

Presentation

POSTDATA – Towards Publishing European Poetry as Linked Open Data

Mariana Curado Malta
Polytechnic of Oporto, Portugal
UNED-LINHD, Madrid, Spain
mariana@iscap.ipp.pt

Elena Gonzalez-Blanco
UNED-LINHD, Madrid, Spain
egonzalezblanco@flog.uned.es

Abstract

POSTDATA is a 5 year European Research Council (ERC) Starting Grant Project that began in May 2016 and is hosted by the Universidad Nacional de Educación a Distancia (UNED), Madrid, Spain. The context of the project is the corpora of European Poetry (EP), with a special focus on poetic materials from different languages and literary traditions. POSTDATA aims to offer a standardized model in the philological field and a metadata application profile (MAP) for EP in order to build a common classification of all these poetic materials. The information of Spanish, Italian and French repertoires will be published in the Linked Open Data (LOD) ecosystem. Later we expect to extend the model to include additional corpora.

There are a number of Web Based Information Systems in Europe with repertoires of poems available to human consumption but not in an appropriate condition to be accessible and reusable by the Semantic Web. These systems are not interoperable; they are in fact locked in their databases and proprietary software, not suitable to be linked in the Semantic Web.

A way to make this data interoperable is to develop a MAP in order to be able to publish this data available in the LOD ecosystem, and also to publish new data that will be created and modeled based on this MAP. To create a common data model for EP is not simple since the existent data models are based on conceptualizations and terminology belonging to their own poetical traditions and each tradition has developed an idiosyncratic analytical terminology in a different and independent way for years. The result of this uncoordinated evolution is a set of varied terminologies to explain analogous metrical phenomena through the different poetic systems whose correspondences have been hardly studied – see examples in González-Blanco & Rodríguez (2014a and b). This work has to be done by domain experts before the modeling actually starts. On the other hand, the development of a MAP is a complex task though it is imperative to follow a method for this development. The last years Curado Malta & Baptista (2012, 2013a, 2013b) have been studying the development of MAP's in a Design Science Research (DSR) methodological process in order to define a method for the development of MAPs (see Curado Malta (2014)). The output of this DSR process was a first version of a method for the development of Metadata Application Profiles (Me4MAP) (paper to be published). The DSR process is now in the validation phase of the Relevance Cycle to validate Me4MAP (for more information and detail on DSR see Hevner (2007)). The development of this MAP for poetry will follow the guidelines of Me4MAP and this development will be used to do the validation of Me4MAP.

The final goal of the POSTDATA project is: i) to be able to publish all the data locked in the WIS, in LOD, where any agent interested will be able to build applications over the data in order to serve final users; ii) to build a Web platform where: a) researchers, students and other final users interested in EP will be able to access poems (and their analyses) of all databases; b) researchers, students and other final users will be able to upload poems, the digitalized images of manuscripts, and fill in the information concerning the analysis of the poem, collaboratively contributing to a LOD dataset of poetry.

Aknowledgements

The work presented in this abstract 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.

References

- Curado Malta, M. (2014). Contributo metodológico para o desenvolvimento de perfis de aplicação no contexto da Web Semântica. University of Minho. Retrieved from <http://hdl.handle.net/1822/30262>
- Curado Malta, M., & Baptista, A. A. (2012). State of the Art on Methodologies for the Development of a Metadata Application Profile. *Metadata and Semantics Research*, 343(July), 61–73. doi:10.1007/978-3-642-35233-1_6
- Curado Malta, M., & Baptista, A. A. (2013a). A method for the development of Dublin Core Application Profiles (Me4DCAP V0.2): detailed description. In M. Foulonneau & K. Eckert (Eds.), *Proc. Int'l Conf. on Dublin Core and Metadata Applications 2013* (pp. 90–103). Lisbon: Dublin Core Metadata Initiative. Retrieved from <http://dcevents.dublincore.org/IntConf/dc-2013/paper/view/178/81>
- Curado Malta, M., & Baptista, A. A. (2013b). Me4DCAP V0. 1: A method for the development of Dublin Core Application Profiles. In *Information Services and Use* (Vol. 33, pp. 161–171). Dublin Core Metadata Initiative. doi:10.3233/ISU-130706
- González-Blanco, E., & Rodríguez, J. L. (2014a). ReMetCa: A Proposal for Integrating RDBMS and TEI-Verse. *Journal of the Text Encoding Initiative*, (8).
- González-Blanco, E., & Seláf, L. (2014b). Megarep: A comprehensive research tool in medieval and renaissance poetic and metrical repertoires. In *Humanitats a la xarxa: món medieval / Humanities on the web: the medieval world*, eds. L. Soriano - M. Coderch - H. Rovira - G. Sabaté - X. Espluga. Oxford, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Wien: Peter Lang, 321-332.
- Hevner, A. R. (2007). The three cycle view of design science research. *Scandinavian Journal of Information Systems*, 19(2), 87.