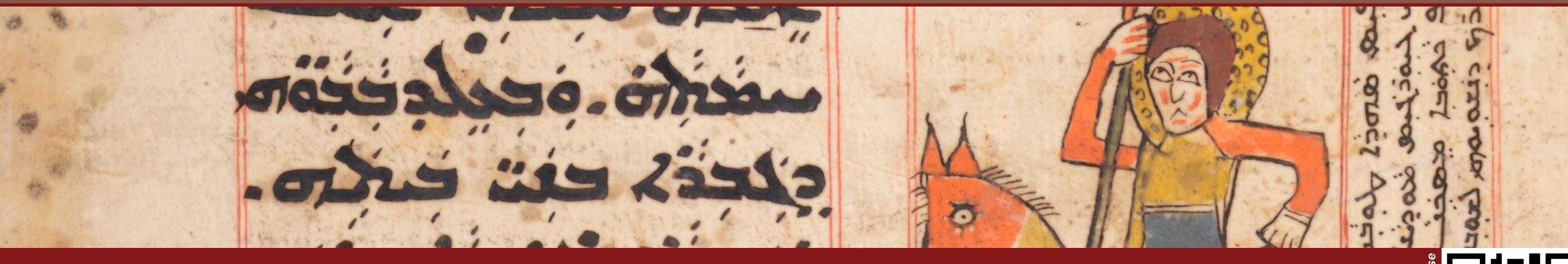# Modeling a Body of Literature in TEI

by Nathan P. Gibson and David A. Michelson, Vanderbilt University

Syriaca.org

The New Handbook of Syriac Literature

edited by Nathan P. Gibson, David A. Michelson, and Jeanne-Nicole Mellon Saint-Laurent

Vol. 1, *Bibliotheca Hagiographica Syriaca Electronica* (J.-N. Mellon Saint-Laurent, D.A. Michelson, U. Zanetti, and C. Detienne, eds.)

Abstract

The New Handbook of Syriac Literature (*NHSL*) is a born-digital TEI-encoded reference work for the study of Syriac literature. The first volume, *Bibliotheca Hagiographica Syriaca Electronica,* was published by Syriaca.org in 2016 using a simple TEI schema to describe a single genre (hagiography).[1] In preparation for expanding the *NHSL* to include other genres, Syriaca.org is revising this TEI schema. The authors actively seek feedback, suggestions, and criticism concerning this revised schema.

The Work Entity

The core entity being modeled in the *NHSL* is a "work", which is "a distinct intellectual or artistic creation" as defined by FRBR.[3] In *NHSL* each work is encoded within a <bibl> element comprising the entire <body> of the TEI document. FRBR work entities are abstract concepts distinct from any historical exemplar. Thus Plato's "Allegory of the Cave" would be a "work" distinct from any extant individual manuscript copies or printed editions of it. The benefit of the FRBR model for *NHSL* is its capability for grouping related manuscript items, editions, and translations around a conceptual work. For practical reasons, *NHSL* has simplified the FRBR model into two entities: "works" and "citations". This two-fold organization is useful because it renders the *NHSL* easily compatible with *de facto* descriptive practices (as reflected in RDF vocabularies) of major online catalogues such as worldcat.org, openlibrary.org, dnb.de, and catalog.perseus.org.[4] Such compatibility is necessitated by our goal of linking each "work" to related digital, print, and manuscript citations. Theoretically, our work could be expanded in the future to include the full FRBR taxonomy.

TEI as a Metadata Format

Given our project goals, we found a number of advantages in using TEI over other metadata and bibliographic formats such as MARC XML, Dublin Core, MADS, and EAD. In methodological terms, *NHSL* follows the documentation norms of the discipline of History, requiring extensive use of footnotes to indicate the provenance of information. Of the above data formats, only TEI's @source attribute combined with TEI's model.biblLike class of elements provided the comprehensive provenance mechanism needed to meet this standard. While other formats do have some ability to indicate sourcing (including @source in EAD), their sourcing mechanisms are not universally available throughout the data model nor able to reproduce a bibliography of citations. Because TEI permits customization such that @source could be used in a standard way with any element, we are able to record and attribute multiple and even contradictory historical claims to serve our scholarly users. A second benefit to using TEI comes from the relative compatibility of the TEI data model with our project-specific needs for RDF serialization. TEI customization provided the simplest method of embedding URIs directly in the data (via <idno>, @ref, <relation>, etc.). While other metadata formats are also capable of serialization into RDF, the extensibility and flexibility of TEI was a particular fit. Third, we found that the generic nature of the TEI offered advantages over more domain-specific data models for metadata. MARC XML, Dublin Core, MADS and EAD are primarily designed with the needs of library or archive use in mind, specifically the creation of standards or "authority" fields. Because in many cases the historical methodology of our project explicitly precludes privileging a uniform title or attributed author for a work above others, we needed the ability to record multiple valid titles and authors for scholarly purposes without having to designate one as authoritative or preferred. Only the TEI allowed such a neutral approach. (We do also recognize the need of catalogues for authority files. By using TEI as a base format, we are able to serialize MADS records with uniform titles for such use. The reverse would not, however, be possible had we begun with MADS. In sum, the granularity of our data in TEI makes crosswalks possible to other, less-granular bibliographic formats.) Finally, TEI's flexible <note> and <bibl> mechanisms allow us to include a variety of semi-structured textual, descriptive, and bibliographic information including excerpts from the texts of the works ("incipits", "explicits", "colophons", etc.).

Rationale

Syriac is a language which once flourished on the Mesopotamian plateau. A dialect of Aramaic, Syriac was widely used during much of the first millennium of the common era and continues to be used today by a world-wide diaspora. There is currently no comprehensive inventory of Syriac literature. Syriaca.org is a research collaborative building digital tools to "re-sort" and "re-orient" the field of Syriac literature.[2] One such tool is *NHSL*. Each *NHSL* entry uniquely identifies a "work" (see below) and links to related digital and print resources if possible. The *NHSL* also seeks to describe works even if they have never been edited or published, by providing titles and excerpts, author information, manuscript attestations, bibliography, and descriptions of language, genre, and/or subject. The *NHSL* also documents the relationships between works, authors, manuscripts, and the geographic places with which they are associated. Although there is a long scholarly precedent for using TEI to encode ancient and medieval texts, past practice has focused on describing specific manuscripts (or text-bearing objects) or creating editions of works. We found that the use of TEI to encode born-digital metadata about works was less common. Similarly, in the TEI community at large the <bibl> element and its sibling <biblStruct> are often used to represent bibliographic information for specific publications of a work, but rarely for "works" in the abstract or conceptual sense. This poster presents our approach to modeling such a body of literature in TEI.

TEI

```
<text>
  <body>
    <bibl xml:id="work-1553" type="lawd:ConceptualWork"
      ana="http://syriaca.org/documentation/work-subject-
      classifications.xml#hagiography
      http://syriaca.org/documentation/work-subject-classifications.xml#memra">
      <title xml:lang="name1553-1" xml:lang="syr" source="#bib1553-1">
        syriaca-tags="#syriaca-headword"></title>
      <title xml:id="name1553-2" xml:lang="fr-x-bhs"
        source="#bib1553-1">Maryam Magdalāytā</title>
      <title xml:id="name1553-3" xml:lang="fr" source="#bib1553-1">
        Marie Madeleine (texte)</title>
      <title xml:id="name1553-4" xml:lang="en" source="#bib1553-1"
        syriaca-tags="#syriaca-headword">Mary Magdalene (text)</title>
      <author
        source="#bib1553-1" ref="http://syriaca.org/person/13"
        role="http://syriaca.org/documentation/author-editor-roles.xml#attributed">
        <persName xml:lang="en-x-gedsh">Ephrem</persName> —
        <persName xml:lang="syr">ܐܦܪܝܡ</persName>
        <persName xml:lang="fr">Éphrem</persName>
      </author>
      <idno type="URI">http://syriaca.org/work/1553</idno>
      <idno type="URI">http://viaf.org/viaf/0123456789</idno>
      <idno type="BHS">1466</idno>
      <idno type="CPG">3952</idno>
      <textLang mainLang="syr"/>
      <date notBefore="0250" notAfter="0550">3rd/4th century</date>
      <extent>8 strophe</extent>
      <note xml:lang="en" type="abstract">One of the memre "On Reproof."
      </note>
      <note xml:lang="syr" type="incipit" source="#bib1553-1">
        <quote>ܩܘܡ ...</quote>
      </note>
      <note xml:lang="syr" type="explicit" source="#bib1553-1">
        <quote>...</quote>
      </note>
      <note xml:lang="en" type="explicit" source="#bib1553-1">
        <quote>Amen, amen, I say to you, I will write your name in the Gospel; go
        woman, your debts have been pardoned and all your sin is erased.
        </quote>
      </note>
      <bibl type="lawd:Edition" source="#bib1553-1" xml:id="bib1553-2">
        <title>Des heiligen Ephraem des Syrers Sermones II</title>
        <ptr target="http://syriaca.org/bibl/1815"/>
        <citedRange unit="vol">2</citedRange>
        <citedRange>col. 313-338</citedRange>
        <listRelation>
          <relation type="mssWitnesses" active="#bib1553-2" ref="dct:source"
            passive="#bib1553-4" source="#bib1553-1"/>
        </listRelation>
      </bibl>
      <bibl type="lawd:Translation" source="#bib1553-1" xml:id="bib1553-10">
        <title>CSCO 312 = Script. Syri 135</title>
        <ptr target="http://syriaca.org/bibl/2085"/>
        <citedRange unit="pp">99-113</citedRange>
        <lang>de</lang>
        <listRelation>
          <relation type="translationSource" active="#bib1553-10"
            ref="dct:source" passive="#bib1553-3" source="#bib1553-1"/>
        </listRelation>
      </bibl>
      <bibl type="lawd:Translation" source="#bib1553-1" xml:id="bib1553-6">
        cf. <title>CPG</title>
        <ptr target="http://syriaca.org/bibl/2081"/>
        <lang>ar</lang>
        <listRelation>
          <relation active="#bib1553-6" ref="lawd:embodies"
            passive="http://syriaca.org/work/1553"/>
        </listRelation>
      </bibl>
      <bibl xml:lang="fr" type="lawd:WrittenWork" source="#bib1553-1"
        xml:id="bib1553-4">Londres, British Library, Add. 14651, f. 196-201
        <ptr target="http://syriaca.org/manuscript/01234#a1"/>
        <listRelation>
          <relation active="#bib1553-4" ref="lawd:embodies"
            passive="http://syriaca.org/work/1553" source="#bib1553-1"/>
        </listRelation>
      </bibl>
      <bibl type="lawd:Citation" xml:id="bib1553-1">
        <title level="m" xml:lang="la">Bibliotheca Hagiographica Syriaca</title>
        <ptr target="http://syriaca.org/bibl/649"/>
        <citedRange unit="entry">1466</citedRange>
      </bibl>
      <listRelation>
        <relation active="http://syriaca.org/work/1553"
          ref="syriaca:commemorated"
          passive="http://syriaca.org/person/1884" source="#bib1553-1"/>
        <relation part" active="http://syriaca.org/work/1553"
          ref="dct:isPartOf" passive="http://syriaca.org/work/5432"
          sortKey="1"/>
      </listRelation>
    </bibl>
  </body>
</text>
```

RDF Class as @type

To aid RDF serialization, @type attributes throughout the document may be prefixed RDF classes.[5]

Subjects & Genres

Syriac literature does not yet have an established subject or genre classification. One or more links in tei:bibl/@ana point to informal tags that describe the work in terms of relevant subjects, genres, and Syriac-language terms. This preliminary tagging may lead to more formal taxonomies in the future.

Titles

Multiple titles attested in manuscripts or in scholarly use are collected. There is no standard authority defining uniform titles for Syriac works, and Syriaca.org's identifier for the work is a URI rather than a title. However, the custom syriaca-tags="#syriaca-headword" is used to label titles preferred for display in page headers and browse lists.

Language & Script Codes

Syriaca.org uses the ISO 639 macrolanguage code "syr" ("Syriac") for all literary Syriac. Transliterations that reflect a particular system are marked by the code of the target language, plus a custom extension representing the transliteration system (e.g., "fr-x-bhs").[6]

Sources

To meet the demands of its scholarly audience (and in anticipation of a proposed change in the TEI to allow @source globally), Syriaca.org uses @source to identify the sources of any information encoded in text nodes including titles, authorship claims, and other information about the work.

Authorship

Authorship of Syriac works is often disputed. @role designates authorial or editorial responsibility for a work with whatever level of precision editors choose. Competing or multiple claims of authorship can also be indicated. Pseudonymous works can point to both the URI of a real author (usually anonymous) and to the URI of the person whose name was assumed (with @role="pseudo").

Identifiers

<idno> with @type="URI" declares the Syriaca.org URI for the work entity, as well as any URIs in other datasets that identify the same work. Stable identifiers from standard print reference works are also included here using standard abbreviations for their @type values.

Note Types

Several types of descriptive <note> elements may be included, distinguished by @type. "abstract" provides the briefest description of the work useful for identification. "scope" describes its relationship to larger works or collections containing it. "content" summarizes its contents. "disambiguation" explains how to distinguish it from works with which it has been or could be confused.

Incipits, Explicits, and Other Excerpts

The <note> element is also used, with @type and @source, to provide incipits, explicits, or other excerpts that may be needed to identify the work. (The TEI Guidelines do not allow <incipit> and <explicit> as children of <bibl> and define them exclusively for manuscripts, not abstract works.) When available, translations are provided and tagged with the appropriate language code. One open question is how to distinguish two levels of citation: a direct quotation from the item indicated by @source versus a translation made from that item.

Relationships to Textual Manifestations/Embodiments

Manuscript witnesses, editions, and translations (ancient or modern) of the work are contained in <bibl> elements nested inside the work <bibl> and marked with RDF classes in @type. <ptr> elements identify the URI of the relevant entity in Syriaca.org's Works Cited or Manuscripts modules, which is used to supply a full citation on the HTML page.

A <relation>, placed here for ease of editing, relates editions and translations to their sources using the RDF property "dct:source", and manuscript items to the work entity using "lawd:embodies". These can be serialized easily into RDF.[7]

A remaining question is how to encode the language of translations, particularly ancient versions. <bibl> elements for ancient versions typically refer to a published edition/translation or to a reference work describing the version. Neither @xml:lang nor <textLang> can be used to show the version language, since the former would indicate the language of the citation itself, while the latter would refer to the language of the cited item, which is not necessarily the same as that of the version. At present, we have provisionally included a <lang> element inside the <bibl>.

Secondary Literature

<bibl> elements with @type="lawd:Citation" supply references to secondary literature using a <ptr> to the URI of the cited entity in the Works Cited module.

Work-Person and Work-Work Relations

The <listRelation> element placed directly inside the work <bibl> records additional relations between the work and other entities, again defined by RDF properties. In *NHSL* vol. I, the persons commemorated in hagiographic works are recorded here with the URI from their entry in Syriaca.org's *Qadishe: Guide to the Syriac Saints.* Works may also have relations to other works, for example as part of a larger containing work ("dct:isPartOf") or as belonging to a hagiographical tradition that includes multiple works ("syriaca:isPartOfTradition"). For maintainability, these are encoded only in the child work, but visualized for both child and parent works.

Notes

1. See also Jeanne-Nicole Mellon Saint-Laurent, "Gateway to the Syriac Saints: A Database Project," *The Journal of Religion, Media and Digital Culture* 5, no. 1 (May 3, 2016): 183–204, https://www.jrmdc.com/journal/article/view/78.

2. David Allen Michelson, "Mixed Up by Time and Chance? Using Digital Methods to 'Re-Orient' the Syriac Religious Literature of Late Antiquity," *The Journal of Religion, Media and Digital Culture* 5, no. 1 (May 3, 2016): 136–82, https://www.jrmdc.com/journal/article/view/80; "Syriaca.org as a Test Case for Digitally Re-Sorting the Ancient World," in *Ancient Worlds in Digital Culture*, ed. Claire Clivaz, Paul Dilley, and David Hamidović (Leiden: Brill, 2016).

3. IFLA Study Group on the Functional Requirements for Bibliographic Records, "Functional Requirements for Bibliographic Records: Final Report," International Federation of Library Associations and Institutions, 1998 (revised 26 December 2007), http://archive.ifla.org/VII/s13/frbr/frbr_current3.htm#3.2.

4. Nathan P. Gibson, David A. Michelson, and Daniel L. Schwartz, "From Manuscript Catalogues to a Handbook of Syriac Literature: Modeling an Infrastructure for Digital Humanities," *Journal of Data Mining and Digital Humanities* (forthcoming), https://arxiv.org/abs/1603.01207.

5. For the ontology, see "The LAWD Ontology," last modified September 3, 2014, http://lawd.info/.

6. "Language and Script Identifiers," Syriaca.org, last modified August 17, 2016, http://syriaca.org/documentation/isostandards.html.

7. Anna Jordanous et al., "Exploring Manuscripts: Sharing Ancient Wisdoms across the Semantic Web," in *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics* (New York: ACM, 2012), doi:10.1145/2254129.2254184.

Funding