisted variously with Greek, Coptic, Arabic, Latin, and other languages, the Syriaca.org

project often links its data to digital projects in these neighboring fields.[1] For each Syriac-related entity, Syriaca.org creates a separate XML file, mints a Uniform Resource Identifier (URI), collects information about that entity, links to related URIs for other Syriaca.org entities, and links out to external resources such as VIAF, Wikidata, Pleiades, and others. Moreover, these URIs are so-called cool-URIs that are not only unique strings but are also resolvable by a web browser into HTML pages with a human-readable version of the data about each entity (Sauerman and Cyganiak 2008). The extensive pointers to URIs for other entities in each file result in a graph of data. Syriaca.org Senior Programmer Winona Salesky has also developed the open-source Srophé Application. Built on eXist-DB, it renders the TEI data as HTML, facilitates advanced browse and query features, and also renders the TEI data into various forms of the Resource Description Framework (RDF).[2] On the whole, Syriaca.org can be understood as a reference work that leverages the linked open data environment.

3 Syriac Persons, Events, and Relations (SPEAR) is a related project that relies on the Syriaca.org authority files and the data graph that results from their linked encoding, but focuses on analyzing primary sources by deriving assertions from them rather than providing reference information. SPEAR encoders create a separate born-digital TEI file for each primary source text and employ a customized TEI schema to encode prosopographical data extracted from texts. Thus the result is not a model of the source texts themselves but actually a model of prosopographical factoid data derived from those texts.

2. Of Facts and Factoids

4 The modern historical methodology known as prosopography began in the late nineteenth century, flourished throughout the twentieth century, and continues today. Definitions of this research methodology are varied, some emphasizing the descriptions of individual persons, others emphasizing connections between the individuals in a prescribed group. K. S. B. Keats-Rohan discusses this history of prosopographical research and offers the following useful definition, “Prosopography is about what the analysis of the sum of data about many individuals can tell us about the different types of connection between them, and hence about how they operated within and upon the institutions – social, political, legal, economic, intellectual – of their time” (2000, 2). Initially, this work focused on the production of magisterial print volumes containing brief

descriptions of known persons with a shared characteristic (i.e., occupation) or socio-economic status. The chronological focus of the earliest work was ancient and medieval Mediterranean history and these works have proven invaluable for historians seeking to understand social networks and the political trajectories of elite Greeks and Romans (Cameron 2003). While historians have long used structured databases for historical sources such as census data whose original form was already a table and thus anticipated database technology, it was only very recently that scholars began to work out how to structure prosopographical data to leverage the potential of a structured database. Over the last couple of decades, a group of scholars at King's College London, originally part of the Centre for Computing in the Humanities but now refashioned into the Department of Digital Humanities (CCH/DDH), pioneered a particular approach to prosopographical research designed to model prosopography in a relational database. The result has been projects such as the Prosopography of the Byzantine World (PBW), the Prosopography of Anglo-Saxon England (PASE), and the Clergy of England Database (CCEd).[3] The shift from narrative descriptions to databases required a conceptual move from prosopography as textual descriptions of persons to structured data about persons. This move was anticipated by print prosopographies such as the Prosopographia Imperii Romani and Prosopographia Ptolemaica (both subsequently digitized) which made use of identifying numbers, cross-referencing, and source quotations in individual entries. Prosopographical factoids develop this methodology by focusing on small pieces of data about persons: name variants, occupations, places of residence, events in which they played a role, personal and professional relationships, and possessions. Factoid prosopographers then take these discrete pieces of information and put them into a relational database that creates a graph of relationships between the pieces.

5 Unfortunately, the term *factoid* can cause confusion. Where facts are generally understood to be confirmed truths, the term *factoid* was introduced by Norman Mailer in his biography of Marilyn Monroe to describe pieces of information, true or otherwise, that have a life only on account of being repeated by the media (Marsh 2014). This idea is not what factoid prosopographers have in mind, however. In adopting the term to describe their approach to prosopography, researchers at CCH/DDH instead courted the irony of the term (Pasin and Bradley 2015, 89). Rather than asserting that the result of factoid prosopography is a database of fabricated or insignificant information, the researchers who applied the term intended to emphasize the contingency of

information asserted in textual sources. Factoids capture assertions made in primary source texts and not necessarily confirmed truths about the past. The factoid approach to prosopography thus captures the scholar's interpretation of what a historical source asserts about persons. As a result, factoids may or may not be verifiable and in some cases are demonstrably false. Section 4.6 discusses how SPEAR handles these kinds of uncertainties. The scholarly interpretation required to understand and reconstruct the source's assertions is a first step toward historical criticism, but factoid prosopographies stand in contrast to some prosopographical databases that are intended to contain verified historical facts. It might be helpful to think of a factoid prosopography as a sophisticated index of the source's assertions, as understood and annotated by the encoder.[4] The resulting born-digital work provides a space situated between the primary source text and the historical-critical evaluation of that text where researchers can encode and make publicly available in a linked data framework their interpretation of the text. As we will show below, the TEI offers an excellent standard for the kind of nuanced interpretive work required to meet this purpose.

6 The projects using this approach to prosopography have so far focused on relational databases for modeling prosopographical factoids. While highlighting this fact back in 2005, John Bradley and Harold Short also noted the possibility of using XML for factoid-based prosopography, especially given the potential of eXist-DB and XQuery to index, search, and browse XML data, thus allowing suitably encoded XML documents to produce the kind of data graph generally associated with the tables of a relational database (Bradley and Short 2005, 15; Bradley 2005; Bradley 2014, 18–19). SPEAR uses these technologies in many of the ways anticipated by Bradley and Short but ultimately sidesteps some of the difficulties they discuss with using XML for factoid prosopography, particularly the limitations of the XML tree structure. This article describes a customization of the TEI that models born-digital prosopographical factoids and integrates them with an LOD infrastructure.

3. Factoids, Texts, and the TEI

7 Syriaca.org adopted the TEI for expressing its data model prior to the inception of SPEAR. Eventually, project members decided that in addition to the Syriac Biographical Dictionary (SBD), with each individual XML file containing information on one individual, we wanted to pursue a factoid approach to doing prosopography as well. Several things made this approach desirable. SBD

offers an authority file for persons relevant to Syriac studies, including a URI minted on the basis of at least one or two references to the person in either primary or secondary source material. These URIs can then be used for authority control (such as in library catalogues) and as identifiers in digital projects.[5] In some instances, a record on an individual contains many data points while other records are quite sparse, serving the purpose of identifying a person and offering relevant citations but not necessarily providing extensive data beyond a basis for disambiguation. For SBD, encoders are expected to bring their knowledge of Syriac studies and scholarly judgment to bear on the accuracy of the information contained in SBD entries. SPEAR, on the other hand, has different aims and demands a different kind of interpretive work from encoders. It does not produce data to be used for authority control, nor does it offer its own canonical identifiers for persons. Rather, SPEAR uses the system of person URIs established in the SBD authority files and employs them in the encoding of prosopographical factoids—factoids which, as noted above, may in fact contain false assertions. The difference between the two data sets is most clear at the level of workflow. SBD begins from a reference to a distinct person relevant to Syriac studies, usually in a reference work or other secondary literature, and mints a URI identifying that person, at the same time recording information helpful for disambiguation, such as biographical dates and name variants. SPEAR works through individual primary source texts and employs the URIs from the SBD authority files to label persons with unique identifiers while encoding prosopographical data from those texts. To date, these primary texts have consisted of late-antique chronicles, letters, and hagiographies, but they could theoretically include any Syriac source. Not only the process but also the structure of the TEI-encoded XML files differs between SPEAR and SBD: each XML file in SPEAR contains prosopographical data about many persons from an individual text rather than information about a single person. Table 1 summarizes the differences between these two data sets.

Table 1. Comparison of SBD and SPEAR data types.

	SBD	**SPEAR**
Goal	Authority control	Factoid prosopography
Starting Point	Lists of persons	Close reading of primary source text

Usage	Identifying persons	Creating structured data out of primary-source assertions about persons
Granularity	Identifying information with primary or secondary bibliography	Any primary-source assertion about persons
Data Format	One XML file per person	One XML file per text or text part

8 Even though SPEAR uses TEI XML and focuses on texts, it does not use traditional text encoding to produce its data. The reasons for this are quite straightforward. It is extremely difficult to produce highly structured prosopographical data through the encoding of texts. Text-encoding projects seeking these kinds of data about the persons in their texts usually encode such data in some sort of stand-off markup.[6] The work of John Bradley and Harold Short on factoid prosopography and structured data has emphasized the benefits of relational database technology over XML for quite some time. They argue that a text-encoding approach to prosopography will run up against the limitations inherent in the hierarchical tree structure of XML (Bradley and Short 2005, 13–15; see also Jannidis and Flanders 2018, 73). They also note the frequent disconnect between the words and phrases employed in a text and the meaning that a scholar can derive by reading the text against the grain (Bradley and Short 2005, 13–15). In its nascent stages, SPEAR briefly attempted to achieve the kind of factoid-based prosopographical data we sought by encoding primary source texts in the TEI. We quickly experienced the types of problems discussed by Bradley and Short. It proved extremely difficult to get the granularity of data desired given the constraints of the hierarchy imposed by XML.

9 Though discussions of problems regarding the tree-structure of XML are commonplace, it is worth giving a bit more attention to the discussion of Bradley and Short. One can read their article as a defense of a relational database (RDB) model for prosopographical factoids. Given the textual nature of the source base and strong preferences among textual scholars for encoding editions of texts in TEI XML, their article aims to make a case for moving from text to database as the preferred

approach for modeling factoids. They summarize their motivations for using relational databases for the factoid projects arising from CCH/DDH as follows: "It is exactly because prosopographical projects are involved in the creation of a model of their material that is perhaps not explicitly provided in the texts they work with that a purely textual approach is in the end not sufficient in and of itself. Instead, it is exactly this kind of structuring which makes our projects particularly suitable for the relational database model" (Bradley and Short 2005, 14). They argue that the structure of text does not necessarily lend itself to the kind of data they seek to produce but also that the encoding of texts does not offer a robust enough model for their data. The results of the CCH/DDH prosopographies clearly demonstrate that rich prosopographical data can be modeled in a relational database. However, their argument in favor of RDBs over XML only considers one application of XML to the work of collecting factoids. Although the use of XML to encode texts is widespread, that is not the only way to apply XML in general and TEI in particular to the encoding of prosopographical factoids. The TEI can also be used to model the prosopographical factoids themselves.

10 SPEAR combines the TEI encoding of factoids with the LOD infrastructure of Syriaca.org and other partner projects to maintain and make explicit the close connection between prosopography and the historical source providing the material for the prosopography. This begins in the `<teiHeader>` element. The file description for each document contains a title statement with a `<title>` element indicating the title of the TEI document. The `<title>` element with a `@type` attribute of `"main"` reflects the historical work from which the data have been derived. A subtitle indicates that it is a prosopography and not an encoding of the document itself.

Example 1. Title statement.

```
<titleStmt>
 <title level="m" type="main">Chronicle of Edessa</title>
 <title type="sub">A <title>SPEAR</title> Prosopography</title>
...
</titleStmt>
```

Likewise, the `<publicationStmt>` contains an `<idno>` element defining the URI for the document. This URI identifies the document as part of SPEAR, but also employs the numerical portion of the corresponding Syriaca.org work URI (which takes the form `http://syriaca.org/work/{\d+}/tei`).

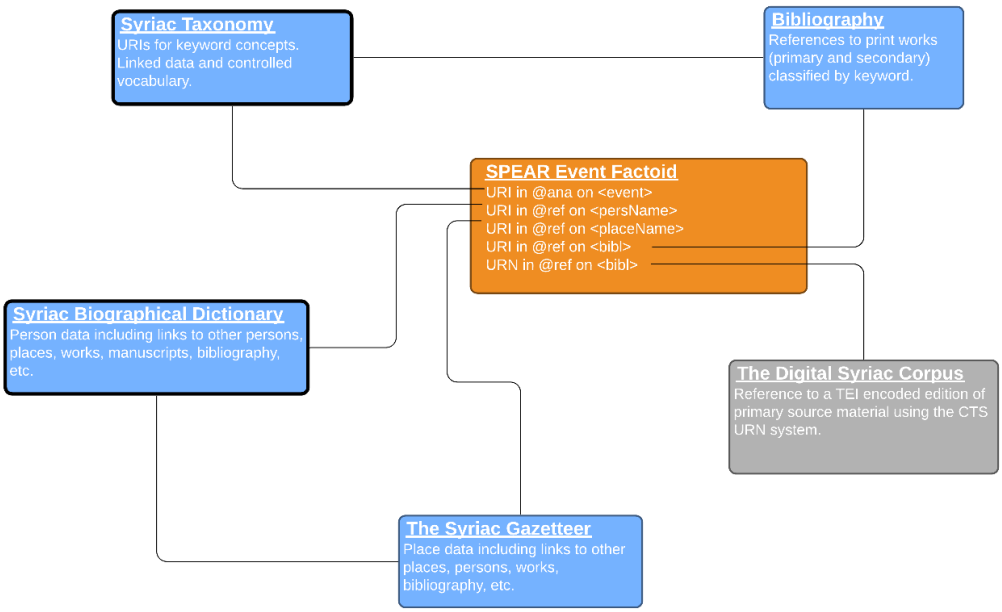
Example 2. Publication statement.

```
<publicationStmt>
 ...
 <idno type="URI">http://syriaca.org/spear/8559/tei</idno>
...
</publicationStmt>
```

This same number is also used as the filename for the XML document.

11 The intent behind this practice is to maintain a connection between the historical source from which researchers encode factoids and the document in which they encode the individual factoids. From the perspective of the data themselves, this practice is not particularly significant. Each factoid exists on a specific path and the document is nothing more than one point on that path. However, from the perspective of workflow, textual interpretation, and data entry, the document level collection does hold significance. The encoding workflow reinforces the textually oriented nature of the data as encoders work document by document to encode the prosopographical material contained in that document. The textual orientation of the TEI nicely facilitates this workflow even though SPEAR does not produce encoded historical texts. This practice is not only followed at the level of the `<teiHeader>`, textual connections are important at the factoid level also, as will be demonstrated below. Figure 1 shows some examples of the kinds of connections between factoids and other related data sets. These connections will be explored in the discussion of the factoid modeling that follows.

Figure 1. Connections between SPEAR and other LOD resources.



4. Prosopographical Factoids in the TEI

12 Encoding a SPEAR factoid entails using the XML structure, the semantic value of the TEI elements and attributes, and the linking abilities of Syriaca.org URIs to express the encoder's interpretation of a prosopographical datum derived from a historical text in the Syriac literary tradition. While SPEAR factoids mostly validate against the TEI P5 schema, the customization significantly constrains the document in order to ensure SPEAR's required formatting for factoid data (see ODD and RelaxNG). It also makes extensive use of Schematron rules to validate attribute values against the appropriate Syriaca.org URIs for different entities. These constraints offer a blueprint for modeling prosopographical data in ways very similar to those described by Bradley and Short when expressing their preference for the RDB model.

13 The "Names, Dates, People, and Places" module of the TEI Guidelines contains a wealth of elements for encoding data about these entities. Moreover, the motivation behind this module is the markup of precisely the kind of prosopographical data contained in SPEAR: "The main intended applications for this module are in biographical, historical, or geographical data systems such as gazetteers and biographical databases, where these are to be integrated with encoded texts" (TEI

Consortium 2020, 13). Nevertheless, modeling a factoid prosopography using the TEI required a lengthy iterative process of moving back and forth between Syriac sources in various genres and the TEI Guidelines. No prosopography can capture every piece of data from every source, nor can it anticipate everything users will want to know about the persons in the resulting data set. These challenges would exist even if the TEI Guidelines anticipated the encoding of prosopographical factoids.

4.1 Containing and Identifying Factoids

14 Modeling prosopographical factoids required the identification of an element to serve as a container for each factoid. We chose the `<ab>` element since it offers a generic container for textual material. Each factoid appears in its own `<ab>` element and must also include a `@type` attribute with a value of `"factoid"` indicating how the `<ab>` element is being used. As described above, Syriaca.org's linked data model requires that entities receive a cool URI, a unique identifier that is resolvable to a web page containing a human-readable version of the data. The `<ab>` contains the textual expression of the conceptual entity that is the factoid. SPEAR uses an `@xml:id` on each `<ab>` element in order to identify this expression of the factoid. However, we also declare the full URI for the factoid by requiring an `<idno>` element as a child of `<ab>`. Schematron rules in SPEAR's customized schema require that the portion of the URI immediately preceding the hyphen match the numerical portion of the `<idno>` in the `<publicationStmt>` which in turn must match the file name for the current document.

Example 3. Factoid container and identifier.

```
<ab type="factoid" subtype="event" xml:id="factoid-5" resp="#jwalters">
 <idno type="URI">http://syriaca.org/spear/798-5</idno>
...
</ab>
```

15 The `<ab>` elements in SPEAR must also include a `@resp` attribute. While the header contains `<editor>` and `<respStmt>` elements detailing contributions at the document level, we feel it is important that the intellectual contribution constituted by the interpretation of the text and encoding of the prosopographical data out of that text requires attribution at the factoid level. This information is useful for the encoder in terms of being able to see the number of factoid

contributions to the project as well as to the user of the data who might find that information significant for determining the credibility of the way the assertion is interpreted in the factoid. In the process of encoding factoids, the encoders contributing to a given factoid select their personal identifiers from those allowed by SPEAR's customized schema.

16 Each factoid in SPEAR shares the structure described so far; a containing `<ab>` element, `@type`, `@xml:id`, and `@resp` attributes, along with a child `<idno>` element indicating the full URI for the factoid. The highest level of divergence has to do with the three basic types of data that SPEAR collects: data on persons, data on events of significance to persons, and data on relationships between persons. A `@subtype` attribute on the factoid `<ab>` offers the first indication of this divergence. In the previous example, the `@subtype` indicates that this is an event factoid. A relation factoid is indicated by the value `"relation"` while the person factoid values specify the type of person data encoded. Those values, mostly derived from TEI element names, include: `"birthDate"`, `"birthPlace"`, `"citizenship"`, `"deathDate"`, `"deathPlace"`, `"education"`, `"ethnicLabel"`, `"gender"`, `"langKnown"`, `"mentalState"`, `"nameVariant"`, `"occupation"`, `"physicalTrait"`, `"residence"`, `"sanctity"`, and `"socecStatus"`. The `@subtype` values are not, strictly speaking, necessary as they can be derived from the context of the XML. However, we include these values because it facilitates easier processing of SPEAR's different data types.

17 As the TEI does not allow `<person>`, `<event>`, or `<relation>` to appear as direct children of `<ab>`, SPEAR uses `<listPerson>`, `<listEvent>`, and `<listRelation>` followed by the respective elements. A factoid is by definition an assertion of personal information sourced to a primary source text and every `<ab>` must also have one or more `<bibl>` elements. While not technically necessary for a factoid prosopography, SPEAR also includes the opportunity for encoders to add a `<note>` element as a child of `<ab>`. This element may hold a discursive note for an encoder to express additional information about the factoid assertion contained elsewhere. These notes may be used to express uncertainty and doubts about the veracity of the assertions made in the primary source text. We will return to a fuller discussion of the `<bibl>` and `<note>` elements after discussing how SPEAR uses the `<relation>`, `<event>`, and `<person>` elements. While all of the data encoded using these elements are prosopographical, we speak in terms of person factoids, event factoids,

and relation factoids. This division is based on elements in the TEI used to encode these different types of data. In the following sections we will discuss our methodologies for encoding each of the three types of factoids, moving from the simplest to the most complex.

4.2 Relation Factoids

18 Relation factoids encode relationships between people using the `<relation>` element and attribute values that leverage the same linked data standards seen elsewhere in factoids. Relation factoids can either stand alone or as discussed in section 4.3.2, they can be included in an event factoid using the `<relation>` element as a child of the `<desc>` element. Following the lead of the TEI Guidelines (TEI Consortium 2020, 13.3.2.3), the `<relation>` element may take either a `@mutual` attribute or both an `@active` and a `@passive` attribute, depending on the type of relationship. The relationship *spouse of* would take a `@mutual` attribute with two values while the relationship *parent of* would take the parent in the `@active` attribute and the child in the `@passive` attribute. In all cases, acceptable values for the attributes are Syriaca.org person URIs established in the SBD authority files, as discussed in section 3 and section 4.4. The `@type` attribute on this `<relation>` element takes `"person"` indicating that this encodes a personal relationship. While this is clear enough from the context of the XML, including this attribute value facilitates processing of the factoid data by distinguishing this from the encoding of a relationship between two events as discussed in section 4.6. SPEAR also encourages encoders to provide date information for the element in a factoid using attributes from the `att.datable` class provided by the TEI Guidelines (TEI Consortium 2020, 3.5.4 and 13.3.7).

19 A `@ref` attribute on `<relation>` provides a pointer to a keyword URI in The Syriac Taxonomy, where various identifiers and metadata about the relationship are encoded.

Example 4. SPEAR `<relation>` element.

```
  <ab type="factoid" subtype="relation" xml:id="factoid-15" resp="#ktorabi
#dschwartz">
   ...
   <relation type="person" ref="http://syriaca.org/keyword/sibling-of-parent-of"
active="http://syriaca.org/person/3056" passive="http://syriaca.org/person/2553">
    <desc>
     <persName ref="http://syriaca.org/person/3056">Anonymous 3056</persName> was
the sibling of the parent of
```

```
    <persName ref="http://syriaca.org/person/2553">Antoninus</persName>. </desc>
  </relation>
 ...
</ab>
```

The Taxonomy contains fifty-four relationships divided into ten categories: family relationships, personal relationships, professional relationships, clerical relationships, monastic relationships, intellectual relationships dealing with schooling and influence, relationships constituted through legal proceedings, relationships constituted through military service, relationships constituted through slavery, and epistolary relationships dealing with references to third parties and transport of letters. Additionally, SPEAR allows an optional `@ana` attribute that again points to keyword URIs in the Taxonomy. These keywords qualify relationships in the `@ref` attribute, further specifying family relationships as maternal, paternal, in-law, step, adoption, half, foster, claimed, or ritually based (i.e., baptismal godparent). Taxonomy relationships derive from the relationship ontology produced in the SNAP:DRGN project.[7] The SNAP ontology was designed to create a data standard for networking ancient prosopographical material. Not only is SNAP a project with which SPEAR has a strong interest in sharing data, but SNAP's focus on the ancient world means that many of the relationships expressed in SPEAR's source base are already present in the SNAP ontology. For example, modern relationship ontologies such as Friend of a Friend (FOAF) do not provide robust mechanisms for describing the relationships constituted by the practice of slavery. SPEAR uses most of the relationship types included in the SNAP ontology.[8] Given the nature of the source base and the expectations of researchers in Syriac studies, SPEAR has also created a number of relationships that can be expressed as subsets of SNAP relationships. In particular, SPEAR captures a range of relationships from within the ranks of Christian clerics and monastic communities. Example 5 shows a type of relationship not included in the SNAP ontology.

Example 5. Ordination.

```
  <relation type="person" ref="http://syriaca.org/keyword/ordained"
active="http://syriaca.org/person/2562" passive="http://syriaca.org/person/51">
   <desc>
    <persName ref="http://syriaca.org/person/2562">Epiphanius</persName>
    ordained <persName ref="http://syriaca.org/person/51">Severus of Antioch</
persName>. </desc>
```

```
  </relation>
```

When SPEAR prepares RDF to share with the SNAP project, this clerical relationship can be expressed as a SNAP professional relationship. This practice allows for consistent encoding within SPEAR that can also be serialized as RDF according to the needs of other linked open data projects.

4.3 Event Factoids

20 Event factoids contain an encoding of any event deemed relevant by the encoder of the file. The <event> element takes an @ana attribute with one or more values pointing to the Taxonomy. This keyword tagging creates links between this event and other SPEAR events but also creates links between the event and scholarly bibliography described with the same keywords. The <event> element also contains a <desc> element offering the encoder's interpretation of the event described in the primary source text. The SPEAR encoding manual encourages encoders to describe events making liberal use of <persName> and <placeName> elements with @ref attributes pointing to Syriaca.org person and place URIs. These URIs link different factoids but also link factoids to other Syriaca.org data sets.

Example 6. An event factoid.

```
  <ab type="factoid" subtype="event" xml:id="factoid-555" resp="#dschwartz">
   <idno type="URI">http://syriaca.org/spear/8559-555</idno>
   <listEvent>
    <event ana="http://syriaca.org/keyword/emperors http://syriaca.org/keyword/
bishops http://syriaca.org/keyword/natural-disasters">
     <desc> The fourth time the walls of <placeName ref="http://syriaca.org/
place/78"> Edessa </placeName> were broken down was in the days of the king
      <persName ref="http://syriaca.org/person/2268"> Justin </persName> and of
the bishop <persName ref="http://syriaca.org/person/2236"> Asclepius </persName>
in
      <choice>
       <reg><date notBefore="0524-10" notAfter="0525-09"
calendar="#gregorian">524/5</date></reg>
       <orig><date when-custom="0836" datingMethod="#seleucid">A. Gr. 836</date></
orig>
      </choice>
     </desc>
```

```
    </event>
   </listEvent>
  ...
 </ab>
```

21 Encoders have a great deal of discretion regarding the encoding of events. It is up to the scholarly expertise of the encoders as they engage with the text and draw upon field-specific knowledge to determine what constitutes an event. Different encoders can disagree on the scope of the events they encode and exactly how many factoids they assign to a closely related series of events. Encoding instructions encourage the liberal use of keyword identifiers using the `@ana` attribute but the application of keywords remains a part of the interpretive work of the encoder. SPEAR seeks a rather general application for this data model, but other projects might forgo a more general approach to encoding events and focus instead on encoding only events that serve to address more specific research questions.

22 It is important also to recall that the factoid approach offers descriptions of textual assertions on events. The factoid describing a historical event does not function as an authority file of that historical event. Entities like persons, places, and works can be difficult enough to isolate for the purpose of assigning a URI and establishing an authority file.[9] These issues become significantly more complicated when it comes to historical events since what constitutes an event can vary so dramatically depending upon the source, the perspective of the encoder, and the research questions motivating the encoding.

4.3.1 Dates in Factoids

23 Example 6 demonstrates SPEAR's preference for applying dates whenever indicated in the source text. In so far as a factoid is an interpretation of a text, the encoder may simply use the `<date>` element with a properly formatted `@when`, `@notBefore`, or `@notAfter` attribute by itself. However, the SPEAR encoding manual expresses a preference for using the TEI's ability to capture the calendar system used in the text. Example 6 shows how encoders use `<choice>` with a child `<reg>` element indicating `<date calendar="#gregorian">` and a child `<orig>` element with a `<date>` element pointing to another method of dating. Likewise, encoders can also capture non-calendrical dating systems as shown in example 7.

Example 7. Non-calendrical dating systems.

```
  <choice>
   <reg><date notBefore="0379-01-19" notAfter="0380-01-18"
calendar="#gregorian">379/80</date>
    <note ana="#calculated">This regularized date was calculated by the SPEAR
editor from a non-calendrical dating system such as regnal years, etc.</note></
reg>
   <orig><date datingMethod="#regnalYear">the year Theodosius the Great became
king</date></orig>
  </choice>
```

A <note> as child of <reg> allows the HTML serialization of the data to indicate to the user that this date has been calculated by the encoder.

4.3.2 Relationships Established through Events

24 An additional complication presents itself when it comes to dealing with the intersection of events and relationships in the SPEAR factoid model. How do we encode events that create relationships? Take, for example, a birth. A birth is an event, but it also creates a parent-child relationship. While a SPEAR encoder can encode these data using two separate factoids, the SPEAR encoding manual prefers inserting the relationship data in a <relation> element as a descendant of the <desc> element.

Example 8. An event that creates a relationship.

```
  <ab type="factoid" subtype="event" xml:id="factoid-12" resp="#syoung
#dschwartz">
   <idno type="URI">http://syriaca.org/spear/3045-12</idno>
   <listEvent>
    <event ana="http://syriaca.org/keyword/ordination http://syriaca.org/keyword/
bishops http://syriaca.org/keyword/clergy http://syriaca.org/keyword/letters">
     <desc>
      <date notBefore="0513" notAfter="0518" calendar="#gregorian">Sometime
between 513 and 518</date>,
      <persName ref="http://syriaca.org/person/51">Severus</persName> wrote a
letter to <persName ref="http://syriaca.org/person/2546">Entrechius</persName>
```

```
      about the appointment of <persName ref="http://syriaca.org/
person/3064">Anonymous 3064</persName> in
      <placeName ref="http://syriaca.org/place/173">Rhosus</placeName>.
      <listRelation>
       <relation type="person" ref="http://syriaca.org/keyword/sender-of-
letter-to" active="http://syriaca.org/person/51" passive="http://syriaca.org/
person/2546"/>
      </listRelation>
     </desc>
    </event>
   </listEvent>
  ...
  </ab>
```

This practice maintains the tight connection between event and relationship and has the added benefit of allowing the inclusion of these kinds of events into graphs of the relationships depicted in bodies of literature.

4.4 Person Factoids

25 SPEAR encodes the following data within the `<person>` element: birthdate, birthplace, citizenship, death date, death place, education, ethnic label, gender, language, mental state, name variant, occupation, physical trait, place of residence, sanctity, and social rank.[10] Every `<person>` element must indicate the person about whom the item of data is asserted by the primary source text. SPEAR accomplishes this association with a self-closing `<persName>` element as the immediate child of `<person>`, along with a `@ref` attribute pointing to the Syriaca.org URI for that person. There are several important points to be made regarding this connection. To begin with SPEAR does not contain the Syriaca.org authority file for the individual persons. Neither is it the case that some kind of aggregation of SPEAR data points about an individual constitutes such an authority file. SPEAR makes factoid statements indicating what a source asserts about the person entities identified by Syriaca.org authority files. As a consequence, Syriaca.org must have an authority file among its data for any entity an encoder would like to say something about in a SPEAR factoid.[11] SPEAR references Syriaca.org URIs for persons, places, keywords, bibliography, and works.

26 Following the <persName> element, a person factoid must have one of the following: <birth>, <death>, <education>, <nationality>, <residence>, <occupation>, <langKnowledge>, <state>, <trait>, or <socecStatus>. The <birth> and <death> elements contain either <date> with a following-sibling <note> or <placeName> with a following-sibling <note>.

Example 9. Person factoid: date of death.

```
<ab type="factoid" subtype="deathDate" xml:id="factoid-155" resp="#dschwartz">
 <idno type="URI">http://syriaca.org/spear/8559-155</idno>
 <listPerson>
  <person>
   <persName ref="http://syriaca.org/person/13"/>
   <death>
    <choice>
     <reg><date when="0373-06-09" calendar="#gregorian">9 June 373</date></reg>
     <orig><date when-custom="0684-09-09" datingMethod="#seleucid-
SyriacMonths">9 Ḥazirān A. Gr. 684</date></orig>
    </choice>
    <note type="desc">
     <persName ref="http://syriaca.org/person/13">Ephrem</persName> died on 9
June 373. </note>
   </death>
  </person>
 </listPerson>
...
</ab>
```

Example 9 contains the encoding of a date of death.

27 As with the <persName> child of <person> discussed above, the <placeName> in a birth or death element is a self-closing element with a @ref attribute pointing to the Syriaca.org URI for that place.

Example 10. Person factoid: birthplace.

```
<ab type="factoid" subtype="birthPlace" xml:id="factoid-7" resp="#ktorabi">
 <idno type="URI">http://syriaca.org/spear/796-7</idno>
 <listPerson>
  <person>
```

```
      <persName ref="http://syriaca.org/person/1885"/>
      <birth>
       <placeName ref="http://syriaca.org/place/2804"/>
       <note type="desc">
        <persName ref="http://syriaca.org/person/1885">Maro</persName> was born in
<placeName ref="http://syriaca.org/place/2804">Kalesh</placeName>.</note>
      </birth>
     </person>
    </listPerson>
   ...
  </ab>
```

Similar to <birth> and <death>, <nationality>[12] and <residence> both take <placeName> with a following-sibling <note>. Again, this <placeName> element is self-closing and must include a @ref attribute pointing to a Syriaca.org place URI. The <note type="desc"> elements in person factoids provide a human-readable version of the data for the HTML serialization of the factoid. From a purely data-oriented perspective, these <note> elements are not strictly necessary. The semantic value of the TEI elements combined with an ISO date and the URIs in @ref attributes suffice to produce usable, machine-readable data. The Srophé Application is able to ingest this combination of person, date, and place information and facilitate visualization and queries of the data. We will discuss the use of the <note> element below, after we have seen some of the broader context that makes its significance clearer.

28 The remaining types of person factoids collected by SPEAR require a slightly different structure. The <education>, <occupation>, <state>, <trait>, and <socecStatus> elements each have only one child element, a <note type="desc">. Again in this context, the <note> element offers a more prose-oriented expression of the factoid. However, unlike the situation featuring the elements discussed in the previous paragraph, some additional information is necessary to produce the kind of highly structured data required by SPEAR. Here SPEAR again makes use of The Syriac Taxonomy. The <education>, <occupation>, <state>, <trait>, and <socecStatus> elements all take an @ana attribute with a Taxonomy keyword URI as their value. Example 11 contains a factoid asserting the occupation of a person who was a Christian bishop.

Example 11. Person factoid: occupation.

```
<ab type="factoid" subtype="occupation" xml:id="factoid-191" resp="#dschwartz">
 <idno type="URI">http://syriaca.org/spear/8559-191</idno>
 <listPerson>
  <person>
   <persName ref="http://syriaca.org/person/2251"/>
   <occupation ana="http://syriaca.org/keyword/bishops">
    <note type="desc">
     <persName ref="http://syriaca.org/person/2251">Cyrus</persName> was Bishop
of <placeName ref="http://syriaca.org/place/2605"> Edessa </placeName>.</note>
   </occupation>
  </person>
 </listPerson>
...
</ab>
```

The use of the URI for bishops as the value for the `@ana` attribute on the `<occupation>` element provides a controlled vocabulary for describing the specific occupation but also serves to connect this occupation factoid to 238 scholarly books and articles in the Syriaca.org bibliography that researchers have also described using that keyword URI.

29 A similar use of the `@ana` attribute functions the same way on the `<education>` and `<socecStatus>` elements. Moreover, SPEAR uses the `<state>` and `<trait>` elements in a similar fashion for the purpose of encoding ethnic label, gender, mental state, physical trait, sanctity (used as a general category for people described as "saints," "martyrs," etc.), and social rank. SPEAR follows the TEI in using the `<state>` element for qualities more likely to change over time and `<trait>` for qualities that are more permanent or not a matter of the person's will (TEI Consortium 2020, 13.3.1). The `<state>` element requires a `@type` attribute with one of the following values: `"mental"` or `"sanctity"`. The `@type` attribute on the `<trait>` element can be either `"ethnicLabel"`, `"gender"`, or `"physical"`. As with `<education>`, `<occupation>`, and `<socecStatus>` discussed above and shown in example 11, `<state>` and `<trait>` also require an `@ana` attribute pointing to the appropriate keyword.

30 Wherever SPEAR employs keyword URIs from the Taxonomy, the customized schema constrains the `@ana` values, pointing encoders to keywords from the relevant subsets of the Taxonomy. Metadata in the encoding of the Taxonomy entities indicates which keywords are relevant to

an ethnic label, for example, and which to an occupation. Encoders are directed away from `"http://syriaca.org/keyword/bishops"` with an `@ana` attribute in a gender factoid or `"http://syriaca.org/keyword/sogdians"` with an `@ana` attribute in an occupation factoid. Example 12 shows a Schematron rule that uses an index of the Taxonomy to constrain the use of keywords for occupations. The Schematron rule uses a warning message for this purpose and provides suggested keywords for encoders. This allows encoders to use keywords not in the Taxonomy in cases where appropriate keywords do not exist. The warning message flags the need for an editor to reassess the decisions made by encoders but it also offers a mechanism for the ongoing development of the Taxonomy by collecting concepts encoders cannot find in the controlled vocabulary.

Example 12. Schematron rule in ODD constraining the use of Taxonomy keywords.

```
  <constraintSpec ident="ana-in-occupation" scheme="schematron">
   <constraint>
    <sch:rule context="//tei:occupation/@ana">
     <sch:let name="anaValues" value="tokenize(., ' ')"/>
     <sch:let name="ti" value="doc('https://raw.githubusercontent.com/srophe/
syriaca/master/documentation/indexes/taxonomyIndex.xml')"/>
     <sch:let name="occupations" value="$ti//listURI[@type = 'occupations']/uri"/>
     <sch:let name="error" value="."/>
     <sch:assert test="every $i in (tokenize(., ' ')) satisfies $i = $occupations"
role="warning">
      <sch:value-of select="$error"/> is not currently in use. SPEAR has a strong
preference for using established URIs when possible but appropriate URIs are not
always available. Please
      consult the Taxonomy. If no appropriate keyword exists in the
following list, please use your own preferred descriptive term: <sch:value-of
select="string-join($occupations, '; ')"/>.</sch:assert>
    </sch:rule>
   </constraint>
  </constraintSpec>
```

4.5 Sources and Citations

31 The definition of a factoid discussed earlier focused on the prosopographer's interpretation of what a historical source asserts about persons. As with other aspects of the project, SPEAR encoding relies heavily on the linked data infrastructure produced by Syriaca.org. Two data sets

come into play in this regard, the Syriaca.org bibliography and the New Handbook of Syriac Literature (NHSL). Every SPEAR factoid `<ab>` element requires one or more `<bibl>` elements. Each `<bibl>` element must take a `@type` attribute, showing whether the source referred to is a bibliographic item or a URN in the Canonical Text Services (CTS) system (Blackwell and Smith 2015).

32 A `<bibl type="primary">` indicates a bibliographic reference to a primary source text in the Syriaca.org bibliography. Even though all factoids have a primary source text as the source, the designation distinguishes this source from the URN approach to sourcing factoids. Because Syriaca.org has robust linked data on bibliographic items, a `<bibl type="primary">` only contains two elements, `<ptr>` and `<citedRange>`.

Example 13. Citing the Syriaca.org bibliography.

```
<ab type="factoid" subtype="deathDate" xml:id="factoid-155" resp="#dschwartz">
 <idno type="URI">http://syriaca.org/spear/8559-155</idno>
 <listPerson>
 ...
 </listPerson>
 <bibl type="primary">
  <ptr target="http://syriaca.org/bibl/633"/>
  <citedRange unit="p">5</citedRange>
 </bibl>
</ab>
```

The `<ptr>` is a self-closing element with a `@target` attribute pointing to a Syriaca.org bibliography URI. Additional information such as author, editor, date, publisher, etc. can be displayed in the HTML serialization of the factoid by drawing on that data from the Syriaca.org bibliography. The factoid does not need to repeat this information in the `<bibl>` element. The `<citedRange>` includes a `@unit` attribute and a text node indicating the appropriate place within that source.

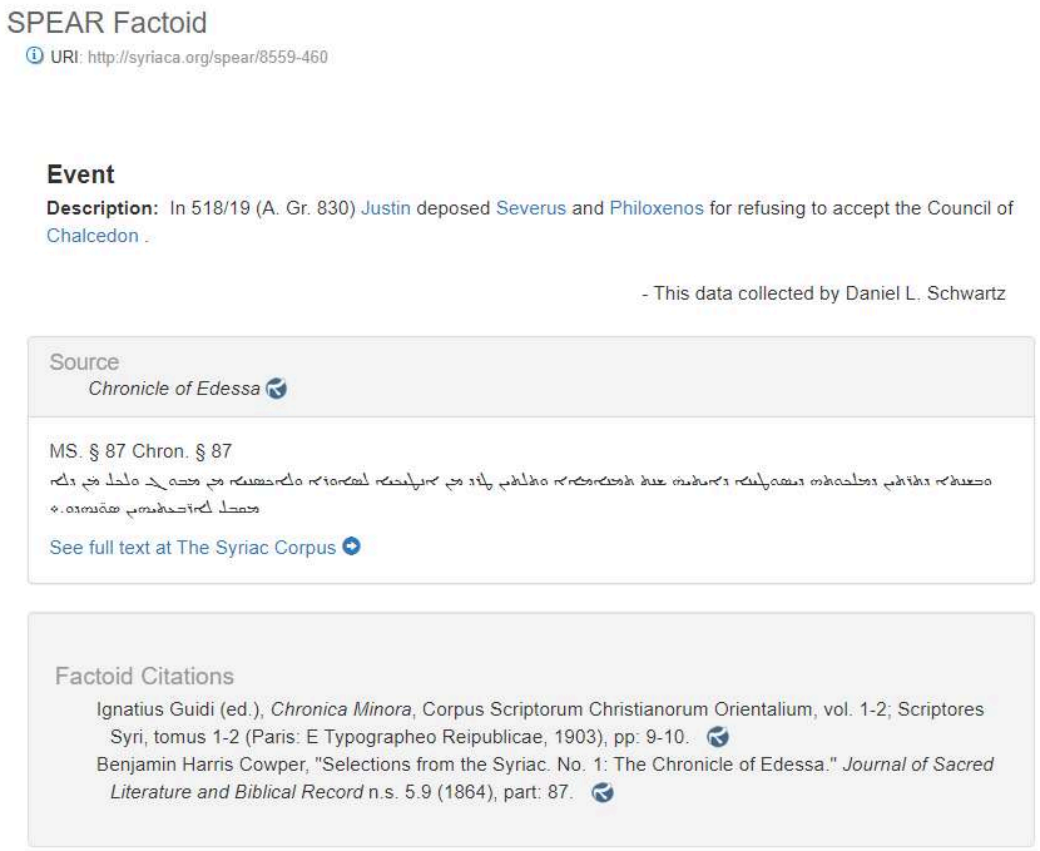
33 A `<bibl type="urn">` indicates that the citation employs the URN system established by the Canonical Text Services (CTS). [13] The URNs used by SPEAR point to the Digital Syriac Corpus, a sister project also using the Srophé Application that encodes the full text of Syriac sources and integrates them with a suite of online lexical tools.

Example 14. A factoid citing the Digital Syriac Corpus.

```
  <ab type="factoid" subtype="event" xml:id="factoid-460" resp="#dschwartz">
   <idno type="URI">http://syriaca.org/spear/8559-460</idno>
   <listEvent>
    <event ana="http://syriaca.org/keyword/council-of-chalcedon-451 http://
syriaca.org/keyword/deposition-of-clergy">
     <desc>
      In <choice>
       <reg><date notBefore="0518-10" notAfter="0519-09"
calendar="#gregorian">518/19</date></reg>
       <orig><date when-custom="0830" datingMethod="#seleucid">A. Gr. 830</date></
orig>
      </choice>
      <persName ref="http://syriaca.org/person/2268"> Justin </persName> deposed
<persName ref="http://syriaca.org/person/51"> Severus </persName> and
      <persName ref="http://syriaca.org/person/44"> Philoxenos </persName> for
refusing to accept the Council of <placeName ref="http://syriaca.org/place/622">
Chalcedon </placeName> .
     </desc>
    </event>
   </listEvent>
   ...
   <bibl type="urn">
    <ptr target="urn:cts:syriacLit:nhsl8559.syriacCorpus57:87"/>
   </bibl>
  </ab>
```

The inclusion of URNs allows the Srophé Application to make a call to the CTS API and return the piece of text cited in the factoid. This source text is then displayed on each factoid page along with a link that users can follow to see this piece of text in the context of the whole encoded work and to make use of linked lexical tools.

Figure 2. HTML serialization with Digital Syriac Corpus text displayed.



34 This use of URNs highlights SPEAR's departure from more traditional practice in using the TEI. Most often, the TEI is used to encode a primary source text. A robust project will generally produce a personography that exists as stand-off markup containing a list of identifiers, names, and personal information. SPEAR uses URNs pointing to externally encoded texts in such a way as to invert the typical practice of TEI encoding. Born digital prosopographical factoids are the SPEAR data and, from the perspective of SPEAR, the historical source encoded in the Syriac Corpus functions as stand-off markup. Of course, from the perspective of the Syriac Corpus, SPEAR and the collection of Syriaca.org authority files function as stand-off markup. Using the TEI along with robust linked data standards allows various semi-independent digital projects to produce rich sets of linked resources using a de-centralized workflow. Each of these resources can maintain their own research agenda and intended audience while also providing data that benefit the others.

4.6 Notes and Descriptions

35 Though they appear in several different contexts within SPEAR, it is best to consider the <note> and <desc> elements in one place. As mentioned earlier, the TEI Guidelines have limitations on which elements can take a child <note> or a child <desc>. From the perspective of the data encoded, SPEAR does not understand a significant difference between these two. We considered the option of customizing the schema to allow one of the two elements to appear in parallel contexts for each type of factoid. Rather than take this course, we decided to conform the data model to the TEI Guidelines but within those parameters strive to be as consistent as possible. SPEAR uses the <note> element inside person factoids because it is allowed in each of the requisite contexts: <birth>, <death>, <education>, <nationality>, <residence>, <occupation>, <langKnown>, <state>, <trait>, and <socecStatus>. The <desc> element is allowed as a child for some of these elements but in order to maintain consistency across the person factoids, SPEAR made the choice to use <note type="desc"> consistently instead of mixing <note> and <desc> elements. SPEAR uses <desc> as a child of <event> and <relation> elements in those factoid types. SPEAR also has a second use for the <note> element. When appearing as a child of <occupation>, <state>, etc. in a person factoid it offers a prose expression of that factoid content; a <note> element can also appear as the child of the factoid <ab>. Here it functions as a discursive note applied to the entire factoid. We will look at these uses in turn.

36 The use of <note> or <desc> in the first context mentioned above highlights an important aspect of the decision to use TEI for this factoid prosopography. TEI offers a well-established and widely used standard for encoding texts. As such, it allows users to take heterogeneous corpora of textual material and mark it up in such a way as to make it amenable to computation in a wide variety of ways. As mentioned above, SPEAR briefly attempted to accomplish the kind of highly structured data it sought by encoding texts directly in the TEI. This proved prohibitively difficult and, in some aspects, not possible. The options were either to abandon the TEI in favor of an RDF data set or a relational database, or to model factoids in TEI instead of modeling the texts themselves. The decision to employ the TEI and model the factoids came down to the ability of the TEI to combine structured data via the hierarchy of the tree structure and the extensive use of pointers to linked

data with prose descriptions of that data. The use of <note> and <desc> elements play an important role in this regard. SPEAR uses these elements to offer prose descriptions of the data that can capture and display the requisite nuance.

Example 15. Event description.

```
  <event ana="http://syriaca.org/keyword/council-of-chalcedon-451 http://
syriaca.org/keyword/patriarchs">
   <desc>
    After <choice>
     <reg><date notBefore="0525-06-27" calendar="#gregorian">27 June 525</date></
reg>
     <orig><date notBefore-custom="0836-09-27" datingMethod="#seleucid-
SyriacMonths">27 Ḥazirān A. Gr. 836</date></orig>
    </choice>,
    <persName ref="http://syriaca.org/person/2296"> Pawlā </persName> repented of
his opposition to the Council of <placeName ref="http://syriaca.org/place/622">
Chalcedon </placeName>,
    petitioned <persName ref="http://syriaca.org/person/2284"> Justinian </
persName> for the return of his episcopal see, and wrote a libellus affirming the
council to
    <persName ref="http://syriaca.org/person/2278"> Euphrasius the Patriarch </
persName>. </desc>
  </event>
```

Example 15 shows how the various parts of an event factoid come together to produce data out of a reference to an event in a historical source. The encoders have attached keywords on the <event> element, and then offered a rich description of the event that captures the dating system employed in the source document along with references to relevant persons and places. The pointers applied in this factoid as well as elsewhere in the data set create a graph of relationships between these concepts, persons, and places amenable to various forms of computation. These machinable data are at the same time prose descriptions capturing the encoder's interpretation of the factoid—an interpretation that would be difficult or impossible to represent in a relationship graph alone.

37 SPEAR also allows a <note> element to appear as a child of <ab>. In such instances it is used to express the scholarly assessment of the encoder regarding the veracity of the content of the factoid.

Example 16. A note on the factoid.

```
  <note type="errata">
   Since the first session of the Council of Chalcedon began on 8 October (Teshri
I) of 451, the correct year
   according to the Greek calendar (A. Gr.) would be 763, not 762.
   <bibl type="secondary">
    <ptr target="http://syriaca.org/bibl/660"/>
    <citedRange unit="p">113</citedRange>
   </bibl></note>
```

Recall that factoids contain the encoder's interpretation of a primary source text. This `<note>` element requires a `@type` attribute with `"dubia"`, `"errata"`, or `"incerta"` for expressing different types of uncertainty encoders can use to express their position on the prosopographical datum contained in that factoid. A text node contains a prose description of that uncertainty and optional `<bibl>` elements allow the encoder to indicate a scholarly source discussing the issue.

38 Occasionally more than one factoid deals with the same or a closely related event. SPEAR cannot make direct links between events themselves because events do not receive URIs and related authority files. When encoders are aware of another factoid dealing with a related event, they can include a `<note type="relatedEvent">`.

Example 17. A note indicating multiple factoids encoding related events.

```
  <ab type="factoid" subtype="event" xml:id="factoid-388" resp="#dschwartz">
   <idno type="URI">http://syriaca.org/spear/8559-388</idno>
   <listEvent>
    <event ana="http://syriaca.org/keyword/natural-disasters">
     <desc>
      In <choice>
       <reg><date notBefore="0498-10" notAfter="0499-09"
calendar="#gregorian">498/9</date></reg>
       <orig><date when-custom="0810" datingMethod="#seleucid">A. Gr. 810</date></
orig>
      </choice>
      the city of <placeName ref="http://syriaca.org/place/2574"> Nicopolis </
placeName>
     was destroyed by an earthquake. </desc>
```

```
    </event>
   </listEvent>
   ...
   <note type="relatedEvent">
    <listRelation>
     <relation ref="http://syriaca.org/keyword/proximate-event" type="event"
mutual="http://syriaca.org/spear/8559-388 http://syriaca.org/spear/8559-389
http://syriaca.org/spear/8559-396">
      <desc> The following factoids deal with closely related events:
       <rs ref="http://syriaca.org/spear/8559-388">8559-388</rs>,
       <rs ref="http://syriaca.org/spear/8559-389">8559-389</rs>,
       <rs ref="http://syriaca.org/spear/8559-396">8559-396</rs>
      </desc>
     </relation>
    </listRelation>
   </note>
  </ab>
```

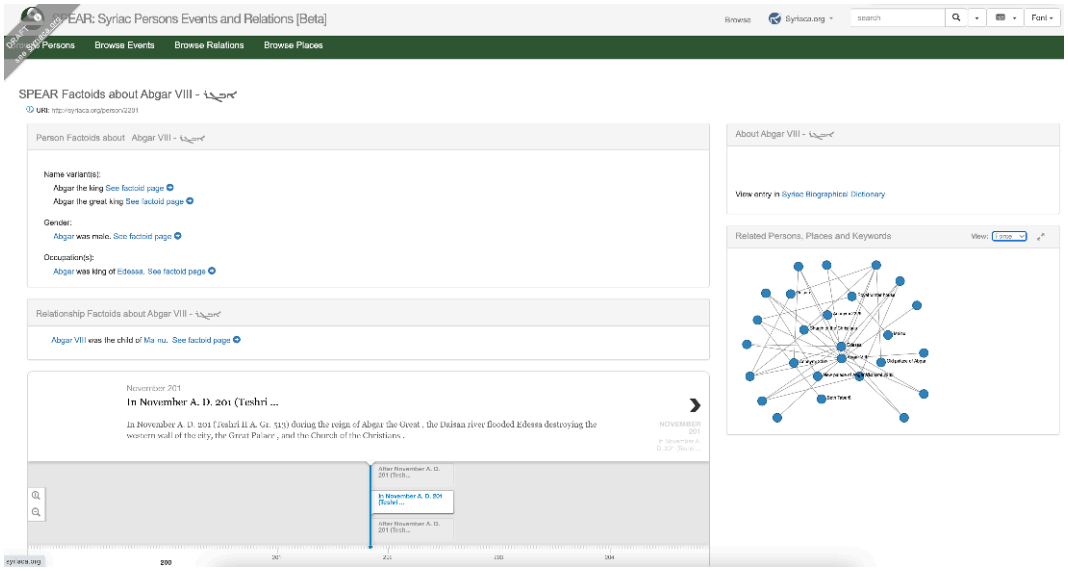
This `<relation>` element must have a `@mutual` attribute with the URIs for two or more factoids. SPEAR uses a `@ref` attribute on `<relation>` to indicate the type of `<relation>`. The acceptable values are either `"http://syriaca.org/keyword/same-event"` or `"http://syriaca.org/keyword/proximate-event"`. The former is reserved for explicit references to the same event encoded in different event factoids while the latter creates a link between factoids that deal with closely related events. This connection would include things like causal connections between events and containing relationships between events.

5. Visualizing Factoid Data in the Srophé Application

39 Structuring factoids and embedding URIs in them as described above makes it possible to visualize and query the resulting data graph in a variety of ways. The scope of this paper does not allow a full discussion of these issues, but a demonstration of the utility of using the TEI requires at least a brief discussion of how the Srophé Application uses eXist-DB to display SPEAR data. Factoids can be viewed individually with links to related data as shown above in figure 2, but they can also be aggregated by source text or across multiple source texts according to person, keyword, or

place URIs. Figure 3 shows the aggregation of factoids about one individual. The events associated with this person are displayed on a timeline. Related persons, places, and keywords are also displayed in a force graph (shown), sankey graph, or a list. The Srophé Application also facilitates faceted browse of factoid data and search by multiple fields. Thus, TEI-encoded factoids can not only capture nuanced interpretations in structured, transparent assertions containing linked-data pointers; they can also benefit from the same kinds of aggregate, dynamic visualizations that relational databases provide.

Figure 3. Aggregation of factoid data about a person showing force graph and event timeline.



6. Conclusion

40 SPEAR demonstrates the value of bringing the TEI to the task of modeling prosopographical factoids. While this may not fit the process all factoid prosopography projects use, it is a viable alternative to RDBs and one that fits the ethos of the TEI, that is, open and well-documented community-based encoding. The ability of the user to view the native TEI data, combined with the use of a customized TEI schema for validation and documentation, enhances the transparency of the data. Applying linked data standards to the modeling of prosopographical data in the TEI allows for the integration of structured data about persons and semi-structured, marked-up prose interpretations of the historical sources from which they are derived. These prose interpretations

enrich the resulting data graph and easily display a human-readable version of the data with hyperlinks that connect the user to additional linked data. Using the TEI brings transparency to the model by maintaining a close connection between each factoid datum and the text from which the encoder derives it. Each factoid makes a prosopographical assertion about a text, indicates the scholar responsible for that textual interpretation, and points the user to the text. In this way, SPEAR uses the TEI both to model prosopographical factoids and to model the scholarly process by which the factoids are produced.

41 Combining this approach with the LOD infrastructure of Syriaca.org further enhances the data-oriented approach to the textual interpretation of SPEAR. Pointers in `@ref` attributes do not point to project-specific `@xml:id` attributes in a stand-off markup file but to linked data that have an independent purpose. These URIs are also used by other projects: digital research projects, encoded reference works, etc. The Taxonomy URIs SPEAR applies to events, occupations, etc. are the same URIs used to catalog and describe bibliography, thus creating links between these and other data sets. This approach demonstrates a way that autonomous and semi-autonomous research projects can maintain their distinctive research agendas and methodologies while providing critical metadata for general use. This level of linking results in a different kind of born-digital application for the TEI, one in which the encoding of primary source texts remains important but can also be transcended through the encoding of born-digital interpretations of those texts.

BIBLIOGRAPHY

Blackwell, Christopher and Neel Smith. 2014. "The Canonical Text Services URN Protocol Specification." https://cite-architecture.github.io/ctsurn_spec/.

Bodard, Gabriel. 2021. "Linked Open Data for Ancient Names and People." In Linked Open Data for the Ancient Mediterranean: Structures, Practices, Prospects, edited by Sarah E. Bond, Paul Dilley, and Ryan Horne. ISAW Papers 20.http://hdl.handle.net/2333.1/zs7h4fs8.

Bradley, John. 2005. "Documents and Data: Modelling Materials for Humanities Research in XML and Relational Databases." *Literary and Linguistic Computing* 20, no. 1: 133–51. https://doi.org/10.1093/llc/fqh048.

———. 2014 “Silk Purses and Sow’s Ears: Can Structured Data Deal with Historical Sources?” *International Journal of Humanities & Arts Computing: A Journal of Digital Humanities* 8, no. 1: 13–27. https://doi.org/10.3366/ijhac.2014.0117.

Bradley, John and Harold Short. 2005. “Texts into Databases: The Evolving Field of New-style Prosopography.” *Literary and Linguistic Computing* 20, no. 1: 3–24. https://doi.org/10.1093/llc/fqi022.

Cameron, Averil. 2003. *Fifty Years of Prosopography: The Later Roman Empire, Byzantium and Beyond*. Proceedings of the British Academy 118 Oxford: Oxford University Press.

Gibson, Nathan P., David A. Michelson, and Daniel L. Schwartz. 2017. “From Manuscript Catalogues to a Handbook of Syriac Literature: Modeling an Infrastructure for Syriaca.org.” *Journal of Data Mining and Digital Humanities* Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages: 1–15. https://jdmdh.episciences.org/1395.

Jannidis, Fotis and Julia Flanders. 2018. “A Gentle Introduction to Data Modeling.” In *The Shape of Data in Digital Humanities: Modeling Texts and Text-Based Resources*, edited by Jannidis Fotis and Julia Flanders, 26–95. London: Routledge.

Keats-Rohan, K. S. B. 2000. “Prosopography and Computing: A Marriage Made in Heaven?” *History and Computing*. 12, no. 1: 1–11. https://doi.org/10.3366/hac.2000.12.1.1.

King, Daniel. 2018-12-14*The Syriac World*. London: Routledge. https://doi.org/10.4324/9781315708195.

Marsh, David. 2014–01–17. “A Factoid is not a Small Fact.” *The Guardian*. Mind Your Language. https://www.theguardian.com/media/mind-your-language/2014/jan/17/mind-your-language-factoids.

Michelson, David A. 2021. “Using Linked Open Data to Model Cultural Heritage Information: The Research Questions and Data Structures of the Syriaca.org Knowledge Graph.” In Linked Open Data for the Ancient Mediterranean: Structures, Practices, Prospects, edited by Sarah E. Bond, Paul Dilley, and Ryan Horne. ISAW Papers 20. http://hdl.handle.net/2333.1/69p8d8cc.

Pasin, Michele and John Bradley. 2015. “Factoid-based Prosopography and Computer Ontologies: Towards an Integrated Approach.” *Digital Scholarship in the Humanities*. 30, no. 1: 86–97. https://doi.org/10.1093/llc/fqt037.

Patton, Glenn E. 2009. *Functional Requirements for Authority Data: A Conceptual Model*. Berlin and Boston: De Gruyter.

Sauerman, Leo and Richard Cyganiak. “Cool URIs for the Semantic Web.” W3C Interest Group Note. https://www.w3.org/TR/2008/NOTE-cooluris-20081203/.

TEI Consortium. 2020. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 4.1.0. Last updated August 19. N.p.: TEI Consortium. https://tei-c.org/Vault/P5/4.1.0/doc/tei-p5-doc/en/html.

Viglianti, Raffaele. 2016. "Why TEI Stand-off Markup Authoring Needs Simplification." *Journal of the Text Encoding Initiative*. 10. http://journals.openedition.org/jtei/1838.

NOTES

1 For an extended discussion of the Syriaca.org data model, see Michelson 2021. On Syriac cultural heritage, see, accessed January 31, 2022 https://gallery.library.vanderbilt.edu/exhibits/syriaca and King 2018.

2 For more on the Srophé Application, see section 5.

3 For a list of CCH/DDH prosopographies with brief introductions to each, see, accessed January 31, 2022, Factoid Prosopographies at CCH/DDH KCL at: https://factoid-dighum.kcl.ac.uk/factoid-prosopographies-at-cchddh-kcl/.

4 However, Bradley and Short (2005, 17) are careful to note that the database is a model of prosopography itself, not a model of the text(s) from which the prosopographical material is drawn. SPEAR seeks to accomplish the same goal using the TEI.

5 Syriaca.org data for persons, places, and works meet the criteria for authority files articulated by Patton 2009, 83.

6 On some different approaches to TEI stand-off markup see Viglianti (2016).

7 For more on the SNAP:DRGN project, see Bodard, 2021.

8 For a helpful graph of the SNAP ontology, see http://snap.dighum.kcl.ac.uk/img/OwlVizImage.png.

9 For a discussion of how Syriaca.org handles these challenges in establishing authority files for works, see Gibson, Michelson, and Schwartz (2017).

10 Some person factoids (like birth or occupation) might imply that an event occurred to constitute that personal characteristic. Thus, they could alternatively be modeled as events (see section 4.1). However, SPEAR encourages encoders to stay close to the text, modeling the assertion as a person factoid when they interpret the source text to be giving a personal description rather than relating an event.

11 Syriaca.org employs webforms that SPEAR encoders can use to create the required entities when they do not already exist in Syriaca.org data. These forms reserve a URI in the proper data set and produce a "stub" TEI XML record that editors of the relevant data sets can develop and incorporate into the Syriaca.org data graph.

12 It is unfortunate that the TEI employs the element `<nationality>`. For the vast majority of recorded history (even if not for the vast majority of extant texts) there was no such thing as a nation-state. SPEAR uses the `<nationality>` element even though it allows `<nationality>` to contain a `<placeName>` with a `@ref` attribute pointing to any type of place entity of which someone could be a citizen, "citizenship" being a more meaningful category for the primary source base of SPEAR than is "nationality." Factoids of this kind are contained in a `<ab subtype="citizenship">`. In this way, SPEAR encoding does not distinguish between a person being a citizen of a state-level entity, for example the Roman or Persian Empire, and a settlement-level entity, for example the city of Antioch.

13 See The CTS URN · The CITE Architecture and the formal CTS URN specification at https://cite-architecture.github.io/ctsurn_spec/. A related and more recent specification not yet implemented by the Digital Syriac Corpus is (DTS) - Distributed Text Services. On the relationship between CTS and DTS, see https://distributed-text-services.github.io/specifications/FAQ.html#what-is-the-relationship-between-dts-and-cts-are-they-redundant.

AUTHORS

DANIEL L. SCHWARTZ

Daniel L. Schwartz is Associate Professor of History and Associate Director of the Center of Digital Humanities Research at Texas A&M University. He co-directs (with David A. Michelson) Syriaca.org, a TEI-encoded linked-open data project dedicated to advancing scholarship in the field of Syriac Studies and to preserving Syriac cultural heritage. He is the general editor of Syriaca.org's *Syriac Persons, Events, and Relations* (*SPEAR*, an online prosopographical research tool, and the co-editor (with Sergey Minov) of *The Syriac Taxonomy*, a domain ontology for Syriac Studies.

NATHAN P. GIBSON

Nathan P. Gibson is a Research Associate in the Institute of Near and Middle Eastern Studies at Ludwig-Maximilian University of Munich. He directs the project Communities of Knowledge (Usaybia.net, funded by the "Kleine Fächer – Große Potenziale" program of the German Federal Ministry of Education and Research), which uses TEI factoids to encode and analyze networks of interreligious collaboration in the medieval Middle East. He is a co-editor of the TEI-based clavis *New Handbook of Syriac Literature* (with David A. Michelson and Jeanne-Nicole Mellon Saint-Laurent at Syriaca.org) and an associate editor of the Bibliography of the Arabic Bible (with Ronny Vollandt, general editor, at Biblia-Arabica.com).

KATAYOUN TORABI

Katayoun Torabi is an Instructional Assistant Professor in the Department of English at Texas A&M University and Project Manager of the New Variorum Shakespeare at the Center of Digital Humanities Research. She specializes in Old and Middle English literature, with strong research and teaching interests in the digital humanities. She is currently working on a digital project that investigates connections between purgation and incarceration in pre- and post-mortem punitive spaces in Old and Middle English literature.