

Linked open data to represent multilingual poetry collections. A proposal to solve interoperability issues between poetic repertoires.

María Gimena del Rio, Elena González Blanco García y Clara Martínez Cantón.

Cita:

María Gimena del Rio, Elena González Blanco García y Clara Martínez Cantón (Mayo, 2016). *Linked open data to represent multilingual poetry collections. A proposal to solve interoperability issues between poetic repertoires. LDL 2016 ? 5th Workshop on Linked Data in Linguistics: Managing, Building and Using Linked Language Resources. LREC, Portoroz.*

Dirección estable: <https://www.aacademica.org/gimena.delrio.riande/51>

ARK: <https://n2t.net/ark:/13683/pdea/xcv>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Linked open data to represent multilingual poetry collections. A proposal to solve interoperability issues between poetic repertoires

Elena González-Blanco¹, Gimena del Río², Clara I. Martínez Cantón¹

¹Universidad Nacional de Educación a Distancia (UNED)

Bravo Murillo, 38 -- Madrid, Spain

²Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

C1033AAJ, Av Rivadavia 1917, Buenos Aires, Argentina

E-mail: egonzalezblanco@flog.uned.es, gdelrio@conicet.gov.ar, cimartinez@flog.uned.es

Abstract

This paper describes the creation of a poetic ontology in order to use it as a basis to link different databases and projects working on metrics and poetry. It has been built on the model of the Spanish digital repertoire ReMetCa, but its aim is to be enlarged and improved in order to fit under every poetic system. The conceptual semantic model, written in OWL, includes classes and metadata from standard ontological models related to humanities fields (such as CIDOC or Dublin Core), and adds specific elements and properties to describe poetic phenomena. Its final objective is to interconnect, reuse and locate data disseminated through poetic repertoires, in order to boost interoperability among them.

Keywords: ontology, TEL, verse, metrics, poetry, CIDOC

1. Introduction

Poetic features have been analysed and classified in the different literary traditions since the beginnings of Literary Theory studies. These features have been organized in form of poetic repertoires (first printed in paper and later web-based) that give account of metrical and rhythmical schemes of each poetical tradition or school. They gather long corpora of poems, which are defined by their main characteristics.

Performing comparative analysis of the existing digital poetic repertoires and databases poses important problems, as data sources are a rich and heterogeneous mosaic of virtual poetry collections integrated by multilingual corpora, such as French lyrical collections (Nouveau Naetebus), Italian (BedT), Hungarian (RPHA), Medieval Latin (Corpus Rhythmorum Musicum, Annalecta Hymnica Digitalia, Pedecerto), Gallego-portuguese (Oxford Cantigas de Santa María, MedDB2), Castilian (ReMetCa), Dutch (Dutch Song Database), Occitan (BedT, Poésie Neotroubadouresque, The last song of the Troubadours), Catalan (Repertori d'obres en vers), Skaldic (Skaldic Project), or German (Lyrik des Minnesänger), among others.

Each repertoire belongs to its own poetical tradition and each tradition has developed its own analytical terminology for years in a different and independent way (González-Blanco & Sélaf, 2014). The result of this uncoordinated evolution is a bunch of varied terminologies to explain analogous metrical phenomena through the different poetic systems whose correspondences have been hardly studied.

From the philological point of view, there is no uniform academic approach to analyse, classify or study the different poetic manifestations, and the divergence of theories is even bigger when comparing poetry schools from different languages and periods. To give an example, the same quatrain of dodecasyllables can be encoded in different ways depending on the philological tradition: 12A12A12A12A, 4x(7pp+7p) or 4aaaa; or even named

with different meaning: “alexandrine” means 14-syllable line in Spanish but only 12-syllables in French.

There are also important technical issues, as these repertoires were created in different periods and most of them are driven by stand-alone collected databases. The ER (Entity-Relationship) data model is the most commonly used for this purpose, together with the data model based on records for the logical implementation (Elmasri & Navathe, 2011), which is widely accepted, but the technological implementation varies from one project to another. So, there are repertoires that use SQL databases, others that are based on XML tagging or even new models based on non-structured databases.

Although the current ICT infrastructures are prepared to harvest such collections and provide access to them by a search engine, it is absolutely necessary to standardize metadata and vocabularies at philological level in order to be able to climb up the semantic layer and link data between different traditions. There are a few studies which deal obliquely with some of the above mentioned aspects (Bootz & Szoniecky, 2008; Zöllner-Weber, 2009), but there is not yet a conceptual model of ontology referred to metrics and poetry.

The closest related works to this topic are probably the conceptual model of CIDOC¹, the vocabularies of the Getty Museum², as they are designed to express relations and artistic manifestations in the field of humanities, the controlled vocabularies of English Broadside Ballad Project³ and the linked data relations offered by the Library of Congress⁴, which do not offer a deep information on metrics vocabulary.

2. Our Proposal

The aim of this paper is to present a model able to serve as a uniform solution for terminological issues in order to

¹ <http://www.cidoc-crm.org>

² <http://www.getty.edu/research/tools/vocabularies>

³ <http://ebba.english.ucsb.edu/>

⁴ <http://id.loc.gov/>

2.1 The Basis for our Proposal

2.2 The Data Model

The implementation of the model makes use of one of the most recognized standards for the Semantic Web description: the Ontology Web Language (OWL), developed by W3C as an extension of RDFS⁶. The ontology integrates sets of predefined metadata using namespaces and it has been built using WebProtégé (Tudorache et al. 2011).

The resulting first version of this ontology is hosted at various places⁷.

Figure 2 below shows a sample of the classes, properties and data properties of the poetry ontology related to the previous ReMetCa data model (Figure 1):

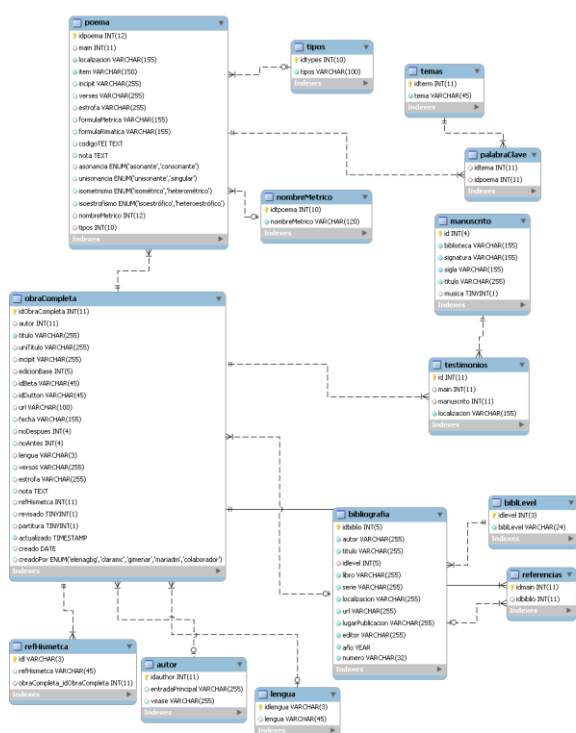
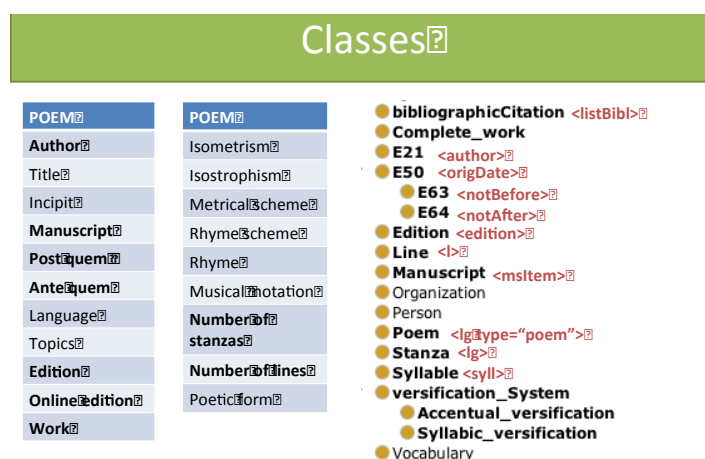


Figure 1: ReMetCa data model



⁶ <https://www.w3.org/TR/owl-features/>

7 www.purl.org/net/ReMetCa;
<http://datahub.io/dataset/ReMetCa-ontology>;
<http://lov.okfn.org/dataset/lov/vocabs/ReMetCa>.

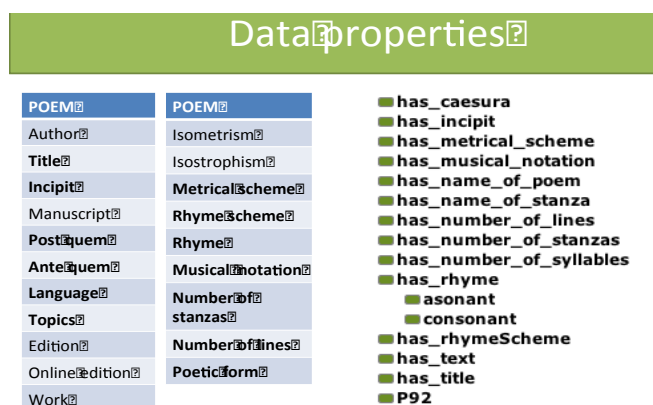
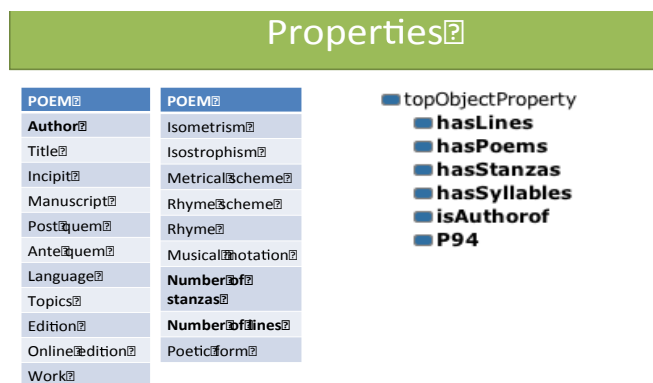


Figure 2: Sample of classes, properties and data properties related to the ReMetCa data model

This model is based on previous works that combine linked data and TEI. There are some preliminary approaches in the field of Philology developed by Christian-Emil Ore and Øyvind Eide (2007). Although the authors focus on the use of Topic Maps, they also point at the creation of a Conceptual Referential Model (CRM) model based on the TEI document and filled with all instances of mapped elements, having in mind that although TEI provides a richer vocabulary than EAD (Encoded Archival Description) or DCMI (Dublin Core Metadata Initiative), it is less abstract than RDF or METS (Metadata Encoding and Transmission Standard). Taking all these reflections into consideration, our ontology includes some elements of CIDOC into its classes and properties. For example, the entity “author” can be linked with DC: creator, FOAF: agent and CIDOC E21, as this is shown among other mappings in Figure 3 below.

As this is a relatively small ontology with interoperability purposes, there are not many elements shared with other ontologies. We have 13 entities, 5 object properties and 14 data properties or attributes. Almost all of them have a TEI equivalent or origin (especially the entities), but only a few are linked with other ontologies or vocabularies, such as CIDOC or DC. Only in these cases, our ReMetCa URIs have been substituted by their existing URIs.

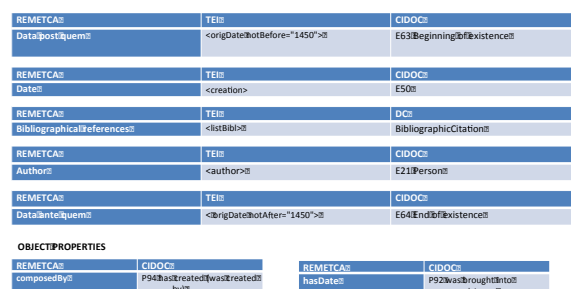


Figure 3: Mapping among ReMetCa, TEI and CIDOC models

Another issue is the number of attributes used for issues like “topics”, “names of the poem types” or “functionality”. All these categories are defined by the content of attributes like <poem type="">, or @subtype. The solution has been including TemaTres terminological software both to work as a lexical/content provider for TEI tags and to organize metadata. A general controlled vocabulary on Medieval Castilian Poetry at CAICYT-CONICET’s Semantic Server in order to create a more consistent categorial prototype has been set at <http://vocabularios.caicyt.gov.ar/pmc/> which is one of the most useful applications of Linked Data in combination with TEI of this proposal, as it complements the XML structure with enriched content semantically organized and structured.

3. Conclusion

To sum up, this project of a poetic and metrical ontology intends to be much more than a repository of datasets, thesauri or controlled vocabularies. It aims to create a semantic standardized structure to describe, analyze and develop logical operations through the different poetic digital repertories and their related resources. Its final objective is to interconnect, reuse and locate the data disseminated through poetic databases in order to get interoperability among projects, to perform complex searches and to make the different resources “talk” to each other following a unique but adaptive model.

4. Acknowledgements

This paper has been developed thanks to the research projects funded by MINECO and led by Elena González-Blanco: Acción Europa Investiga EUIN2013-50630: Repertorio Digital de Poesía Europea (DIREPO) and FFI2014-57961-R. Laboratorio de Innovación en Humanidades Digitales: Edición Digital, Datos Enlazados y Entorno Virtual de Investigación para el trabajo en humanidades, and the Starting Grant ERC-2015-STG-679528 POSTDATA.

5. Bibliographical References

- Bootz, P. & S. Szoniecky, (2008). "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." <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>. Accessed 30-10-2015.
- Ciula, A., P. Spence & J. M. Vieira. (2008). "Expressing complex associations in medieval historical documents: the Henry III Fine Rolls Project", *Literary and Linguist Computing* 2008, 23 (3): 311-325.
- Eide, Ø. & C.-E. Ore. (2007). "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
- Elmars, R. & S. B. Navathe. (2011). *Fundamentos de Sistemas de Bases de Datos*, Madrid, Pearson, Addison Wesley, 2011.
- González-Blanco, E. & J. L. Rodríguez. (2013). "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.
- González-Blanco, E. & L. Seláf. (2014). "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, 2014.
- Tudorache, T., C. I. Nyulas, N.F. Noy & M.A. Musen. (2011). "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: 30/10/2015.
- Zöllner-Weber, A. (2009). "Ontologies and Logic Reasoning as Tools in Humanities?", *DHQ* 2009, 3: 4. <http://www.digitalhumanities.org/dhq/vol/3/4/000068/000068.html> Accessed: 30/10/2015.

