When TEI Verse becomes linked data: using TEI tags as a model to build a linked poetry system

Keywords: verse module, poetry, linked data, metrics, controlled vocabulary

The Text Encoding initiative is conceived as a standard system designed to let interoperability between different projects working on analogous topics. However, interoperability is sometimes not easy to reach due to the different interpretation of tags made by each project or group. Given the variety of tags, there are several concepts that can be easily identified for many projects, such as <msIdentifier> or <incipit>, which referent is pretty clear and recognizable. There are some other tags, however, which can experience multiple interpretations, limited by the content of the attributes that modify them.

The case of Verse module is controversial, especially in its parts related to metrics, due to several issues: first of all, there are not many projects using the full set of Verse tags (most projects use <lg> and <l> and perhaps <rhyme> but do not add many more metrical and rhythmic information), and some of them complete the information related to specific kind of rhythms and stanzas using attributes and new customized tags. This situation makes it even more difficult to make poetry projects interoperable, which becomes a dream, especially after taking into account that many of the projects working in this area -such as metrical repertoires as The Cantigas de Santa Maria Database: http://csm.mml.ox.ac.uk/, Analecta Hymnica Digitalia: database on Medieval Latin poetry: http://webserver.erwin-rauner.de/crophius/Analecta_conspectus.htm, Bibliografia Elettronica dei Trovatori: http://w3.uniroma1.it/bedt/BEdT_03_20/, Le Nouveau Naetebus: database on French narrative Medieval poetry : http://www.nouveaunaetebus.elte.hu/ Répertoire de la Poésie Hongroise Ancienne (RPHA) : Repertoire on Medieval Hungarian poetry: http://rpha.elte.hu/ , MedDB: Lírica Profana Galego-Portuguesa http://www.cirp.es/pls/bdo2/f?p=MEDDB2, Corpus rhythmorum musicum (IV-IX secolo): database on Latin Medieval poetry accompanied with music http://www.corimu.unisi.it/, Skaldic poetry of Scandinavian Middle Ages https://www.abdn.ac.uk/skaldic/db.php , or Dimev: Digital index of medieval English verse: http://www.cddc.vt.edu/host/imev/record.php?recID=6768, use a variety of technological systems, mostly consisting of databases which do not tag their texts with XML.

This paper presents a model to link poetry projects based on TEI tags. It consists of a theoretical schema of linked data, whose entities are extracted from the different tags used by the project REMETCA www.remetca.uned.es, the Spanish Digital Repertoire on Medieval Poetry, conceived and created on a TEI basis. This contribution wants to show how each tag of the xsd schema can be translated and transposed into RDF (following standards like Dublin Core for elements of the <teiHeader> such as <author> or <date>) and published later via linked data in order to create a macro-schema in which other poetical repertoires can see many of their terms identified with an existing category, which is equivalent both to a TEI tag and to a URI.

In this translation from TEI to linked data, there are some tags which have quick and easy identifiers in standards like Dublin Core http://dublincore.org/ or CIDOC-CRM http://www.cidoc-crm.org/ (Eide & Ore 2007), but the main problem arrives when dealing with the tags of the Verse module, since there are no equivalents in other metadata systems and it is necessary to create a new entity (this is the case, for example of <met> or <rhyme>). Difficulties get bigger after looking at the different digital resources working on poetry, completely conditioned by their own literary traditions which have no equivalent in other literatures and a very different way of classification.

Our proposal consists of an abstract system of entities and properties which can be "translated" and identified in the different poetical traditions and languages. This linked data model is represented in RDF in two levels: the abstract level in which TEI tags are transformed into "entities", and each tag is supposed to have a possible translation into every literary tradition, and a "vocabulary" level, populated with controlled terms whose content varies depending on the different poetic traditions. The words of these vocabularies are equivalent to the content of the attributes in a TEI system, which varies depending on the project. This interpretation, which at the moment consists of a prototype represented as linked data using Webprotege software (Tudorache et al. 2011), is conceived at a double level equivalent to the structure of tags and attributes of the TEI. This representation of TEI tags will make it easier for poetry projects to identify their own needs and common fields and will let them to link their contents within this system, even if they are built in different software not XML based.

References

Bootz, P. & S. Szoniecky, "Towards an ontology of the field of digital poetry", paper presented at Electronic Literature in Europe, 2008. Full text available at http://elmcip.net/node/415

Burnard, L. & S. Bauman, eds., "TEI P5: Guidelines for Electronic Text Encoding and Interchange. Ver. 2.5.0." Accessed 30-10-2013. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/

Eide, Ø. & C.-E. Ore, "From TEI to a CIDOC-CRM Conforming Model: Towards a Better Integration between Text Collections and Other Sources of Cultural Historical Documentation", paper presented at the DH conference 2007. Abstract available at: http://www.edd.uio.no/artiklar/tekstkoding/poster_156_eide.html

González-Blanco, E. & L. Seláf, "Megarep: A comprehensive research tool in Medieval and Renaissance poetic and metrical repertoires", Humanitats a la xarxa: món medieval / Humanities on the web: the medieval world, edited by L. Soriano, M. Coderch, H. Rovira, G. Sabaté & X. Espluga., Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien, Peter Lang, 2013 (forthcoming).

González-Blanco, E. & J. L. Rodríguez, "ReMetCa: a TEI based digital repertory on Medieval Spanish poetry", at The Linked TEI: Text Encoding in the Web, Book of Abstracts - electronic edition. Abstracts of the TEI Conference and Members Meeting 2013: October 2-5, Rome edited by Fabio Ciotti & Arianna

Ciula, DIGILAB Sapienza University & TEI Consortium, Rome 2013, 178-185. http://digilab2.let.uniroma1.it/teiconf2013/abstracts/. Accessed 30-10-2013.

TEI SIG ontologies group http://wiki.tei-c.org/index.php/SIG:Ontologies. Accessed 22/05/2014.

Tudorache, T., C. I. Nyulas, N.F. Noy & M.A. Musen, "WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web", Semantic Web Journal, IOS Press, 2011. http://www.semantic-web-journal.net/content/webprot%C3%A9g%C3%A9-distributed-ontology-editor-and-knowledge-acquisition-tool-web. Accessed: 30/10/2013.