Computer - Assisted Processing of Intertextuality in Ancient Languages

Mark Hedges, Anna Jordanous, K. Faith Lawrence, Charlotte Roueché, Charlotte Tupman

Journal of Data Mining and Digital Humanities

Computer-Assisted Processing of Intertextuality in Ancient Languages

Mark Hedges[1*], Anna Jordanous[2], K. Faith Lawrence[1], Charlotte Roueché[1], Charlotte Tupman[3]

1 King's College London, UK

2 University of Kent, UK

3 University of Exeter, UK

*Corresponding author: Mark Hedges mark.hedges@kcl.ac.uk

Abstract

The production of digital critical editions of texts using TEI is now a widely-adopted procedure within digital humanities. The work described in this paper extends this approach to the publication of gnomologia (anthologies of wise sayings), which formed a widespread literary genre in many cultures of the medieval Mediterranean. These texts are challenging because they were rarely copied straightforwardly; rather, sayings were selected, reorganised, modified or re-attributed between manuscripts, resulting in a highly interconnected corpus for which a standard approach to digital publication is insufficient. Focusing on Greek and Arabic collections, we address this challenge using semantic web techniques to create an ecosystem of texts, relationships and annotations, and consider a new model – organic, collaborative, interconnected, and open-ended – of what constitutes an edition. This semantic web-based approach allows scholars to add their own materials and annotations to the network of information and to explore the conceptual networks that arise from these interconnected sayings.

keywords

linked data; semantic web; digital edition; manuscripts; ontology; RDF; TEI; gnomologia; anthologies

INTRODUCTION

The TEI (Text Encoding Initiative) XML format has been widely adopted as the standard encoding for marking up textual data with semantic content [Mylonas & Renear, 1999; Pierazzo, 2011; Sperberg-McQueen, 1991]. The adoption of this standard in principle facilitates interoperability between different resources, enabling them to be used in combination for new research, and this publication strategy has been embraced widely in the digital humanities community.[i] Lack of communication and failure to share research can still result in consolidation rather than expansion of information, the so-called 'digital silo' [Nichols, 2009; Zorich, 2008], but sometimes we also need to be able to do more with our texts than TEI currently allows.

We are therefore extending the TEI model through our work on editing medieval gnomologia. Our manuscripts contain collections known as gnomologia, which consist of wise sayings containing moral or social advice, or expressing philosophical ideas. These are collections of extracts from earlier works, but were rarely straightforward copies. Sayings were selected from other manuscripts, reorganised, and modified or reattributed. The genre also crossed linguistic barriers, e.g. from Greek into Arabic, again rarely in straightforward translation. These changes often reflected a change of social context, especially between different cultural traditions. In later centuries, collections were translated from Arabic into western European languages.[ii] In all languages, such collections also informed the writing of continuous texts, of

a kind that are more readily perceived as literary. The Sharing Ancient Wisdoms project (SAWS), funded by HERA from 2010 to 2013, aimed to examine such texts, publishing several gnomologia in Greek, Arabic and Latin, as well as a number of continuous texts that used gnomologia as sources.[iii] These complex traditions themselves call into question the simple concept of citation: many examples of familiar passages may come not from full copies of the original text, but though a chain of texts which may extend over centuries. Moreover, the compilation of collections, which to the modern reader appears a second order activity, required a creative process; each collection was shaped in some way, and must be taken into account if we are seeking to understand the preoccupations and concerns of particular periods. While we could not hope to present more than a fraction of this rich and abundant material, we aimed to develop tools and protocols for doing so.

In this paper we describe a methodology and framework for publishing gnomological manuscripts that addresses and exploits their high degree of connectivity, without imposing a false concept of hierarchy. The paper considers the rationale for the work in Section 2. After a survey of related research in Section 3, in Section 4 we describe our information model, including a brief description of the ontology. We describe our approach to implementation in Section 5; our evaluation and future work in Section 6; and our current conclusions in Section 7.

I GNOMOLOGIA AND DIGITAL EDITIONS

There has long been interest in the relationships within and between these manuscripts [Gutas, 1981; Richard, 1962; Rodríguez Adrados, 2001], as the analysis of these interrelations can reveal much about the dynamics of the cultures that created and used these texts. The large number of manuscripts, the complexity of their interrelationships, and the fact that a certain critical mass of material is required to carry out such research, has hindered their exploitation in the past; these very factors, however, make research in this area particularly susceptible to digital methods.

The nature of the material suggests a fundamentally different approach to creating an edition. Instead of considering variant witnesses to an 'original' text[iv] of which we are trying to create a single edited version, we have a number of interrelated texts of equal standing. The similarities between these texts need to be represented, but as these similarities take various forms, have various degrees, and operate at various levels of granularity, a more nuanced approach is required.

We are extending rather than rejecting the standard paradigm. We envisage that any particular text will be edited using TEI, and that such editions will continue to be published in digital libraries. We are, however, concerned not *only* with creating digital editions of these texts; we are building on current best practice, to publish the gnomologia in a manner that enables a better understanding of these texts as a network of information rather than as isolated documents. As long ago as 1990, DeRose and his co-authors reflected on how electronic text documents could best be structured for flexibility in use and reuse.[v] We are addressing this by better facilitating linking and comparisons, using an approach based on semantic web technologies.

The texts that we are editing form a small subset of these manuscripts, the majority of which will be edited and published by others; we thus envisage scenarios in which other groups will link their texts to ours (and to those of still other scholars). Many will identify relationships

between texts that they publish independently as semantic triples. Consequently we are creating a framework of tools and methods that will enable researchers to add texts and relationships of their own, to create a corpus whose value will increase with its size and interconnectivity. We envisage an eventual network of marked-up texts and textual excerpts, linked together to allow researchers to represent, identify and analyse the flow of knowledge and transmission of ideas through time and across cultures.

This will enable scholars to create a more detailed picture of anthological sources, and provides a clearer picture of what was read and deemed important at a particular time and place. The extensibility of our approach means that others will be able to link the remaining unidentified text passages to sources they have identified, thus building and strengthening the corpus.

This publishing model also serves to contribute to the wider debate around the somewhat ambiguous notion of citation. Traditional editions can often indicate in a broad-brush manner that one section of text is a citation of another earlier text, without a deeper discussion of the nuances of the term, or without questioning the author's access to the earlier works. This model will enable us to understand better how often Byzantine - and indeed many medieval authors - use citations from collections, and how these collections may influence the thinking of an author. Some continuous texts may in fact be driven by the shape and selection of the collection with which they worked. It also enables us to address the question of what exactly constituted a 'learned' author in medieval Byzantium. This leads to a new concept of what constitutes an edition – a corpus that is organic, collaborative, interconnected, and open-ended.

While our work has focused on gnomologia, and in particular on Greek and Arabic gnomologia from the ninth to twelfth centuries AD, the methods and tools developed are applicable to other groups of manuscripts with analogous characteristics, such as medieval mathematical, medical or scientific texts.

II RELATED WORK

A key aspect of SAWS is to represent relationships between and within collections of gnomic sayings. The RDF[vi] (Resource Description Framework) format is appropriate for this purpose, particularly when supported by an ontology of relevant information. We want to use RDF-like syntax to mark up relations between the text and links to external entities, and while RDFa (RDF with annotations) allows RDF to be encoded directly in marked-up documents, it has primarily been deployed in XHTML documents. It would be desirable to extend the scope of RDF to a wider scale [Eide et al., 2008][vii] to TEI XML documents, without extensive changes being required to the XML or to the workflow. This last point is of particular concern for the growing community of non-technical users of TEI [Pierazzo, 2011]. Keeping structural, syntactic and semantic information in the same documents also makes the markup process simpler and potentially less error-prone. To date, no method for accommodating TEI and RDF in the same document has been adopted as standard by the TEI community, though several approaches have recently been offered.

RDFTEF [Tummarello, 2005] is a Java-based tool for converting TEI files to a form which can incorporate and output RDF/XML markup. It implements a basic ontology for representing structural and syntactical elements and allows additional ontologies to be added. Queries need to be relatively complex and standard XML tools cannot be deployed within the

RDFTEF environment [Portier et al., 2012]. RDFTEF has been criticised as '[o]nly a "toy" experiment' [Portier et al., 2012] for these limitations and lack of ongoing maintenance (last source code update 2007). Also, RDFTEF introduces a new stage of work to the existing editing workflow and requires extra software.

RDFa has been used to encode RDF in a TEI document [Jewell, 2010; Lawrence 2011]. This work primarily used the OntoMedia (OM) ontology [Lawrence, 2007][viii] to describe elements within the textual narrative and to annotate the TEI XML with explicit reference to the ontological class of the typed event or entity. This typing was on an automatic basis, processing information extracted from the TEI via a conceptual mapping between the TEI and OM. A second script was used to generate RDF linked data from the extended TEI. By drawing on the ontological data held in the RDFa as well as the information in the structure and elements of the TEI, triples were created that could be cross-referenced internally and to external data resources while retaining a link back to their textual context. Specialised scripts had to be deployed to extract the RDF to add to a triple store. Deploying such scripts is non-trivial for non-technical users, in setting up the appropriate environment and in executing the scripts. The scripts used by Jewell's and Lawrence's work were also highly specific to those documents. These issues were also seen in a similar script-based approach to automated creation of RDF triples from TEI documents, in work performed by the Supporting Productive QueRies (SPQR) project [Blanke et al., 2012]. There is a more user-friendly alternative of transformations through XSLT stylesheets, incorporated into the user interface of tools like the Oxygen XML editor. Another tool is available to represent document structure(s) with RDF: the EARMARK OWL ontology [Peroni and Vitali, 2009]. This uses RDF to model structural information, but does not model the text and additional semantic information, so again structure, data and markup become separated. To reduce over-specificity and encourage re-use of our materials, the adoption of a more generic underlying model for transformations is explored in this present paper.

The inclusion of RDF in TEI documents is a current area of interest in the TEI-Ontologies Special Interest Group (SIG),[ix] which is using XSLTs to convert TEI to RDF [Ore and Eide, 2009] by relating TEI markup to vocabulary in the CIDOC-CRM cultural heritage model [Doerr, 2003]. CIDOC-CRM is the Cultural Reference Model for the museum heritage organisation CIDOC. The inclusion of FRBRoo, the Functional Requirements for Bibliographic Records (FRBR) model harmonised with CIDOC-CRM [Doerr and LeBoeuf, 2009], has also been discussed. However work in this area is progressing slowly and development has concentrated around TEI/CIDOC-CRM harmonisation, for example see [Ciula and Eide, 2014]. Some mappings have been drafted (last updated 2007/8) and stylesheets (last updated 2011) and guidelines (last updated 2010) have been published, all by the SIG, but two issues are worth noting:

- The size of the current TEI P5 tagset raises practical difficulties in providing comprehensive mapping from TEI to alternative representations. The TEI ontologies SIG has identified a subset of elements to map to CIDOC-CRM, choosing only those which represent semantically meaningful elements within the text, "such as persons, places, dates and events"[x]. This approach is practical but disregards many triples of potential interest such as document structure and metadata.
- The only direct representation of lexical material within CIDOC-CRM is through one class (E33 Linguistic Object) and its two subclasses (E34 Inscription, E35 Title). This choice of CIDOC-CRM as base model is acknowledged to be influenced by the research interests of the SIG members in cultural heritage and museum

documentation[xi]. For our interests in structural information and metadata, the Dublin Core (DC) model seems a more natural choice and is highly developed and widely adopted. A mapping from TEI to DC has been tackled in stylesheets created by the SIG but does not appear in reports of their main approach.

The aim of SAWS is to represent semantic relationships between passages of text identified during the editorial process; it was consequently desirable to mark these up directly in TEI. As TEI did not provide a standard mechanism for encoding all such relationships, we proposed the adoption of the **<relation>** element, which was subsequently accepted by the TEI, for encoding RDF relations in a TEI document[xii], representing the Subject-Predicate-Object triple format of RDF through the following attributes: **@active**, **@ref** and **@passive**. We describe and explain our encoding of RDF relations using **<relation>** in Section 5.2.

III CONCEPTUAL MODEL

In this section we describe our conceptual model, addressing:

- The base ontology and the motivations for its selection.
- The extensions added to form the SAWS ontology.

3.1 Base Ontology

Our ontology reuses the FRBRoo ontology [Doerr & LeBoeuf, 2007], a combination of the CIDOC-CRM and FRBR ontologies. The CIDOC Conceptual Reference Model (CRM)[xiii] is an ontology of the information and relationships relevant for cultural heritage documentation [Doerr, 2003]. CIDOC-CRM is a common vocabulary (ISO 21127) for discussing information on cultural heritage and mapping it to a digital equivalent representation [Binding et al., 2008; Doerr, 2003; Eide et al., 2008; Eide & Ore, 2007; Sinclair et al., 2006; Varnienè-Janssen & Juskys, 2011]. The Functional Requirements for Bibliographic Records model (FRBR) was devised as an entity-relationship model of bibliographic data and publications [Madison, 2000; Tillett, 2004]. It documents and distinguishes between the concepts forming the basis of a *Work*; the *Expression* of such *Works* in a fixed but abstract form; the *Manifestation* of such *Expressions* in physical form; and single *Items* that are exemplars of such *Manifestations*.

The CIDOC and FRBR ontologies were originally developed independently. Recognising the potential of combining these ontologies, the communities collaboratively produced FRBRoo [Doerr & LeBoeuf, 2007]. FRBRoo is the FRBR ontology expressed in an object-oriented form more compatible with that of the CIDOC-CRM, extending the CIDOC-CRM with the FRBR vocabulary. Given the relevance of CIDOC-CRM and FRBR, particularly in the repeated transmission of ideas expressed in written works, FRBRoo was the most appropriate ontology on which to base our vocabulary, suitably extended to support the description of, and the recording of relationships between, our physical and information objects. In particular, FRBRoo clarifies how the CIDOC *LinguisticObject* class and the FRBR *Expression*/*Manifestation* classes relate to each other, allowing greater clarity in representing our relationships.

Other relevant ontologies were considered:

- An extension of the CIDOC-CRM, CRM_{dig}, has been proposed for documenting digital objects [Doerr & Theodoridou, 2011]. While it makes significant enhancements

to the CIDOC-CRM for dealing with *digital* documents, the standard form of CIDOC-CRM was more relevant for our purposes.

- We considered other ontologies documenting bibliographic resources[xiv], as well as an ontology for documenting scholarly works[xv], but they lacked sufficient depth for describing their content. The SPAR suite of ontologies for Semantic Publishing And Referencing[xvi] offer more depth and scope for detailed referencing, but the term "manuscript" in SPAR's FRBR-aligned Bibliographic Ontology (FaBiO) is explicitly used to refer to a textual work that is not 'a handwritten historical document on paper or parchment'[xvii], which is exactly the type of manuscript that we need to model.
- The OntoMedia [Lawrence, 2007] and Stories[xviii] ontologies focus on the content of a text at the expense of information about the document itself.
- SKOS could be used to represent the hierarchical structure of information content, and Dublin Core metadata provides a vocabulary for describing information about the manuscript[xix].

Each of these is relevant in part to our data; however, rather than using several ontologies representing different aspects, it was decided to adopt FRBRoo as the base ontology, borrowing terms from relevant ontologies as and where necessary, as FRBR-oo represented most aspects of the manuscript information.

Both the CIDOC-CRM model and FRBRoo have been implemented as OWL ontologies[xx] by the University of Erlangen, Germany. We use and extend these OWL implementations as extensions for the SAWS vocabulary.

3.2 The SAWS ontology as an extension of FRBRoo

3.2.1 Items of Interest

We use the term *Manuscript* to refer to the physical objects in which our texts are contained. Typically, a manuscript will contain more than just the collections of wise sayings; conversely, a collection of sayings may span several manuscripts. Consequently, in our model the fundamental unit is the *CollectionInstance*, which is an extension to the FRBRoo ontology as a combination of a *LinguisticObject* (CIDOC) and *Expression* (FRBR), corresponding to the physical instantiation of a collection of sayings in one or more manuscripts.

The other fundamental object is the *ContentItem*, which corresponds to the individual sections of interest - i.e. a saying - within a *CollectionInstance*. These may be simple assertions:

> 'One cannot cover a fire with a cloak nor a shameful sin with time.'

The gnomologia also contain longer anecdotal sections:

> 'Diogenes was asked by someone why people give to beggars but not at all to philosophers, and he said, "Because, perhaps, they expect to become lame or blind but not to become philosophers."'

Here there are two components of interest: the statement itself ('Because, perhaps, they expect ... philosophers'), and a narrative text ('Diogenes was asked...'). Consequently we introduced two corresponding objects, *statement* and *narrative*.

3.2.2 Relationships

The definition of a vocabulary of relationships has been a key component of the research. Relationships may occur:

- within a single *CollectionInstance*
- between *CollectionInstances*
- between a *CollectionInstance* and a 'source text' (e.g. a Greek classical work or the Bible)
- between a *CollectionInstance* and subsequent texts that drew upon it.

We need a vocabulary that is not only capable of representing relationships among a specific set of texts, but is sufficiently flexible to be extended or refined to cover relationships in analogous materials. This vocabulary has been developed through collaboration between information scientists and scholars within digital humanities and manuscript studies, and is published at the permanent URL http://purl.org/saws/ontology.

Relationships that were identified include:

Manuscript *isWrittenAt* Scriptorium
Manuscript *isInLanguage* Language
CollectionInstance *isWrittenBy* Scribe
CollectionInstance *isTranslationOf* CollectionInstance
Section *isSequentiallySimilarTo* Section[xxi]
ContentItem *isShorterVersionOf* ContentItem
ContentItem *isVerbatimOf* ContentItem

3.3 Example of the conceptual model in application

The following examples - translated into English for clarity[xxii] - illustrate how sayings develop. Item 1 is a saying attributed to Alexander the Great in a medieval Greek gnomological text, the 'Gnomologium Vaticanum'; Item 2 is an extract from Plutarch's 'Life of Alexander' (8.4.1), identified as a potential source of the saying. The text is not a direct quotation, but has been paraphrased.

1. Alexander, asked whom he loved more, Philip or Aristotle, said: 'Both equally, for one gave me the gift of life, the other taught me to live the virtuous life.'
2. Alexander admired Aristotle at the start and loved him no less, as he himself said, than his own father, since he had life through his father but the virtuous life through Aristotle.

The second example is a narrative only.

Item 3, below, is an extract from an Arabic anthology, and is attributed to Pythagoras ['Selections from the Sayings of the Four Philosophers: (B) Pythagoras' saying 18 (ed. Gutas)]. Here the source text seems to be Diogenes Laertius's 'Life of Aristotle' (5.19), shown as Item 4, although not only has the saying been translated from Greek into Arabic, it has become more pithy in translation, and the saying has been re-attributed from Aristotle to Pythagoras. Several relationship assertions may need to be used to represent the connection between the two sayings.

3. He said: “Fathers are the cause of life, but philosophers are the cause of the good life.”
4. Aristotle said that educators are more to be honored than mere begetters, for the latter offer life but the former offer the good life.

IV IMPLEMENTATION

The implementation approach has three main aspects:

- The encoding and publication of a digital archive of editions of a selected number of these texts;
- The identification and display of the links between the anthologies, their source texts, and their recipient texts;
- The building of tools to allow scholars outside the SAWS projects to link their texts to ours.

4.1 Encoding Individual Texts as TEI Documents

Each text is marked up in TEI XML schema developed at King’s College London for the encoding of gnomologia, which is based on the TEI Manuscript schema. The structural markup reflects as closely as possible the way in which the scribe laid out the manuscript. We use the **<seg>** element[xxiii] to mark up base units of intellectual interest, such as the *statement* and its *narrative*. These base units need not have been identified as units by the scribe, but are the result of an editorial decision. Consider the case described in 4 above:

> ‘Alexander, asked whom he loved more, Philip or Aristotle, said: “Both equally, for one gave me the gift of life, the other taught me to live the virtuous life.”’

```
<seg type="contentItem">
        <seg type="narrative">
    Alexander, asked whom he loved more,
    Philip or Aristotle, said:
        </seg>
        <seg type="statement">
    Both equally, for one gave me the gift of
    life, the other taught me to live the
    virtuous life.
  </seg>
</seg>
```

Each of these **<seg>** elements was given an **@xml:id** to provide a unique identifier that differentiates them from all other examples of **<seg>**:

```
<seg type="statement" xml:id="AppGnomVat001s2">
```

This provided the means to refer to a specific section of the text, and these internal identifiers were then used to generate URIs.

4.2 Encoding Using RDF within TEI <relation>

Our schema allows editors to publish the texts in accordance with TEI for Manuscripts while also supporting the identification and description of the relationships between individual units

of interest. Because our data model extends the FRBRoo ontology to model the classes and relationships, a more flexible approach to encoding relationships within the TEI document is required, one which can cope with the generic and extensible nature of the ontology.

RDF triples are encoded within the TEI documents to include ontological information that is not present in the TEI markup itself, using the TEI element **<relation>**:[xxiv]

> **@ref** states the relationship type (from the list of relationships in the ontology);
> **@active** points to the URI of the resource that is being linked from;
> **@passive** points to the URI of the resource being linked to;
> **@resp** is used to identify the individual (or bibliographic source) responsible for asserting the relationship.

The TEI Guidelines now include the SAWS usage of the **<relation>** element as one of their examples:

```
<relation
  resp="http://viaf.org/viaf/44335536/"
  ref="http://purl.org/saws/ontology#isVariantOf"
  active="http://www.ancientwisdoms.ac.uk/cts/urn:cts:greekLit:tlg3017.Syno298.sawsGrc01:divedition.divsection1.o14.a107"
  passive="http://data.perseus.org/citations/urn:cts:greekLit:tlg0031.tlg002.perseus-grc1:9.35"/>
                                                                    Show all bibliography
```

> **<relation**
> **resp="**http://viaf.org/viaf/44335536/**"**
> **ref="**http://purl.org/saws/ontology#isVariantOf**"**
> **active="**http://www.ancientwisdoms.ac.uk/cts/urn:cts:greekLit:tlg3017.Syno298.sawsGrc01:divedition.divsection1.o14.a107**"**
> **passive="**http://data.perseus.org/citations/urn:cts:greekLit:tlg0031.tlg002.perseus-grc1:9.35**"**
> **/>**

In other words, this example records an assertion that there is a relationship "isVariantOf", as defined by the SAWS ontology, between a passage of text in the SAWS corpus and a passage of text in the Perseus Digital Library. It also gives details of the assertion having been made by an individual who has an entry in VIAF (Virtual International Authority File), in this case Charlotte Roueché. Each value is a resolvable URI. The use of all four attributes is required in the SAWS usage of **<relation>**, as the project has been particularly concerned to enable users to trace responsibility for these assertions either to a specific person or to a bibliographic reference. It also serves as a means of ensuring that individual credit for making the assertions can be given.

If we want to express uncertainty about whether or not a relationship actually exists, we can use **<certainty match="../" locus="name" cert="low"/>** as a child of the **<relation>** element (where ".." points to the parent element). To express uncertainty about an attribute within **<relation>**, for instance about the type of relationship that exists, we can point to a specific attribute, e.g.: **<certainty match="../@ref" locus="value" cert="low"/>**.

4.3 Publishing Digital Editions

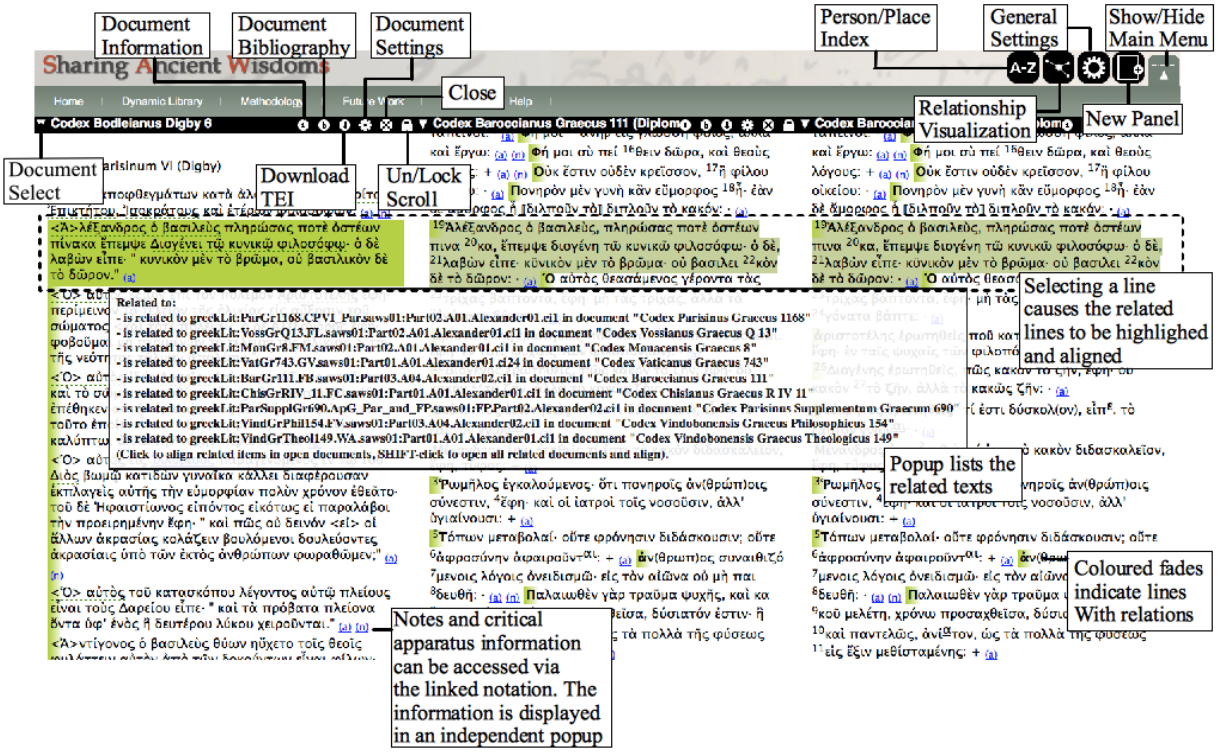
The creation and publication of a digital edition represents one of the main goals of this project. In its design it was important that we not only present the texts and related

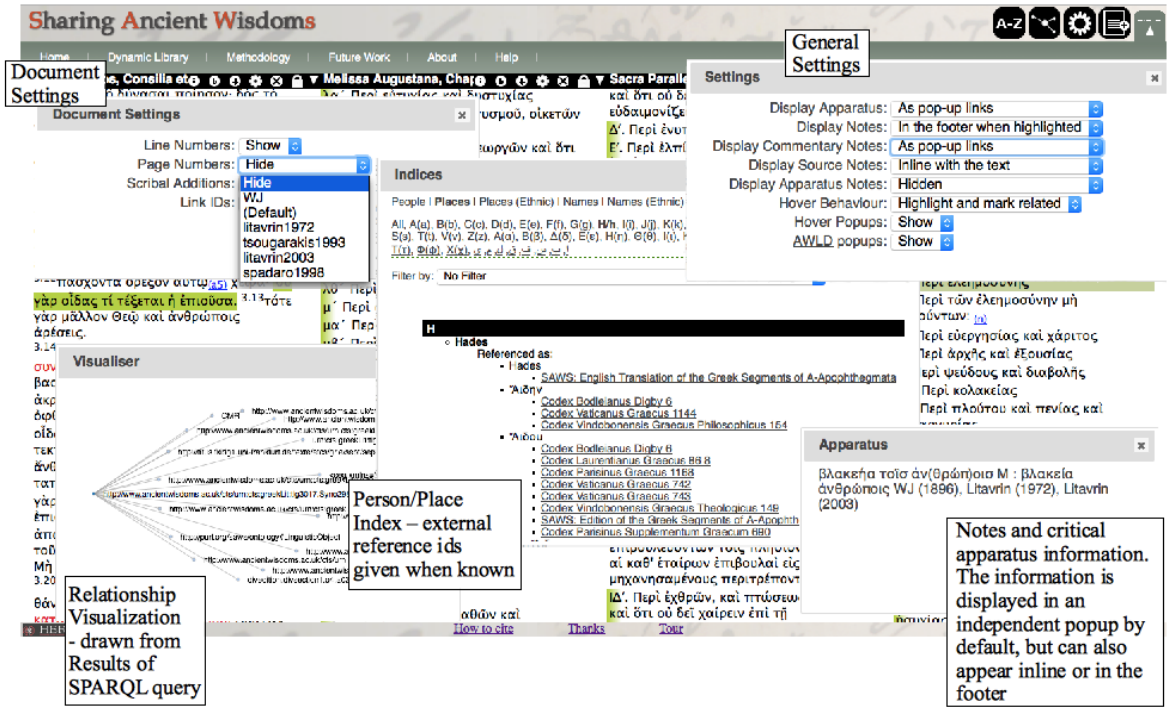
commentaries, but enable the user to visualise and explore the data. In doing so we provide additional contextual information, alternate navigation options and multiple display options.

As comparisons between the texts are vital to the project, the main priority for the text display was to allow multiple texts in parallel. Initially using a modified version of the Versioning Machine (VM), developed at the University of Maryland, we were forced by its limitations dealing with long documents to look for alternatives. A number of parallel text displays developed for Biblical texts were reviewed but, while they showed promise, they were fundamentally linked to the unique structure of Biblical books and therefore not suitable for our purposes.

The Folioscope parallel text view is a lightweight front end integrating both the text drawn from the TEI and the relationship data provided via the sesame triplestore. The traditional text display is provided via the Kiln framework. Developed at the Department of Digital Humanities at King's College London, Kiln uses XSLT stylesheets to create dynamic websites from XML-encoded sources. This allows for a traditional digital edition with each source and commentary presented to the reader supported by search, indexing and contents pages. At the document (panel) level, the user is able to set display options for line and page numbers, scribal additions and identifier display while the global settings offer multiple display options for notes, commentary, popups (for example, those produced by AWLD for Pelagios integration), and hover behavior.

Text integration is provided through direct querying of of the triplestore, using display cues in combination with tooltips to indicate lines with known relations and display the related information. Different commands align related lines in open panels or open all the related documents, documents in the Folioscope are opening in new panels and aligned while external documents are opened in separate tabs of the browser. This allows the system to support the exploration of the texts within the library as well as those available outside it.





4.4 Linked Data

SAWS scholars also need to record links to external sources, particularly collections of Linked Data on the ancient world. To date we link to collections of ancient data through Pelagios references to the Pleiades historical gazetteer [Barker et al., 2012], and to documents in the Perseus Digital Library. We also link to information on people mentioned in a selection of the texts, through the Prosopography of the Byzantine World resource, and other

resources[xxv]. We plan to extend this through participation in the Standards for Networking Ancient Prosopographies (SNAP) initiative[xxvi].

The ability to traverse links between sets of data and discover related information serendipitously is one of the major benefits of adopting linked data for this project. It is a key part of the academic research underpinning this project and is further justified elsewhere [Solomon, 1993]. This is particularly useful where potential sources are geographically scattered, difficult to access or not widely known. As an example, the Perseus Digital Library holds a collection of Classics-related documents, which collectively contain over 68 million words, as well as an Arabic collection containing over 5 million words. Navigating such quantities of potential research material is one of the challenges faced by Classics researchers. Digitisation and cataloguing of the sources through projects like Perseus is an important step in facilitating this research, and is being enhanced further by semantic navigation such as that undertaken in by SAWS.

For example, information in the Pleiades historical gazetteer can be consulted when constructing queries. Researchers can see all texts that refer to that particular geographical location, even if different place names are used over time. This is possible because the Pleiades ontology gives a precise geographical reference for each place[xxvii]. The use of RDF and linking therefore allows us to transcend time and language boundaries to some extent[xxviii].

V EVALUATION

5.1 Overall Evaluation of the Project

Throughout the life of the project, we tested our approach through invited workshops and external presentations; this enabled regular interchange with our peers. For example, we demonstrated the enhancements possible with the RDF information in a workshop in June 2012[xxix]. This highlighted several benefits, in particular the ways in which the manuscripts could be navigated. Also highlighted were ways in which the SAWS editing process could be refined, such as the editing of right-to-left directional languages such as Arabic.

This demonstration prompted useful constructive feedback, leading to the identification of further relationship types. It also prompted scholarly debates following the identification of different interpretations of the notion of translation (generated by the requirement to formalise collaboratively their tacit knowledge). Ongoing consultation with manuscript scholars provided formative evaluative feedback for further developments. These included technical collaborations with the Islandora team in Prince Edward Island, Canada, with whom we developed a TEI to RDF mapping for automatic extraction of the RDF triples inherent in the TEI markup and the triples encoded in **<relation>** elements [Jordanous, Stanley & Tupman 2012, Tupman, Jordanous & Stanley 2013]. This mapping was also deployed in the Islandora Critical Editions Solution Pack, a repository-based software tool for managing digital editions produced by the Editing Modernism in Canada (EMIC) project in conjunction with the Canadian company Discovery Garden. The work was implemented using XSLT, forming a basis for further mappings and transformations. In particular, we explored how the Dublin Core metadata model and the FRBR-oo ontology can be used to enhance the TEI to RDF mappings, for a semantically rich vocabulary.

Overall, the project successfully achieved:

- Digital edition of manuscripts published using TEI and RDF annotations.
- Manuscripts to be navigable through structural and semantic links.
- Semantic content in manuscripts to be searchable and queryable through extraction of RDF information.
- Positive impact within the philological community, particularly those researching medieval manuscripts

5.2 Evaluation of the SAWS Ontology

To evaluate the SAWS ontology, we wish to evaluate both the validity of the ontology and the quality of the ontology for expressing what we want it to.

The validity (logical consistency) of the resulting SAWS ontology was checked using reasoners built in to Protege, the tool used to develop the ontology. Reasoners check that the logical statements within the ontology are consistent, with no contradictions/illogical statements. Application of the reasoners within Protege have highlighted no errors or contradictions in the SAWS ontology; in other words, the ontology has been shown to be valid.

Brank et al. (2005) outline different ways to evaluate how useful and representative an ontology is; we have adopted a number of these to evaluate the quality of the SAWS ontology. One evaluative method identified by Brank et al. (2005) is to compare the ontology to an existing 'standard' of some kind; in our case we identified the FRBRoo ontology (based around CIDOC-CRM) as a useful standard for comparison when recording information about documents using Semantic Web representations. The original ontological requirements from the collaboration between domain experts and technical observers was successfully mapped to the existing FRBRoo ontology, which has undergone extensive review from both the CIDOC and FRBR community as well as users of FRBRoo. After this mapping process, the resulting ontology was presented to the domain experts in the form of a vocabulary they could use to express relationships between and within the manuscripts they studied. Our mapping work has been recognised as a CIDOC-CRM-compliant extension [Alexiev et al., 2013, p. iii].

Formative feedback solicited from the domain experts at various stages was used to refine the ontology further, for another stage of evaluation performed by people (usually domain experts) who 'try to assess how well the ontology meets a set of predefined criteria, standards, requirements, etc.' [Brank et al., 2005, p.1]. This occurred on an ongoing basis with domain experts within the SAWS team, including a number of meetings specifically dedicated to evaluating the ontology as it was at that time. We also conducted a number of workshops with domain experts outside the SAWS team. Typically, in the early stages of the project the ontology underwent significant revision during these evaluation sessions. Towards the end of the project, revisions from internal and external experts were significantly reduced. It is difficult to claim that any ontology can reach a stage of being finished and complete; knowledge develops quickly and terms within a discipline become used in different ways over time. Certainly with the SAWS ontology there are still a small number of terms which are not universally agreed upon by all experts; however some discrepancies were to be expected - we are not dealing with objective facts, but with scholarly opinions, interpretations and assertions which are open to debate. For example, as we mention elsewhere in this section, evaluation of the ontology assisted domain experts in identifying different interpretations of translation of which they had previously been unaware. Overall, from evaluations towards the end of the

project, we are confident that the SAWS ontology has reached a stable state, receiving the approval of many domain experts.

One interesting observation from external domain experts was the ontology was too focussed towards medieval anthologies of wise sayings. The aim of the SAWS ontology was in fact to represent this type of text; however our use of the FRBRoo as the underlying ontology helps provide ways to deal with other types of texts. This issue could then be treated as a way of improving the presentation of the ontology (by including more of the underlying FRBRoo ontology) rather than improving the ontology. This would make the SAWS ontology more widely applicable in strongly related domains, without losing focus on the original scope of the ontology.

The evaluation of the overall project contributes to another method identified by Brank et al. (2005) to evaluate the quality of an ontology: to use the ontology within an application, and to evaluate the results. In the next section we highlight and focus on this particular type of evaluation, which we see as the critical part of assessing this work.

5.3 Evaluation of the Digital Approach of SAWS

As noted above, one way of evaluating the success of an ontology-based approach to representing information is to evaluate how the ontology is used in an application, to assess the usefulness and accuracy of the results. We adopted this approach as part of the ongoing evaluation throughout the lifetime of the SAWS project, as a feedback mechanism to continually develop the ontological work [for example, see the mid-project comments on evaluation in Jordanous et al., 2012]. As reported above, the ontology was refined at various stages throughout the project. Critically, the resulting ontology is successful to the degree that it allows domain experts to record their tacit knowledge and expertise in digital form. Through evaluation of the use of the SAWS ontology in application, overall feedback on the usefulness of our digital approach was identified:

- Through marking up the manuscripts in TEI XML we have made these collections of wise sayings available in digital form with structured content, removing the accessibility problems to the original physical manuscripts. The text of the manuscripts has been supplemented with expert knowledge, much as would happen when producing a critical edition.
- The mark-up process has been undertaken both by experts in this area and non-experts supervised by experts and given brief training. Especially for larger-scale markup projects, the markup process can be time-consuming and it is useful to be able to share this workload without needing to recruit several people with detailed expert knowledge. The process of tagging the digital versions of the collections is modular and can be performed in a distributed way, across a number of people, with the experts being able to add more detail from their specialist knowledge whilst sharing the more repetitive markup with others.
- The markup provided through the SAWS TEI schema caters for different stages of annotation: from quick annotations ‘out in the field', when researchers are actually at the physical location where the manuscript is kept; initial editing of structure and brief observations; detailed analysis; through to the publication of a critical edition of the manuscript. This models the analytical processes and stages that such researchers are already familiar with in their work.

- There is a growing desire to make more of the XML documents by including relationship information within the TEI markup itself, to which the SAWS approach has made a contribution. Information on how documents are related and how links exist within documents is extracted from the analysis to be included in the TEI editions of the manuscripts. This enhances the semantic content of these electronic versions. Highlighting the contribution of the SAWS project to these developments: example markup from SAWS TEI documents has been included in the official TEI guidelines documentation, acting as an example of best practice to follow in incorporating RDF within TEI.
- Further evaluation of the SAWS ontology and the overall approach will be seen over time, through the use of the SAWS ontology in other digital humanities applications. To date, we have collaborated with the *Corpus der arabischen und syrischen Gnomologien* (CASG) project, Halle, Germany to help them adopt a SAWS approach (see xxx, paper submitted to this special issue by CASG). We have also worked with the *Monastic Paideia* (MOPAI) project in Lund, Sweden, to advise on how they can use SAWS [Johnsson & Åhlfeldt, 2015]. From citations we see that SAWS has directly inspired: the approach taken by the BIBLIMOS project for publishing ancient scientific Mauritanian manuscripts [Markhoff et al., 2015]; the use of the FRBRoo ontology as a base for the OMOS (Ontologie sur des Manuscrits Ouest Sahariens) ontology for Western Saharan Manuscripts [Diakite and Markhoff, 2015]; an ontology representing data about proverbs [Zhitomirsky-Geffet et al., 2015]; and a semantic study of Dante Alighieri's philosophical essay 'Convivio' [Bartalesi et al., 2013]. SAWS has also influenced the use of RDF within TEI markup in the BIA-NET project, for archiving Ancient Roman Law texts [Spampinato & Zangara, 2013] and an ontology for Sumerian literary narratives (which also extends FRBRoo) [Nurmikko-Fuller, 2015].

VI FUTURE WORK

The SAWS project has received many indications of interest from the philological community. On a longer-term basis, the success of the SAWS approach will be demonstrated by the future and ongoing adoption of a SAWS-style approach by others across this community, for editing, annotation and publishing of digital manuscript editions. A particular marker for success will be the linking to and from SAWS manuscripts by scholars outside of the SAWS research team, particularly if the SAWS digital editions become the canonical reference point for the manuscripts digitised during the project. Another indicator is the possibility of the SAWS approach being adopted by researchers outside the immediate target audience of manuscript scholars, for example those studying modern texts or other objects represented in TEI, for example the MEI (Music Encoding Initiative) community [Roland, 2002].

The SAWS approach allows us to extract triples from the marked up TEI documents, to be stored in a triple store and queried with SPARQL. With the data in a queryable form, this opens up a whole host of exciting possibilities. The primary aim is to enable the creation of digital analysis and information extraction tools for the immediate target audience (digital humanities researchers), to collect information. Outside the immediate audience, data on wise sayings and how they have evolved over transcription and transmission would also be of interest to linguists, social scientists and historians. The collections of sayings could also be exploited for potential 'pop'-applications outside the academic sphere of interest, such as online or mobile apps to generate wise sayings in appropriate contexts.

In SAWS we have created a framework for others to use and extend; a growing network of interconnected information. As the body of material of interest in this field is potentially very large, we do not view the project as creating just a digital, online edition, although this will be one result of the project, but rather as creating the kernel for a much larger corpus of interrelated digital editions. We envisage this as a SAWS `hub' for enabling related projects to annotate and link their own texts. The research value of such a corpus would be much greater than the sum of its parts, and would increase dramatically once a `critical mass' was reached.

Many of the subsequent contributions to this corpus will, of course, be carried out by other researchers; as described above, we worked closely and creatively with, the Monastic Paideia project in Lund, Sweden, and the CASG project in Halle Germany (for which see the paper by the Halle team in this journal special issue), training researchers from both projects. If such undertakings are to be able to interoperate and contribute to the wider corpus, rather than existing as a collection of separate editions (which would be of much less value to researchers), it is important that everyone 'speaks the same language' regarding how this material is represented in digital form, both semantically and technically; herein lies a significant part of our long-term provision for future scholarship. Moreover, these contributions are likely to be made over a long duration, certainly long in relation to the speed of technical developments, so our approach must be such as to allow migration, without loss of information, as the technological environment changes; our adoption of current standards and reusable ontologies assists this aim.

SAWS has also inspired development outside of philological research. The integration of externally linked information popups showed a clear gap in the (digital) publication. While classical geography is well supported by the Pelagios project, the initial plan to extend the AWLD pop-up library to other sites and provide complementary information about persons mentioned in the text was limited by the lack of suitable sources to which we could link. The lack of a single resource for unique canonical identifiers, such as that which Pleiades offers for places, for classical persons is now being addressed by the SNAP:DRGN project as a direct result of the SAWS initiative.

The development of user-facing tools is another area in which it will be essential to work closely with the scholars who will (or who may) use the tools. We cannot assume that the users will be au fait with the technology, neither can we assume that all scholars will have access to specialists in this area, so the tools must be usable with the help only of standard on-line help and documentation; releasing the tools for reuse by other scholars is important. Further work is required in the community to ensure that the tools are generalised and enabled to deal with a wide variety of textual structures.[xxx]

VII CONCLUSIONS

The work described in this paper allows us to exploit semantic web technologies for a better understanding of the highly interconnected medieval manuscripts known as gnomologia (collections of 'wise sayings'), to obtain a greater understanding of the cultural dynamics of the medieval Mediterranean world.

We achieved this through:

- The publication of the texts as TEI documents, with embedded RDF to record relationships identified by the editors;

- A framework allowing the identification of sections of intellectual interest within texts, relationships between texts, and the recording of these as annotations;
- Links internally within/between documents and externally to relevant sources of linked data outside our collections;
- A methodology that can be used by other scholars to analyse and publish analogous material.

We are thus publishing not only digital editions, but also relationships and semantic annotations within and between those texts, creating a network of relationships and providing a framework upon which others can build. Ultimately we will produce a network of digital editions of these manuscripts, enhanced by a network of semantic annotations and relationships.

We advocate a methodology for making these manuscripts accessible in a way not previously possible, together with tools to support the researcher in studying the collections. By using these textual relationships to analyse the flow of knowledge between texts and cultures, SAWS will enable a better understanding of the processes of cultural exchange between civilisations, and in particular of the cultural dynamics across the centuries of Greek and Arabic thought.

VII ACKNOWLEDGEMENTS

The SAWS team comprised sub-teams from King's College London, UK; Newman Institute, Uppsala; and the University of Vienna. The project was funded by HERA (Humanities in the European Research Area).

References

Alexiev, V., Ivanov, V., Grinberg, M. (2013). 'Foreword'. In *Proceedings of CRMEX 2013 Practical Experiences with CIDOC CRM and its Extensions,* Valetta, Malta.

Barker, E., Simon, R., Isaksen, L. (2012). 'Pelagios: An Information superhighway for the digital world.' In *Proceedings of Digital Humanities* (DH2012), Hamburg, Germany.

Bartalesi, V., Locuratolo, E., Versienti, L., Meghini, C. (2013). 'A preliminary study on the semantic representation of the notes to Dante Alighieri's Convivio.' In *DH-CASE '13 Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities*, Florence, Italy.

Binding, C., May, K., Tudhope, D., (2008). 'Semantic Interoperability in Archaeological Datasets: Data Mapping and Extraction via the CIDOC CRM'. In *Research and Advanced Technology for Digital Libraries*, vol. **5173** of *Lecture Notes in Computer Science*, pp. 280–290. Springer.

Blanke, T., Bodard, G., Bryant, M., Dunn, S., Hedges, M., Jackson, M., Scott, D. (2012). 'Linked data for humanities research — The SPQR experiment', *2012 6th IEEE International Conference on Digital Ecosystems Technologies (DEST),* Campione d'Italia, Italy, 2012

Brank, J., Grobelnik, M., Mladenić, D. (2012). 'A survey of ontology evaluation techniques.' *Proceedings of SiKKD Conference on Data Mining and Data Warehouses*, Ljubljana, Slovenia.

Caxton, W. (1877). *The Dictes and Wise Sayings of the Philosophers*. (originally published London, 1477), reprinted 1877 (Elliot Stock, London).

Ciula, A., Eide, Ø. (2014). 'Reflections on cultural heritage and digital humanities: modelling in practice and theory'. In *Proceedings of the First International Conference on Digital Access to Textual Cultural Heritage* (pp. 35-41). ACM. doi:10.1145/2595188.2595207

DeRose, S. J., Durand, D.G., Mylonas, E., Renear, A.H. (1990). 'What is Text, Really? '. *Journal of Computing in Higher Education* **I**(2):3–26.

Diakite, M.L. & Markhoff, B.B. (2015). 'OMOS: Ontology for Western Saharan Manuscripts'. *Research Report 313,* Université François Rabelais Tours - Laboratoire d'Informatique Li (Upres eA n° 6300)

Doerr, M. (2003). 'The CIDOC CRM - an Ontological Approach to Semantic Interoperability of Metadata'. *AI Magazine* **24** (3):75–92.

Doerr, M. & LeBoeuf, P. (2007). 'Modelling Intellectual Processes: The FRBR - CRM Harmonization'. In C. Thanos, F. Borri, & L. Candela (eds.), *Digital Libraries: Research and Development*, vol. **4877** of *Lecture Notes in Computer Science*, pp. 114–123. Springer, Berlin / Heidelberg.

Doerr, M. & Theodoridou, M. (2011). 'CRM$_{dig}$: A generic digital provenance model for scientific observation'. In *Proceedings of TaPP'11: 3rd USENIX Workshop on the Theory and Practice of Provenance*, Heraklion, Greece.
Eide, Ø., Felicetti, A., Ore, C.-E., D'Andrea, A., Holmen, J. (2008). 'Encoding Cultural Heritage Information for the Semantic Web'. In *EPOCH Conference on Open Digital Cultural Heritage Systems*, Rome, Italy.
Eide, Ø. & Ore, C.-E. S. (2007). 'From TEI to a CIDOC-CRM conforming model: Towards a better integration between text collections and other sources of cultural historical documentation'. In *Digital Humanities (Poster contribution)*, Urbana-Champaign, Illinois.
Gutas, D. (1981). 'Classical Arabic Wisdom Literature: Nature and Scope'. *Journal of the American Oriental Society* **101** (1):49–86.
Jewell, M. O. (2010). 'Semantic Screenplays: Preparing TEI for Linked Data'. In *Proceedings of Digital Humanities*, London, UK.
Johnsson, M. & Åhlfeldt, J. (2015). *Research Libraries and Research Data Management within the Humanities and Social Sciences*. Report, Lund University. Available at http://lup.lub.lu.se/record/5050462.
Jordanous, A., Lawrence, K. F., Hedges, M., Tupman, C. (2012). 'Exploring Manuscripts: Sharing Ancient Wisdoms across the Semantic Web'. In *Proceedings of WIMS '12: International Conference on Web Intelligence, Mining and Semantics*, Craiova, Romania.
Jordanous, A., Stanley, A., Tupman, C. (2012) 'Contemporary transformation of ancient documents for recording and retrieving maximum information: when one form of markup is not enough.' In *Proceedings of Balisage: The Markup Conference* (vol 8), Montreal, Canada.
Lawrence, K. F. (2007). *The Web of Community Trust - Amateur Fiction Online: A Case Study in Community Focused Design for the Semantic Web. Part IV*. Ph.D. thesis, University of Southampton, UK. Available at http://eprints.ecs.soton.ac.uk/14704/.
Lawrence, K. F. (2011). 'Wherefore Art Thou? - Crowdsourcing Linked Data from Shakespeare to Dr Who'. In *Proceedings of Web Science*, Koblenz, Germany.
Madison, M. A. (2000). 'The IFLA Functional Requirements for Bibliographic Records: International Standards for Universal Bibliographic Control'. *Library Resources & Technical Services* **44**(3):153–159.
Markhoff, B.B., Caratini, S., Coreale, F., Diakite, M.L., Ghamnia, A. (2015). 'Semantic Web for BIBLIMOS'. Position paper in *First International Workshop on Semantic Web for Scientific Heritage (SW4SH-15)*.
Mylonas, E. & Renear, A. (1999). 'The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore - But a New Research Community'. *Computers and the Humanities* **33**(1):1–9.
Nichols, S. (2009). 'Time to Change Our Thinking: Dismantling the Silo Model of Digital Scholarship'. *Ariadne* **58**.
Nurmikko-Fuller, T. (2015). *Telling Ancient Tales to Modern Machines: Ontological Representation of Sumerian Literary Narratives.* PhD Thesis, University of Southampton.
Ore, C.-E., Eide, Ø. (2009). TEI and cultural heritage ontologies: Exchange of information? Literary and Linguistic Computing **24**(2): 161-172.
Peroni, S., Vitali, F. (2009) 'Annotations with EARMARK for arbitrary, overlapping and out-of order markup.' In *Proceedings of the 9th ACM symposium on Document engineering*. Munich, Germany. 171-180
Pierazzo, E. (2011). 'A rationale of digital documentary editions'. *Literary and Linguistic Computing* **26**(4):463–477.
Portier, P., Chatti, N., Calabretto, S., Egyed-Zsigmond, E., Pinon, J. (2012). 'Modeling, encoding and querying multi-structured documents'. *Information Processing & Management* 48(5): 931-955.
Richard, M. (1962). 'Florilèges grecs'. In *Dictionnaire de Spiritualité V*. cols. 475-512.
Rodríguez Adrados, F. (2001). *Greek wisdom literature and the Middle Ages: the lost Greek models and Their Arabic and Castilian Translations*, pp. 91–97. English translation by Joyce Greer (2009).
Roland, P. (2002). 'The Music Encoding Initiative (MEI)'. In *Proceedings of the First International Conference on Musical Applications Using XML*, pp. 55–59.
Sinclair, P., Addis, M., Choi, F., Doerr, M., Lewis, P., Martinez, K. (2006). 'The use of CRM Core in Multimedia Annotation'. In *Proceedings of First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006), part of the 15th World Wide Web Conference*, Edinburgh, Scotland.
Solomon, J. (ed.) (1993) *Accessing antiquity: The computerization of classical studies*. Solomon, J. (ed). Tucson: University of Arizona Press.
Spampinato, D., Zangara, I. (2013) 'Classical antiquity and semantic content management on linked open data.' In *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: metadata, vocabularies and techniques in the Digital Humanities*, Florence, Italy.
Sperberg-McQueen, C. M. (1991). 'Text in the Electronic Age: Texual Study and Textual Study and Text Encoding, with Examples from Medieval Texts'. *Literary and Linguistic Computing* **6**(1):34–46.
Tillett, B. (2004). 'What is FRBR? A Conceptual Model for the Bibliographic Universe'. *Library of Congress Cataloging Distribution Service* **25**(5):1–8.
Tummarello, G., Morbidoni, C., Pierazzo, E. 'Toward textual encoding based on RDF.' In *Proceeding of the 9th International Conference on Electronic Publishing* (ELPUB 2005), Kath. Univ. Leuven, June, pages 57-63. 2005.
Tupman, C., Jordanous, A., Stanley, A. (2013). 'Contemporary solutions to retrieve and publish information in ancient documents using RDF and Islandora'. In *Proceedings of Digital Humanities*, Nebraska, US.
Varnienė-Janssen, R., Juskys, J. (2011). 'Strategic, Methodological and Technical Solutions for the Creation of Seamless Cultural Heritage Content: Lithuanian Approach'. In *Proceedings of Summer School in the Study of Historical Manuscripts*, Zadar, Croatia.
Zhitomirsky-Geffet, M., Prebor, G., Bloch, O. (2015). Improving proverb search and retrieval with a generic multidimensional ontology. *Journal of the Association for Information Science and Technology (early view online),* doi: 10.1002/asi.23573.

Zorich, D. M. (2008). 'A Survey of Digital Humanities Centers in the United States'. *Tech. Rep. 143*, Council on Library and Information Resources, Washington, DC.

[i] Examples of such publications include: 'Inscriptions of Aphrodisias', http://insaph.kcl.ac.uk/iaph2007/ ; 'TheArchimedes Palimpsest Project', http://archimedespalimpsest.net/ ; 'Jane Austen's Fiction Manuscripts', http://www.janeausten.ac.uk ; 'Biblioteca Virtual Miguel de Cervantes', http://www.cervantesvirtual.com/ ; 'Perseus Digital Library', http://www.perseus.tufts.edu/. All last accessed October 2012.
[ii] Their significance is underlined by the fact that Caxton's first imprint (the first book ever published in England) was one such collection [Caxton, 1877)].
[iii] http://www.ancientwisdoms.ac.uk
[iv] A concept not always without difficulties, but these can be ignored for the purposes of this discussion.
[v] They concluded that while 'text is best represented as an ordered hierarchy of content object... the hierarchical model can allow future use and reuse of the document as a database, hypertext, or network' (DeRose et al., 1990, p. 3); 'Some structure cannot be fully described even with multiple hierarchies, but requires arbitrary network structures' (DeRose et al., 1990, p. 19).
[vi] RDF (2004). 'RDF/XML Syntax Specification (Revised)'. http://www.w3.org/TR/rdf−syntax−grammar/ . Last accessed October 2012.
[vii] Also 'Text Encoding Initiative / Tracker / Feature Requests / Encoding RDF relationships in TEI - ID: 3309894 (Discussion)'. http://sourceforge.net/tracker/?func=detail&atid=644065&aid=3309894&group_id=106328 . Last accessed October 2012.
[viii] 'OntoMedia Ontology - Working Model', http://www.contextus.net/ontomedia. Last accessed October 2012.
[ix] http://www.tei-c.org/SIG/Ontologies/
[x] As reported in the SIG's meeting minutes at http://wiki.tei-c.org/index.php/SIG:Ontologies
[xi] As acknowledged in the documentation at http://www.tei-c.org/SIG/Ontologies/guidelines/guidelinesTeiMappableCrm.xml
[xii] See discussion at http://sourceforge.net/tracker/?func=detail&atid=644065&aid=3309894&group_id=106328, and guidelines at http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-relation.html
[xiii] 'CIDOC-CRM', http://www.cidoc-crm.org
[xiv] 'Bibliographic Ontology', http://purl.org/ontology/bibo/ ; 'Simplified Ontology for Bibliographic Resources (SOBR)', https://gist.github.com/1331983 . Both last accessed October 2012.
[xv] 'Scholarly Works Application Profile', http://www.ukoln.ac.uk/repositories/digirep/index/Scholarly_Works_Application_Profile. Last accessed October 2012.
[xvi] 'SPAR', http://purl.org/spar . Last accessed October 2012.
[xvii] In the FaBiO documentation for 'fabio:manuscript', http://www.essepuntato.it/lode/http://purl.org/spar/fabio#d4e3689 . Last accessed October 2012.
[xviii] M. O. Jewell, et al. (2010). 'Stories Ontology - Working Draft'. http://contextus.net/stories/ . Last accessed October 2012.
[xix] 'SKOS Simple Knowledge Organization System Reference', http://www.w3.org/TR/skos-reference/ ; 'DCMI Metadata Terms', http://www.dublincore.org/documents/dcmi−terms/. Both last accessed October 2012.
[xx] 'eCRM', http://erlangen-crm.org/current/ ; 'EFRBRoo', http://erlangen-crm.org/efrbroo/ . Both last accessed October 2012
[xxi] A Section is a structural part of a CollectionInstance. One Section has a slightly different sequence to another Section but is related, e.g. through editorial decisions made whilst copying.
[xxii] By the project team.
[xxiii] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-seg.html
[xxiv] http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-relation.html. See also 'TEI Ontologies SIG'. http://www.tei−c.org/Activities/SIG/Ontologies/ and 'Text Encoding Initiative / Tracker / Feature Requests / Encoding RDF relationships in TEI - ID: 3309894 (Discussion)'. http://sourceforge.net/tracker/?func=detail&atid=644065&aid=3309894&group_id=106328 . Both last accessed October 2012.
[xxv] http://www.pbw.kcl.ac.uk/
[xxvi] https://snapdrgn.net/
[xxvii] For example the place "Aphrodisias" (URI http://pleiades.stoa.org/places/638753) was known by the names Ninoe (in the Classical period); Aphrodeisias (Hellenistic-republican, Roman periods); Lelegon polis (unspecified period); Stauropolis (Late-antique period); and Aphrodisias (Roman, Late-antique periods), but these placenames are linked together by Pleiades. Pleiades also disambiguates between Aphrodisias located in modern-day Turkey and the Aphrodisias in modern-day Spain (URI http://pleiades.stoa.org/places/255978/).
[xxviii] E.g. the person "Aristotle" can be represented by the URI http://dbpedia.org/resource/Aristotle independently of whether he is referred to as Aristotle, Ἀριστοτέλης, أرسطو , Aristoteles, or Aristóteles.
[xxix] Demo available at http://www.ancientwisdoms.ac.uk/media/data/texts.html.
[xxx] A Django project - with one application serving the Folioscope app and one serving and managing the data - is under development which will allow Folioscope to be deployed easily as part of a Django-based site.