New Models for Collaborative Textual Scholarship

Mark Hedges, Anna Jordanous, Stuart Dunn, Charlotte Roueché
King's College London
London, UK
{mark.hedges | anna.jordanous | stuart.dunn | charlotte.roueche} @ kcl.ac.uk

Marc W. Küster, Thomas Selig, Michael Bittorf, Waldemar Artes
University of Applied Sciences
Worms, Germany
{kuester | selig | bittorf | artes} @ ztt.fh-woms.de

Abstract— Researchers in digital humanities have for many years been producing online editions of texts based on TEI XML, a widely-adopted standard for marking up textual resources with semantic content. However, this has led to a certain isolation of information, the so-called 'digital silo', and such modes of digital publication have not always made best use of the possibilities of digital technologies. The model is also challenged by the need to model texts that are by their very nature interconnected. The paper describes a collaborative environment of tools and techniques for working with texts that allows scholars to work with such highly-interconnected material.

Keywords—textual editing; collaboration; linking texts; eResearch; digital humanities; repositories; TextGrid; TEXTvre

I. Introduction

Researchers in digital humanities have for many years been producing online editions of texts and manuscripts, commonly encoded using TEI XML, which has by now been widely adopted as a standard for marking up textual resources with semantic content. While it has been beneficial to have these texts more freely available, a lack of communication and sharing has led to a certain isolation of information, the so-called 'digital silo', and such modes of digital publication have not always exploited the possibilities of the digital to best advantage, especially the possibilities offered by the semantic web.

A particular challenge to this model is raised by the existence of texts that are by their very nature interconnected with other texts, and cannot be analysed or understood by scholars to their full extent unless their representations incorporate a commensurate degree of interconnectedness. Examples of such texts are provided by gnomologia, a genre of text that was widespread throughout the antique and mediaeval Mediterranean world. These manuscripts are collections of 'wise sayings' ('gnomes' or 'gnomic sayings') containing moral or social advice, or expressing philosophical ideas, which for the most part can be traced back to older sources (classical authors or to the Bible) and which in turn were used as the sources for later narrative texts.

In the current paper we describe an environment of tools and techniques for editing and analysing with such texts, which allows scholars to work in collaborative fashion towards such highly-interconnected representations. The vision is a broader digital ecosystem for the humanities, containing texts, annotations, tools and, of course, the scholars themselves.

II. Related Work

In 1990, DeRose and co-authors reflected on how electronic text documents could best be structured for flexibility in use and reuse, concluding that "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." [6].

The Text Encoding Initiative (TEI)[1] is an international standard for the exchange of data, particularly for encoding information about texts. It has been widely adopted as the standard encoding for projects marking up textual data with semantic content [27, 17, 20] and has inspired similar XML encoding standards such as MEI, the Music Encoding Initiative [25]. The popularity of TEI within digital humanities research is due to various factors, such as how it allows the researcher to embed structure and metadata within the transcribed text and produce a variety of useful outputs and indices. That TEI has been adopted as standard by this community means that TEI is supported and actively developed by a wide range of people with suitable expertise. Additionally interoperability with other projects is enhanced if TEI is used for a particular project.

To represent relationships between resources in texts, RDF[2] is appropriate, particularly when supported by an ontology of domain-relevant information and knowledge [3]. Taking advantage of the potential of Linked Data [10] to represent semantically notable inter-relations in textual documents, RDF linking is attracting much recent research attention. RDFa[3] allows RDF to be expressed as attributes within markup language documents, but has been restricted primarily to XHTML documents to date. It is desirable, however, to be able to record links and relations within and between XML documents on a wider scale, including in TEI XML documents [7], so that semantic relation information can be recorded. Previous attempts have been made to accommodate this [5, 11,

[1] TEI: http://www.tei-c.org/
[2] RDF: http://www.w3.org/TR/REC-rdf-syntax/
[3] RDFa: http://www.w3.org/TR/rdfa-syntax/

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

12, 15, 19, 29] but none have been adopted as standard, for various reasons:

- Restricting markup to represent structural information rather than semantic links (EARMARK [19], GODDAG [5] and MCT [12]);
- Barriers in dynamically updating the RDF models as needed, restricting the information sharing and flexibility of updating that we wish to promote (GODDAG [5] and MCT [12]);
- Implementations being available only as prototypes that are problematic to install but are no longer actively maintained (RDFTEF [29]: last source code update 2007, leading it to be dismissed as "[o]nly a "toy" experiment" [22]);
- Implementations being too specifically hard-coded to the approach of a single project, without being more generally applicable [11,15].

A recent development within the TEI community[4] sees the <relation> element used to encode RDF triple information in a TEI document. It encodes the Subject-Predicate-Object triple format through the attributes @active (Subject), @ref (Predicate) and @passive (Object). The major advantage of this development is that in the spirit of RDFa it allows RDF to be encoded directly within the manuscript transcriptions that researchers are already working on, rather than requiring files to be transformed. For the scholar undertaking the editing process, working on the structural metadata of a document integrates seamlessly with capturing relations to other, external objects, even though the former is typically expressed in TEI proper and the latter encoded in RDF. The RDF can be automatically extracted and both can then follow separate dissemination chains. For example, the structural metadata might be stored, disseminated and queried in an XML database, whereas the RDF triples might be handled in a triple store.

Encoding relations and exposing them as RDF triples only leads to truly interoperable Linked Open Data if the properties exposed are themselves well-defined and documented, typically by reference to a published ontology, and the external objects are referred to through a commonly-accepted reference system. This could, for example, be a well-identified part of another document or an identifier of a well-known authority list, e.g. for names. In this way, the exposed RDF triple can abstract from many project-specific encoding decisions at the markup level and support cross-project queries.

The SAWS project (see below) is exploring this use of the <relation> element. Whilst this approach has its benefits, for interoperability with other annotation approaches it requires that appropriate links are specified and published in advance, most usually through the domain ontology, and new relations cannot easily be added on the fly. Also, this textual solution does not include a front-end tool that allows annotators to easily add, remove, edit or visualise links, but instead requires proficiency in text editing and application of encoding markup and schemas.

Although many texts are available in the TEI XML format, some technical expertise is necessary to navigate these texts and annotate them with RDF links, even if a text has already been converted to TEI format by others. It would be beneficial to remove some of these technical barriers to text editing and provide a more intuitive environment and methodology for annotating texts. Additionally, a tool that neatly interacts with a storage repository would be useful, for facilitating permanent storage and sharing of texts and links. The SharedCanvas project [26] provides an online GUI-based environment for representing a manuscript visually using images. Manuscript pages can be annotated directly in the SharedCanvas environment using a 'point-and-click' approach to freely enter annotations of interest, which are stored in Open Annotation[5] format. The SharedCanvas model sees these annotations as links between the Canvas object representing the appropriate (page of the) manuscript and the information added in that annotation. This may be contrasted with an alternative model that allows links to be made between two (or more) pages, with the information on that type of link also being recorded. Also, SharedCanvas targets scenarios where information on the manuscript is mostly available through images rather than transcriptions, hence "the focus [of SharedCanvas] is on the relationships between text and image" [26]. In situations where images are unavailable or unobtainable (for example if the digitisation process may damage an ancient document), or alternatively where the text has already been transcribed or is the main focus of interest, SharedCanvas becomes less appropriate.

In the TextGrid ecosystem this role is taken by the Text Image Link Editor (TBLE [6]), a provenance component primarily developed at the University of Applied Sciences Worms [13] (cf. below)[7]. The TBLE is primarily used for linking the transcriptions of facsimiles or manuscripts with their digitized sources at the sub-page level, though it can also be used to build image annotations. Scholars can link segments of text with sections on the corresponding image. The information on the linking between manuscript fragments and the corresponding transcription is itself stored a format compliant with TEI P5. The TBLE is not the first tool of its kind; other tools include the web-based Text-Image Linking Environment (TILE) [21], currently under development, and Tapor's / the University of Victoria's Image Markup Tool[8], the latter created more or less in the same time frame and also using an XML schema extending TEI P5 to store linking information.

[4] Sourceforge.net discussion: Encoding RDF relationships in TEI - ID: 3309894, at http://sourceforge.net/tracker/?func=detail&aid=3309894&group_id=106328&atid=644065

[5] Open Annotation: http://www.openannotation.org/ . SharedCanvas: http://www.shared-canvas.org/

[6] Text-Bild-Link-Editor, German for Text Image Link Editor

[7] TBLE was presented in [1,14] from which part of this description is taken

[8] http://tapor.uvic.ca/~mholmes/image_markup/index.php

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

III. Digital Editions and the Scholarly Environment

Throughout antiquity and the Middle Ages, anthologies of extracts from larger texts, containing wise or useful sayings, were created and circulated widely, as a practical response to the cost and inaccessibility of full texts in an age when these were all in manuscript form. A particular genre of such texts was the so-called gnomologia, which collected 'wise sayings' (`gnomes' or `gnomic sayings') expressing moral or social advice, or philosophical ideas. There has long been interest in the study of this literature and the relationships between manuscripts and within collections [9, 23, 24]. The key characteristics of these manuscripts are that they are collections of smaller extracts of earlier works, and that, when new collections were created, they were rarely straightforward copies. Rather, sayings were reselected from various other manuscripts, reorganised or reordered, and modified or reattributed. The genre also crossed linguistic barriers, in particular being translated into Arabic, and again these were rarely a matter of straightforward translations; they tend to be variations.

Thus the corpus of material can be regarded as a very complex directed network or graph of manuscripts and individual sayings that are interrelated in a great variety of ways. Analysis of these interrelations can reveal a great deal about the dynamics of the cultures that created and used these texts. Such scenarios offer challenges for more conventional approaches to working with digital texts online, and lends themselves well to applications of linked data approaches [10]. Such approaches can help researchers to gain a clearer understanding of these texts as well as to publish them, tracing cultural dynamics by identifying and marking up relationships and links between and within documents.

The work described in this paper is thus producing a framework for representing these relationships, using an RDF-based semantic web approach, as well as tools for creating these complex resources, and for visualising, analysing, exploring and publishing them. We also envisage scenarios where other projects will want to link their own materials to these texts, thus SAWS will provide a `hub' for future scholarship in this field and in related areas. The number of manuscripts of this type is large, and we regard our work as creating the kernel of a much larger corpus of interrelated material, being shared and distributed to facilitate and enhance research. Many of the subsequent contributions will be made by others; consequently we will create a framework of tools and methods that will enable researchers to add texts and relationships of their own, which will be managed in distributed fashion. Data provenance – the "understanding of the origins of data, as well as the transformations that the data has undergone in order to arrive at its current state" [8] – is a major issue in such environments. In this way we will create an interactive environment that enables researchers not only to search or browse this material in a variety of ways, but also to process, analyse and build on the material.

IV. TextGrid and TEXTvre: ecosystems for textual scholarship

The call for papers defines Digital Ecosystems as "open, loosely coupled, demand-driven, domain clustered, agent-based self-organised collaborative environments where species/agents form a temporary coalition (or longer term) for a specific purpose or goals".[9] In the digital humanities, ecosystems are, in fact, inhabited by human agents – notably scholars and content providers – interacting with the "biotic landscape" of digital agents – tools and services – and the "abiotic landscape" of digital resources (both terms following [16]).

TextGrid is as a virtual research environment (VRE) for the humanities dealing with texts in a wide sense (philology, epigraphy, linguistics, musicology, art history etc.). It brings together eight German institutions from both academia and the commercial sector to "create a community grid for the collaborative editing, annotation, analysis and publication of specialist texts". Building on a distributed grid infrastructure (cf. the architecture diagram[10]), it offers access to an ecosystem of services and content for typical tasks.

Facing towards the end user, TextGrid provides an Eclipse-based rich user interface called TextGridLab that integrates a set of interactive tools such as the XML Editor, the TBLE and the Text-Text Link Editor (discussed below).

TEXTvre[11] [2] builds on TextGrid to embed the TextGrid Virtual Research Environment (VRE) in the research ecosystems of the participating institutions, specifically at King's College London. It integrates the VRE with the currently-used Fedora repository and the institution's data management infrastructures. In addition, it interfaces with other relevant tools and services that the users in these particular research ecosystems regularly use. These include alternatives to standard TextGrid tools, such as the commercial XML editor oXygen as an alternative to the existing open source XML editor, but also additional services for linguistic analysis based on the Open Source framework GATE [4] and interaction with various other content and service resources.

A key characteristic of both of these platforms is that they are *collaborative*, enabling geographically-dispersed researchers to work together on the same corpus of material.

The editing of gnomologia required new tooling to be integrated into TextGrid and TEXTvre, beyond that which was previously supported. In the remainder of the paper, we describe some of the detailed scenarios arising from the SAWS (Sharing Ancient WisdomS) project, which is working with a number of these texts (Section V), and describe a tool developed for the TextGrid/TEXTvre environments that supports these additional requirements (Sections VI and VII).

V. The SAWS Project: networks of gnomologia

The SAWS (Sharing Ancient WisdomS) project is investigating the transmission of information in medieval

[9] http://sesar.dti.unimi.it/DEST2012/

[10] http://www.textgrid.de/typo3temp/pics/68951eb1ac.png

[11] http://textvre.cerch.kcl.ac.uk/

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

manuscripts. The primary focus of the project is on Greek gnomologia, from the ninth to twelfth centuries AD, and on Arabic collections of sayings from the same period. Examples of (respectively) a simple saying or 'gnome' from one of these gnomologia, and of a longer anecdotal section, are:

1. 'One cannot cover a fire with a cloak nor a shameful sin with time.'
2. '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." '

Within this second section there are two parts of interest to the manuscript scholar: the statement itself, by Diogenes ('Because, perhaps, they expect …'), and the narrative text surrounding the statement ('Diogenes was asked...'). We refer to a basic unit of interest within the text (e.g. 1.and 2. above) as a ContentItem.

Often manuscripts are large in size and contain more content than the wisdom sayings in which we are interested. Also, a collection of sayings, which we represent with the term CompilationInstance, can span several (parts of) different manuscripts.

Over the centuries, manuscripts were often transcribed by various scribes. Different compilers organised the collections in different ways; perhaps according to author, or alternatively according to themes within the sayings, and then according to author within each theme. During the transcription process, there were also many discrepancies, misattributions, mistakes in transmission, or sections missed out.

We wish to explore relationships within a particular CompilationInstance (between different manuscripts and within a single manuscript), between CompilationInstances, between languages, between CompilationInstances and source texts (e.g. the original transcriptions of the sayings) and between CompilationInstances and edited literary texts which made use of them. Example relationships include:

- Manuscript isWrittenAt Scriptorium
- Manuscript isInLanguage Language
- CompilationInstance isWrittenBy Scribe
- CompilationInstance isTranslationOf CompilationInstance
- Section isSequentiallySimilarTo Section [i.e. one Section of a CompilationInstance has a slightly different sequence to another Section but is related, for example through editorial decisions made whilst copying]
- ContentItem isShorterVersionOf ContentItem
- ContentItem isVerbatimOf ContentItem

Clearly, a text may have several relationship statements which can be made about it. The definition of relationships has been a key activity for the SAWS project; it is however not simply a mechanical process, but one from which the researchers will learn more about their own texts. The overarching aim of such a model is to allow researchers to represent, identify and analyse the flow of knowledge across texts and cultures. This not only enriches the texts themselves but also lays the basis for a study of the cultural dynamics across the centuries of Greek and Arabic thought, and cultural exchange across civilisations. In developing our model, our vocabulary[12] is intended to express not only the relationships among the texts and textual excerpts in our set of texts, but also those that may occur in analogous bodies of material.

A key part of the SAWS project is to identify and publish data and inter-relations between data in the manuscripts as Linked Data on the Semantic Web. To this end, we both provide URIs to link into the SAWS data and link out from the SAWS data to external data sources, including:

- Pleiades historical gazetteer of ancient places (http://pleiades.stoa.org/),
- Prosopography of the Byzantine World: an online collection of data on people in the ancient world (http://www.pbw.kcl.ac.uk/),
- Geonames (http://www.geonames.org/) to refer to locations not covered in Pleiades,
- ISO-639-2 (http://id.loc.gov/vocabulary/iso639-2/) , a standard for referring to languages,
- Libraries of digital transcriptions of relevant documents:
- Perseus Digital Library (http://www.perseus.tufts.edu/),
- Thesaurus Linguae Graecae (http://www.tlg.uci.edu),
- Canonical Text Service (http://cts3.sourceforge.net/),
- DBpedia (http://dbpedia.org/) as an additional source of data for items/resources not covered above.

By linking to other sources in this way we encourage more sharing of our scholars' data, provide access points to the data such as for people interested in linked entities e.g. Aristotle, and make our Linked Data part of the Semantic Web.

VI. The Text-Text-Link-Editor

As relationships between texts and text fragments become more important to the work of scholars in the humanities, a way is needed of editing these relationships and storing them in a persistent manner. The Text-Text-Link-Editor (TTLE) is a tool offering these functionalities. TTLE is part of the TextGrid project and as such has access to the services and tools already offered by TextGrid, allowing the textual researcher to do most of their daily work using a single set of tools. As mentioned earlier, TEI is the most commonly used format for digital texts. Therefore TTLE's persistence layer is based on the linking functionality already available in the TEI specification, and extends it at some points to offer a wider range of use cases. Even though there are other standards available to describe text linking (e.g. the Open Annotation Data Model), using TEI allows TTLE to automatically utilize the services already

[12] Available as an OWL ontology at http://purl.org/saws/ontology

978-1-4673-1703-0/12/$31.00 ©2013 IEEE

available in the TextGrid repository (e.g. search services) that are based on TEI. This approach is not only cost-effective but also helps to reduce errors, as the code for these services has already undergone extensive testing.

TTLE is implemented for the most part as a web application, although it can be integrated into TextGridLab as a plug-in component, providing browser functionality and an interface to the rest of TextGridLab. The backend server handles data persistence, caching, data validation and serialization of link collections into TEI.

Being part of TextGrid benefits TTLE in multiple ways, but also obliges it to conform to various restrictions. One of these restrictions is caused by the ability of TextGrid to work with write-protected documents. In its normal operation, TTLE simply inserts specific tags in the original XML document to mark selected text fragments. This is of course impossible for write-protected data, so TTLE offers two additional methods for marking text fragments, both of which store data only in the TTLE TEI file.

One method is to use character offsets, but this only works properly for immutable documents. In mutable documents, any change to the document will change the selected text fragment. Alternatively, one can select any text that is enclosed in an XML tag with a unique identifier. This also works for documents that are still being modified by their owner, but on the other hand it only allows the selection of predefined text fragments. Even with these different ways of addressing text fragments, it is still possible that a document may be modified in some way, in which case the originally-linked text fragment may no longer be referenced properly. To be able at least to inform the user about such a modification, TTLE has to first realise that the text fragment currently referenced is not the one that was originally selected. To facilitate this, when adding a text fragment to a link, the backend calculates a hash value of the selected text and stores this hash value together with the reference in the TTLE TEI file. Every time this file is opened, all such hash values for text fragments are recalculated and compared to the values stored in the file. Any mismatch indicates that a reference is no longer valid.

As mentioned above, the TTLE user interface is embedded in the TextGridLab client. The TTLE plug-in defines a new view within the client which shows only the TextGridLab Navigator and the TTLE itself, providing the scholar with a clean and uncluttered interface. The embedded browser provides two viewports for displaying up to two documents at the same time. If more than two documents are referenced in the TTLE TEI file, any two of them can be selected for display in the two viewports. All links stored in the TTLE TEI file are listed in a sidebar that contains all fragments of a link sorted by target document, to ensure easy access to the data being linked. If the scholar selects a link or fragment in the sidebar, this will be highlighted in a configurable colour in the corresponding viewport. This simplifies finding fragments in the documents. If a selected fragment is located in a document that is not currently visible, one of the viewports will switch automatically to this document. As default, a document is displayed with XML tags, but it is also possible to hide these tags and to display the document as plain text. All basic operations – such as creating a new TTLE TEI file, or opening a document to select new fragments – are handled through the familiar TextGridLab interface, providing scholars with a consistent workflow.

There are still some problems with text sections in languages that are not written from left to right (e.g. Arabic). One of the major right-to-left problems currently is an unpredictable behaviour occurring when the user tries to select text which contains writing in more than one direction. This can be traced back to browser incompatibilities, and a workaround is under development.

The current version of TTLE allows users to add comments for each link, and the next release will also offer a way of specifying link types, for easy sorting, searching and filtering of links. These link types will be user- or project-specific, avoiding an ever-growing list of publicly available link types.

VII. Joining things up

If flexibility in use and reuse is a key motive for creating, structuring and exposing data, it also is a game changer in the research process itself. Having both source (= manuscript) and target (= published) data available and linked up helps to create transparency within the publication process and the provenance of results. As we have seen, provenance information can be exposed using Linked Open Data, and for this purpose suitable constructs can be defined within a core ontology to provide a provenance vocabulary at the level of the research objects. This is the approach already pioneered for many years in the cultural heritage world by the CIDOC Conceptual Reference Model (CRM), aka ISO 21127:2006, for cultural heritage documentation. In [28], the model is extended explicitly to cover provenance information according to the requirements of the Open Archival Information System (OAIS aka ISO 14721:2003).

However, because this approach centres on the provenance of complete objects, it is insufficient for provenancing philological data, where much of the provenance challenges lie at a very fine-granular level (linking of individual image fragments and text passages as well as text passages amongst each other). One of the challenges addressed by SAWS, together with TBLE and TTLE, is the problem of exposing provenance information at this fine level of granularity, so as to trace in full the process from source document to final research results, as well as the relationships between parts of objects.

Acknowledgment

The work described in this paper was undertaken as part of projects supported by a number of funding bodies: SAWS, which is funded by HERA (Humanities in the European Research Area); TEXTvre, which was funded by JISC (Joint Information Systems Committee), with follow-on funding from the Software Sustainability Institute (SSI); TEXT.Link, which was also funded by JISC as a follow-on to TEXTvre; TextGrid, which is is part of the D-Grid initiative, and is funded in its second phase by the German Federal Ministry of Education and Research (BMBF) for the period starting June 1, 2009 to May 31, 2012 (reference number: 01UG0901A).

REFERENCES

[1] Yahya Ahmed Ali Al-Hajj and M.W. Küster, The Text-Image-Link-Editor: a tool for linking facsimiles and transcriptions and image annotations, Digital Humanities 2011: Conference Abstracts; Stanford University, Stanford, CA, USA, June 19 – 22, 2011. Stanford: Stanford Univ. Library 2011, 74 - 76

[2] T. Blanke, M. Hedges: Humanities e-Science: From systematic investigations to institutional infrastructures. In: Proceedings of the 6th IEEE e-Science conference. Brisbane, Australia, 2010.

[3] B. Chandrasekaran, John R. Josephson and V. Richard Benjamin, “What Are Ontologies, and Why Do We Need Them”, IEEE Intelligent Systems, Vol. 14 No. 1, pp.20-26 (1999)

[4] H. Cunningham, D. Maynard, K. Bontcheva, V. Tablan: "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications". Procedures of the 40th Anniversary Meeting of the Association for Computational Linguistics, 2002.

[5] A. Dekhtyar and I. E. Iacob. A framework for management of concurrent XML markup q. Text, Vol. 52 No. 2, pp.185-208, 2005.

[6] S. J. DeRose, D. G. Durand, E. Mylonas, and A. H. Renear. What is text, really? Journal of Computing in Higher Education, Vol. I No. 2, pp.3-26, 1990.

[7] O. Eide, A. Felicetti, C. Ore, A. D’Andrea, and J. Holmen. Encoding Cultural Heritage Information for the Semantic Web. In EPOCH Conference on Open Digital Cultural Heritage Systems, Rome, Italy, 2008.

[8] B. Glavic, G. Alonso: The perm provenance management system in action. SIGMOD ’09: Proceedings of the 35th SIGMOD international conference on Management of data, 2009, pp.1055-1058.

[9] D. Gutas, “Classical Arabic Wisdom Literature: Nature and Scope”, Journal of the American Oriental Society, Vol. 101 No. 1 Oriental Wisdom (Jan. -Mar., 1981), pp.49-86

[10] T. Heath and C. Bizer, Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, Vol. 1 No. 1, pp.1-136. Morgan & Claypool (2011)

[11] M.O. Jewell, “Semantic Screenplays: Preparing TEI for Linked Data”, Digital Humanities 2010, Friday 9 July, London, UK (2010)

[12] H. V. Jagadish, L. V. S. Lakshmanan, M. Scannapieco, D. Srivastava, and N. Wiwatwattana. Colorful XML: One Hierarchy Isn’t Enough. In Proc ACM SIGMOD Int Conf on Management of Data, Vol. 1, pp.251–262. ACM Press, 2004.

[13] M. W. Küster, C. Ludwig, A. Aschenbrenner: Textgrid as a Digital Ecosystem. Digital EcoSystems and Technologies Conference, 2007. DEST '07. Inaugural IEEE-IES. Cairns: IEEE 2007 S. pp.506-511.

[14] M. W. Küster, C. Ludwig, Yahya Ahmed Ali Al-Hajj, TextGrid provenance tools for digital humanities ecosystems, Proceedings of the 5th IEEE International Conference on Digital Ecosystems and Technologies, 31st May 2011-3rd June 2011. Daejeon, South Korea: IEEE 2011, 317-323.

[15] K. F. Lawrence. Wherefore Art Thou? - Crowdsourcing Linked Data from Shakespeare to Dr Who. In Proceedings of Web Science, Koblenz, Germany, 2011.

[16] C. Ludwig, M. W. Küster. Digital Ecosystem of eHumanities resources and service. 2008 2nd IEEE International Conference on Digital Ecosystems and Technologies : [DEST 2008]; Phitsanuloke, Thailand, 26 - 29 February 2008. Piscataway, NJ: IEEE Service Center 2009 S.476-481.

[17] E. Mylonas and A. Renear. The text encoding initiative at 10: Not just an interchange format anymore - but a new research community. Computers and the Humanities, Vol. 33 No. 1, pp.1-9, 1999.

[18] S. Nichols. Time to change our thinking: Dismantling the silo model of digital scholarship. Ariadne, Vol. 58, 2009.

[19] S. Peroni and F. Vitali. Annotations with EARMARK for arbitrary, overlapping and out-of order markup. In Proceedings of the 9th ACM symposium on Document engineering, pp. 171-180, Munich, Germany, 2009.

[20] E. Pierazzo. A rationale of digital documentary editions. Literary and Linguistic Computing, Vol. 26 No. 4, pp.463-477, 2011.

[21] D. C. Porter, D. Reside, J. Walsh: Text-Image Linking Environment (TILE). The 21st Joint International Conference of the Association for Literary and Linguistic Computing, and the Association for Computers and Humanities and The 2nd Joint International Conference of the Association for Literary and Linguistic Computing, the Association for Computers and Humanities, and the Society for Digital Humanities - Société pour l'étude des médias interactifs 2009, pp.388-390.

[22] P. Portier, N. Chatti, S. Calabretto, E. Egyed-Zsigmond, and J. Pinon. Modeling, encoding and querying multi-structured documents. Information Processing & Management. In press.

[23] M. Richard, “Florilèges grecs”, Dictionnaire de Spiritualité V (1962), cols. 475-512

[24] F. Rodríguez Adrados, Greek wisdom literature and the Middle Ages: the lost Greek models and their Arabic and Castilian Translations (2001), English translation by Joyce Greer (2009), pp.91-97 on Greek models

[25] P. Roland. The Music Encoding Initiative (MEI). In Proceedings of the First International Conference on Musical Applications Using XML, pp.55-59, 2002.

[26] Sanderson, R. Albritton, B. Schwemmer, R. Van de Sompel, H. "SharedCanvas: A Collaborative Model for Medieval Manuscript Layout Dissemination". Procs of the 11th ACM/IEEE Joint Conference on Digital Libraries (conference site), Ottawa, Canada, June 2011

[27] C. M. Sperberg-McQueen. Text in the electronic age: Texual study and textual study and text encoding, with examples from medieval texts. Literary and Linguistic Computing, Vol. 6 No. 1, pp.34-46, 1991.

[28] M. Theodoridou, Y. Tzitzikas, M. Doerr, Y. Marketakis, V. Melessanakis: Modeling and querying provenance by extending CIDOC CRM. Distrib. Parallel Databases, Vol. 27, April 2010, p. 169-210.

[29] G. Tummarello, C. Morbidoni, and E. Pierazzo. Toward textual encoding based on RDF. In Proceeding of the 9th International Conference on Electronic Publishing (ELPUB 2005), Kath. Univ. Leuven, number June, pp.57–63, 2005.

[30] D. M. Zorich. A survey of digital humanities centers in the United States. Technical Report 143, Council on Library and Information Resources, Washington, DC, November 2008

978-1-4673-1703-0/12/$31.00 ©2013 IEEE