# Towards an Ontology for European Poetry

María Luisa Díez Platas[1], Helena Bermúdez[1], Salvador Ros[1], Elena González-Blanco[2], Javier de la Rosa[1], Álvaro Pérez[1], Luciana Ayciriex[1]

[1]Dep. Sistemas de Comunicación y Control-Laboratorio de Innovación en Humanidades Digitales (LINHD), (UNED).

[2]Coverwallet

## Introduction

An ontology is a formal and explicit specification of a shared conceptualization (Studer, Benjamins, & Fensel, 1998). From the point of view of ontological engineering, this formalization defines an abstract model that structures the knowledge of a specific domain (agreed upon by experts and interpretable by a machine) from its concepts and relationships. With ontologies, shared and distributed knowledge can be managed in such a way as to allow the integration of information from different data sets (Davies, Fensel, & Harmelen, 2003). An expression of this feature is the DBpedia, which contains multiple links to other data sets such as Freebase and Geonames, among others.

In this paper, we present the case of an ontology for the domain of European poetry. Given the fragmentary access to poetic resources (see González-Blanco & Selaf, 2014), an ontology, and more specifically an ontology network about European Poetry (EP), proposes a coherent and unambiguous semantic conceptualization of the domain. A shared conceptualization enables the integration of disparate poetic data sources belonging to different poetic traditions, overcoming well-known handicaps and idiosyncratic differences, such as the distinct organization of scholarship, or the heterogeneous bibliography.

## Methodology

The starting point of the ontology construction was the analysis of different databases with contents related to one or more EP traditions.[1] As a result of this process, we built a data model comprising 41 entities, 490 attributes, and 407 relations.

In order to reduce some of the inherent complexity of the model, we decided to identify areas that could be organized within themselves. This allowed us to follow a modular design while ensuring higher cohesion and lower coupling of its modules. To this end, we extracted the ontological modules that represented each of these areas and that could form part of the ontologies of a network. Our goals were to make its structure

[1] See Bermúdez et al. (2017) for more details about the resources analyzed and the methodology followed.

comprehensible, to increase its capacity to be extended, and to allow other ontologies to selectively import modules from it. These modules are Work Module, Transmission Module, Structural Elements Module, Literary Analysis Module, Prosodic Elements Module, Music Module and Additional Elements. All of them cover the different aspects referring to the poetic work, its representation, and the features to be analyzed. In addition, an autonomous upper ontology module (UOM) has been created that reuses other ontologies or modules of a similar nature, in which concepts and relationships independent of the domain that are defined, but which are necessary for the representation of complete knowledge such as information referring to Person, Event, Place y Role. In this sense, a module has also been defined to describe the treatment of dates in the ontology of POSTDATA.

For the definition of each module, an analysis of concepts and relationships has been carried out 1) to refine the semantics, 2) to solve design problems, and 3) to take the reutilization of existing ontologies into account.

Subsumption relations have been identified in order to refine the semantics and to help building hierarchies, such as prosodic information patterns.

The constructions identified to support use cases and structural relations are defined by means of design solutions such as ontological design patterns that allow to represent the mereological relations through the "part of" pattern,[2] or through the use of small ontologies such as “The Ordered List Ontology”.[3]

The ontologies identified, in general, have not been completely reutilized. We have selected ontologies defined for analogous domains or related to the POSTDATA domain, or that cover semantic aspects of some of the modules. The ontologies and vocabularies used are, among others, Functional Requirements for Bibliographic Records (FRBRoo),[4] Schema.org vocabulary,[5] Europeana Data Model (edm),[6] CIDOC Conceptual Reference Model[7]. Then, the ontologies Friend of a Friend (foaf)[8] and Time[9] were selected for general purposes.

Much emphasis has been placed on the definition of restrictions that allow the establishment of requirements (mandatory or optional). Moreover, they express the cardinality of the object properties and data properties.

**Conclusions and future work**

[2] http://www.ontologydesignpatterns.org/cp/owl/partof.owl
[3] http://smiy.sourceforge.net/olo/spec/orderedlistontology.html

[4] http://www.cidoc-crm.org/frbroo/home-0
[5] https://schema.org/
[6] https://pro.europeana.eu/resources/standardization-tools/edm-documentation
[7] http://www.cidoc-crm.org/
[8] http://www.foaf-project.org/
[9] https://www.w3.org/TR/owl-time/

A network of ontologies has been built which models the knowledge of European poetry. Each module covers a specific field of the domain. In addition, the use of ontologies and ontological design patterns enabled a good degree of alignment
Our proposal will support researchers in the organization of their information so that computational methods could be applied over their data. It will provide a means for the EP community of practice to easily exchange information, making their data shareable. In addition, it will enable the creation of applications that could retrieve, query, and manipulate these data.
POSTDATA project aims at becoming a reference in terms of philological digital humanities standardization, interoperability by using linked open data(Gonzalez-Blanco et al., 2018). This proposal is open to improvement and dialogue in the community and allows upgrades based on new knowledge derived from the great wealth and possibilities of the domain of poetry.

**Acknowledgements**
Research for this paper has been made possible thanks to the Starting Grant research project Poetry Standardization and Linked Open Data: POSTDATA (ERC-2015-STG-679528) obtained by Elena González-Blanco. This project is funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme.

**Bibliographic references**

Bermúdez-Sabel, Helena, Mariana Curado Malta, and Elena Gonzalez-Blanco. 2017. 'Towards Interoperability in the European Poetry Community: The Standardization of Philological Concepts'. In *Language, Data, and Knowledge: First International Conference, LDK 2017, Galway, Ireland, June 19-20, 2017, Proceedings*, edited by Jorge Gracia, Francis Bond, John P. McCrae, Paul Buitelaar, Christian Chiarcos, and Sebastian Hellmann, 156–65. Cham: Springer International Publishing. (Gonzalez-Blanco et al., 2018).

Davies, J., Fensel, D., & Harmelen, F. van. (2003). *Towards the Semantic Web: Ontology-driven Knowledge Management*. John Wiley & Sons.
Gonzalez-Blanco, E., Ros Muñoz, S., Ruiz Fabo, P., Diez Platas, M. L., Bermudez, H., Caminero, A., … Ayciriex, L. (2018, diciembre 11). Poetry and Digital Humanities making interoperability possible in a divided world of digital poetry: POSTDATA project. Recuperado de https://zenodo.org/record/2203807#.XA_RYWhKjIU

Gonzalez-Blanco, E., Ros Muñoz, S., Ruiz Fabo, P., Diez Platas, M. L., Bermudez, H., Caminero, A., … Ayciriex, L. Poetry and Digital Humanities making interoperability possible in a divided world of digital poetry: POSTDATA project. (2018)

González-Blanco, E., Selaf,L.:Megarep: A comprehensive research tool in medieval and renaissance poetic and metrical repertoires. In: Humanitats a la xarxa: ḿon

medieval/Humanities on the Web: The Medieval World, pp. 321–332. Peter Lang (2014)

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1), 161-197. https://doi.org/10.1016/S0169-023X(97)00056-6