1. Introduction and 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. Since there is currently no comprehensive inventory of Syriac literature, researchers must consult outdated literature histories or dozens of manuscript catalogues in order to locate or identify Syriac texts. Syriaca.org is a research collaborative building digital tools to “re-sort” and “re-orient” the field of Syriac literature (Michelson 2016a, Michelson 2016b). One such tool is The New Handbook of Syriac Literature (NHSL), 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) (Saint-Laurent et al. 2016; see also Saint-Laurent 2016, Zanetti 2016). In preparation for expanding the NHSL to genres besides hagiography, Syriaca.org is revising this TEI schema.
Each NHSL entry uniquely identifies a work 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. In addition, the NHSL 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 paper presents our approach to modeling such a body of literature in TEI and invites feedback, suggestions, and criticism concerning the revised schema.
Figure 1: Cover Image of the *Bibliotheca Hagiographica Syriaca Electronica*

Bibliotheca Hagiographica Syriaca Electronica

New Handbook of Syriac Literature, Volume 1

Jeanne-Nicole Mellon Saint-Laurent,
David A. Michelson, Ugo Zanetti,
and Claude Detienne

Journal of the Text Encoding Initiative,
The core entity being modeled in the NHSL is a work, which is “a distinct intellectual or artistic creation” as defined by FRBR (IFLA 1998). 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 (Gibson,
Michelson, and Schwartz 2017). 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.

3. TEI as Metadata Format

Given our project goals, we found a number of advantages in using TEI over other metadata and bibliographic formats such as MARCXML, Dublin Core, MADS, and EAD. First, 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 the 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. As of version 3.1.0 of the TEI Guidelines, @source can be used in a standard way with any element, enabling us to record and attribute multiple and even contradictory historical claims to serve our scholarly users (TEI 2016). 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. MARCXML, 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.).

4. Example Description of a Syriac Hagiographical Work in TEI

4.1 RDF Classes as @type

To aid RDF serialization, @type attributes throughout the document may be prefixed RDF classes.¹

Example 1: Example of RDF Classes in @type

```xml
<bibl xml:id="work-1553" type="lawd:ConceptualWork"> ... <bibl type="lawd:Edition" source="#bib1553-1" xml:id="bib1553-2"> ... </bibl></bibl>
```

4.2 Subjects and Genres

Syriac literature does not yet have an established subject or genre classification. One or more links in the <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.

Example 2: Example of Subjects and Genres in <bibl> @ana

```xml
<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"> ... </bibl>
```

4.3 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.

¹ Journal of the Text Encoding Initiative,
Example 3: Example of Titles including Headwords

```xml
<title xml:id="name1553-1" xml:lang="en" source="#bib1553-1" syriaca-tags="#syriaca-headword">Mary Magdalene (text)</title>
<title xml:id="name1553-2" xml:lang="syr" source="#bib1553-1" syriaca-tags="#syriaca-headword">MaryamMa#dəlāyţā</title>
<title xml:id="name1553-3" xml:lang="fr-x-bhs" source="#bib1553-1">Marie Madeleine (texte)</title>
<title xml:id="name1553-4" xml:lang="fr" source="#bib1553-1">Marie Madeleine (texte)</title>
```

Figure 3: HTML Visualization of Work Titles

![HTML Visualization of Work Titles](image)

4.4 Language and Script Codes

Syriaca.org uses the ISO 639 macrolanguage code "syr" ("Syriac") to construct a BCP 47-compliant language tag 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").

4.5 Sources

To meet the demands of its scholarly audience (and in line with the change in the TEI Guidelines that allows @source globally [TEI 2016]), 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 (see Example 4).
4.6 Authorship

Authorship of Syriac works is often disputed. The @ref of <author> or <editor> points to the URI of an author from *A Guide to Syriac Authors* (Michelson and Gibson 2016), while the @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 a corresponding @role value).

Example 4: Example of Sources and Authorship

```xml
<title xml:id="name1553-1" xml:lang="syr" source="#bib1553-1" syriaca-tags="#syriaca-headword">########</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">#####</persName> (<persName xml:lang="fr">Éphrem</persName>)
</author>
... <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>
```
4.7 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.

Example 5: Example of Work Identifiers

<idno type="URI">http://syriaca.org/work/1553</idno>
<!-- Example authority file URI -->
<!-- <idno type="URI">http://viaf.org/viaf/0123456789</idno> -->
<idno type="BHS">1466</idno>
<idno type="CPG">3952</idno>

Figure 5: HTML Visualization of Work Identifiers
4.8 Descriptive Notes

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. (See example 6.)

4.9 Incipits, Explicits, and Other Textual 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 [TEI 2016].) 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.

Example 6: Example of Work Abstract and Excerpts

```
<note xml:lang="en" type="abstract">One of the memre "On Reproof."</note>
<note xml:lang="syr" type="incipit" source="#bib1553-1">
  <quote>##### ## #####̈# ### ### ###### #### ##### #### ####̈# ####### ####### ## ### ### #### #### ####### #### ####</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>
```
4.10 Relationships to Textual Manifestations and 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 are used to supply a full citation on the HTML page, with links to catalogue records or scans of the item in external systems.

A `<relation>`, placed here for ease of editing, relates editions and translations to their source texts using the RDF property “dct:source”, and manuscript items to the work entity using “lawd:embodies”. These can be serialized easily into RDF (Jordanous et al. 2012).

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>`. 

---

*Journal of the Text Encoding Initiative,*
Example 7: Relationships among Editions, Manuscripts, Translations, and Ancient Versions

```
<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="syriaca:AncientVersion" 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" source="#bib1553-1"/>
  </listRelation>
</bibl>

<bibl type="syriaca:Manuscript" source="#bib1553-1" xml:id="bib1553-4">
  <title type="shelfmark">London, British Library, Add. 14651</title>
  <ptr target="http://syriaca.org/manuscript/419"/>
  <biblScope unit="folio" from="196" to="201">196-201</biblScope>
  <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>
```
4.11 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. These are cited within the document using @source and visualized in the HTML as superscript numbers linked to footnote anchors at the bottom of the page.

4.12 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* (Saint-Laurent and Michelson 2016). Works may also have relations to other works, for example as part of a larger containing work (@ref="dct:isPartOf") or as belonging to a hagiographical tradition that includes multiple similar works (@ref="skos:broadMatch"). For maintainability, these are encoded only in the child work, but visualized for both child and parent works.
Example 8: Examples of Work-Person and Work-Work Relationships

<listRelation>
  <relation name="syriaca:commemorated" active="http://syriaca.org/work/1553"
  passive="http://syriaca.org/person/1884" source="#bib1553-1"/>
  <relation type="part" active="http://syriaca.org/work/1553" ref="dct:isPartOf"
  passive="http://syriaca.org/work/10525">
    <desc>
      <label type="order" subtype="part" n="1">1</label>
    </desc>
  </relation>
</listRelation>

Figure 8: HTML Visualization of Work-Person Relationship

Relationships

This work commemorates 1 person.
Mary Magdalene
(saint)
Names: Maryam Maghdlayth, Maryam Magd’layth, Marie Madeleine, Mary Magdalene
Mary Magdalene, disciple of Jesus and first witness to his resurrection.
URI: http://syriaca.org/person/1884
Example 9: Example of the <text> of a Work Record

<text>
<body>
<title xml:id="name1553-1" xml:lang="en" source="#bib1553-1" syriaca-tags="#syriaca-headword">Mary Magdalene (text)</title>
<title xml:id="name1553-2" xml:lang="syr" source="#bib1553-1" syriaca-tags="#syriaca-headword">Maryam Ma#dəlāyṯā</title>
<title xml:id="name1553-3" xml:lang="fr-x-bhs" source="#bib1553-1">Maryam Madeleine (texte)</title>
<title xml:id="name1553-4" xml:lang="fr" source="#bib1553-1">Marie Madeleine (texte)</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">#####</persName> (<persName xml:lang="fr">Éphrem</persName>) </author>
=idno type="URI">http://syriaca.org/work/1553</idno>
<! Example authority file URI -->
<! <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">third/fourth 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>## #####̈# ### ### ###### #### ##### #### ####̈# ####### #### ## ### ### #### ####
### ##### #####.</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;</quote>
</note>
</bibl>
go woman, your debts have been pardoned and all your sin is erased.

</note>

</bibl>

<bibl type="lawd:Edition" source="#bib1553-1" xml:id="bib1553-2">
<title>Des heiligen Ephraem des Syriers Sermones II</title>
<ptr target="http://syriaca.org/bibl/1815"/>
<citedRange unit="vol">2</citedRange>
<citedRange>col. 313-338</citedRange>
</bibl>

<bibl type="syriaca:Manuscript" source="#bib1553-1" xml:id="bib1553-4">
<title type="shelfmark">London, British Library, Add. 14651</title>
<ptr target="http://syriaca.org/manuscript/419"/>
<biblScope unit="folio" from="196" to="201">196-201</biblScope>
</bibl>

<bibl type="syriaca:AncientVersion" source="#bib1553-1" xml:id="bib1553-6">
<cf. title="CPG">
<ptr target="http://syriaca.org/bibl/2081"/>
</cf.>
</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>
</bibl>
<relation type="translationSource" active="#bib1553-10" ref="dct:source" passive="#bib1553-3" 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 type="part" active="http://syriaca.org/work/1553" ref="dct:isPartOf" passive="http://syriaca.org/work/10525">
<desc>
<label type="order" subtype="part" n="1">1</label>
</desc>
</relation>
</listRelation>
</bibl>
</body>
</text>


NOTES

1 For the ontology, see “The LAWD Ontology,” http://lawd.info/.
3 http://syriaca.org/documentation/isostandards.html
7 The URIs in this example and those below are unique identifiers, but as of the writing of this article some do not yet resolve to publicly viewable web pages, since their respective modules are still in development.
8 Live version at http://syriaca.org/work/1553.
Some `<bibl>`s have been removed for the sake of conciseness.


**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) (Saint-Laurent et al. 2016; see also Saint-Laurent 2016, Zanetti 2016). Past TEI-encoding practice has focused on describing specific manuscripts or creating editions of works. By contrast, the NHSL seeks to describe abstract or conceptual works (including unpublished ones) and to relate them to people, places, and other works, as well as to the manuscripts, editions, and translations that embody them. Two key features of this encoding model include using `<bibl>` for description of works and fully leveraging `@source` for scholarly citations. In preparation for expanding the NHSL to include other genres, Syriaca.org is revising the TEI schema used for hagiographic works. Among other revisions, the new model will employ RDF classes and properties as `<bibl>` `@type` and `<relation>` `@ref` values, respectively, in order to aid RDF serialization. The authors actively seek feedback, suggestions, and criticism concerning this revised schema.

**INDEX**

**Keywords:** Syriac literature, work description, reference tools, authority files, bibliographical standards, historical corpora, RDF

**AUTHORS**

**NATHAN P. GIBSON**

Nathan P. Gibson is a Postdoctoral Scholar in Syriac Studies and Digital Humanities at Vanderbilt University (Nashville, TN, USA) and a Visiting Researcher at Ludwig Maximilian University (Munich, Germany). He co-edits reference works for Syriaca.org, including the Syriac Biographical Dictionary and the New Handbook of Syriac Literature, resources which will enable scholars to better navigate the corpus of Syriac literature surviving from antiquity and the middle ages. Dr. Gibson is a historian of the Middle East trained in Syriac and Arabic. He received his Ph.D. in Semitic and Egyptian Languages and Literatures from The Catholic University of America in 2015.
DAVID A. MICHELSON

David A. Michelson is Assistant Professor of the History of Christianity at Vanderbilt University (Nashville, TN, USA) and the general editor of Syriaca.org: The Syriac Reference Portal. He is passionate about the use of digital scholarship to preserve the heritage of the minority populations in Syria and Iraq. In addition to digital publications, Prof. Michelson is the author of The Practical Christology of Philoxenos of Mabbug (Oxford, 2014) a study of the prolific Syriac author and theologian, Philoxenos of Mabbug (d. 523). Prof. Michelson earned his Ph.D. in History from Princeton University.