Towards a common model for European poetry: Challenges and solutions

Helena Bermúdez-Sabel
Université de Neuchâtel, Switzerland

María Luisa Díez Platas
Universidad Internacional de La Rioja, Spain

Salvador Ros
Universidad Nacional de Educación a Distancia, Spain

Elena González-Blanco
Coverwallet, Spain

Abstract

This paper stems from the analysis of multiple poetic resources that were available online, as well as the results of methodological discussions with scholars of European Literature. The goal was to retrieve the informational needs of all these different sources in order to build a common data model for European Poetry (EP). Thus, by implementing a reverse engineering method, we have created the Domain Model for EP, which is an important breakthrough for making existent poetry resources interoperable. The lack of a uniform academic approach to analyse and classify poetic manifestations, the divergence of theories when comparing poetry schools from different languages and periods is some of the factors that hinder the modelling process. In this paper, we will present some of the challenges we encountered while conceptualizing the information relevant to poetic analysis and how we have worked around them. Some elements of the ontology will be presented to illustrate our modelling strategies.

Correspondence:
Helena Bermúdez-Sabel, Institut des sciences du langage, Université de Neuchâtel, Rue Pierre-à-Mazel 7, 2000 Neuchâtel, Switzerland.

E-mail:
helena.bermudez@unine.ch

1 Rationale

During the history of European Literature, there have been different cultural centres that irradiated their influence on other cultural systems. Most traditions, due to historic and socio-political reasons, have leaned at some point in their history on other literary models (Even-Zohar, 1978, p. 48). Thus, the relations between the different literary traditions are many and heterogeneous. This poses some difficulties for literary research, since these relations are not always easy to trace. An additional drawback is that this situation demands from researchers to closely know traditions and languages other than the ones of their specialization and the accumulation of all that knowledge is not always humanly possible. The digital environment is our best ally to find resources in other languages or periods that can complement our research, but mere availability is not enough.

Digital Scholarship in the Humanities, Vol. 37. No. 4, 2022.
https://doi.org/10.1093/llc/fqab106 Advance Access published on 14 December 2021

Although online poetic resources exist, their access is fragmentary: there is no way all relevant information for an effective comparison can be retrieved at once (González-Blanco and Selaf, 2014). Researchers need to look for multiple resources and then, for each one of them, carry out different queries in order to retrieve the required information.

To work around this problem, the project 'Poetry Standardization and Linked Open Data' (POSTDATA)[1] has a proposal that focuses on two key concepts: standardization and interoperability, according to the linked open data paradigm (LOD) (Gonzalez-Blanco *et al.*, 2018).

This paper addresses the challenges we faced while creating a common data model that has later been formalized as a network of ontologies (POSTDATA Project, 2020a). After presenting some brief notes about the objectives of the project (Section 2), this paper will concentrate on modelling issues that arose during the elicitation of the informational needs (Section 3). Each issue is illustrated with specific examples to showcase our modelling strategies, that will be further discussed with the presentation of the ontology (Section 4).

2 Contextualization

Linked open data must endorse a semantic data model before being published. This underlying semantic model can be formally specified through an ontology. With ontologies, shared and distributed knowledge can be managed in a way that allows the integration of information from different data sets (Davies *et al.*, 2003). Considering that one of the main aims of POSTDATA is to provide a means to publish European poetry (EP) data as LOD, this project is building a network of ontologies for EP.

Among the domains close to our field, Information Sciences and Linguistics are the ones that have the most linked data resources.[2] Such resources are more scarce for literary studies, although important projects exist (Gerber and Hunter, 2009; Jewell, 2010; Ciotti *et al.*, 2016; Bartalesi and Meghini, 2017; Daquino *et al.*, 2019; Giovannetti, 2020; Ruiz Fabo *et al.*, 2020).

The starting point of the ontology construction was the development of a domain model for EP (DM-EP) (Curado Malta and Bermúdez-Sabel, 2018). This model was done through the analysis of different databases with contents related to one or more EP traditions in order to represent the informational needs of the community of practice of EP.[3]

Our goal is to enhance interoperability between existing poetic repertoires and to facilitate the creation of new resources (Bermúdez Sabel and González-Blanco, 2019). With such an ambitious objective in mind, we must be very exhaustive when eliciting the data needs of our target.

Our sources to draw out the informational needs of the EP community were, on the one hand, a representative sample of poetic repertoires and, on the other hand, a survey that allowed us to directly consult the EP community.[4] In addition, there were different validation processes through which we implemented various tests and also received the direct input of experts in EP in order to refine the model (Curado Malta *et al.*, 2018). The conceptual model obtained presents a great level of complexity due to its exhaustivity and the different domains that it covers.

Considering our heterogeneous inputs, we are dealing with miscellaneous sources of information that incorporate data from multiple languages and cultures. This matter complicates the process of modeling. In the following section, we will present some of the issues we encountered while developing the common data model for EP and how we worked around them.

3 Modelling Challenges

The creation of a domain data model that covers all required concepts to analyse any European poem causes some difficulties. Each of the following subsections exhibit a problematic feature that is illustrated with at least one example and with our solution. The decisions taken during this phase were focused on terminological standardization. We selected terms related to the structural parts of the poem and to literary devices that were representative. For each term, we provided a precise, straightforward description so as to facilitate the transparency of the term.

3.1 Multilingualism

The most obvious problem we ran into arises from working with a multilingual reality. The modellers must analyse online resources in languages with which they are not familiar, and they need to verify that each concept represented in that project has a counterpart in the data model, that is, in the DM-EP.[5] This knowledge gap is covered with either the documentation translated to English by the project being analysed, or with additional bibliography. Nevertheless, the direct contact with the people in charge of that resource is at times inevitable, but the response and willingness to collaborate is, for the most part, extremely positive. In this sense, creating the means to get the direct input from the community is crucial.

In addition, the validation of our work from the early stages is critical. In this way, we can verify whether the elicitation of the informational needs of each online database is being done properly. The first validation of the data model took place during a workshop in which representatives of the projects analysed until that point participated. We also designed a form in which any person could formalize a poetic resource of their choice following our data model. For more details about the validation steps, see Curado Malta *et al.* (2018).

Besides properly representing a multilingual reality, we need to create a resource that is useful and usable by the whole community. Thus, multilingualism is a challenge during the elicitation of concepts, but also during the implementation process. The implementation of the data model needs to establish terminological relations between different languages which poses additional challenges since a one-on-one correspondence is not always possible.[6] Therefore, a more complex formalization is required. In this sense, the linguistics community has designed the means to model the relations between expressions from different languages that are able to describe any semantic nuances. See, for example, the Linguistic Meta-Model (LMM) by Picca *et al.* (2008), specifically designed to address the simultaneous representation of 'multiple linguistic knowledge sources, allowing interoperability and an improved comprehension and exploitation of knowledge' (Picca *et al.*, 2008, pp. 2416–7).[7] A future integration of LMM in the ontology implementation opens the door to go in depth in the meaning of the entities and properties whose name varies from language to language, independently from its functionality in the model. In this manner, the semantics would be reinforced by going beyond a simple textual description of the entity/property that was chosen for the normalized representation.

3.2 Polysemy

Occasionally, the difficulties are due to ambiguities in one language. For instance, we find that many European languages have a term derived from Latin *versus* to describe the poetic line. In English, however, the term *verse* can describe either the line of poetry, a bigger division like the stanza or the whole poetic composition (American Heritage Dictionaries, 2015b). In this case, during the elicitation process, it was clear that those concepts needed to be independent entities in our model. Thus, the DM-EP formalized these different concepts as the entities *Line*, *Stanza* (for any group of lines smaller than the poem, including any group that may be or not a metrical unit; see Fig. 1),[8] and *Redaction/PoeticWork*.[9] We need to discern the referent whenever an ambiguous term is used to evaluate if additional properties need to be added to the model. In addition, documentation needs to be thorough so, if an existing resource uses polysemic terms, they can map the concepts correctly.

It is not uncommon to find the same term in different technical vocabularies but with distinct meanings. For example, the term *dieresis* in syllabic verse traditions describes the separate pronunciation into two syllables of two sounds which usually form one syllable (Larousse, s.d.). However, in quantitative verse, a *dieresis* expresses the pause that occurs when the end of a foot coincides with the end of a word (American Heritage Dictionaries, 2015a). The DM-EP needs to include both concepts in an unequivocal way. In this case, we selected the term *diaeresis* to describe the type of pause and used *hiatus* to express the separation into two syllables, taking the term from its counterpart concept in Linguistics.[10] More specifically, we have the boolean attribute *hasHiatus* in the entity *Syllable* (see Fig. 3) and the boolean attribute *hasDiaeresis* in the entity *LinePattern* (Fig. 2).

Prosody, the study of the patterns of rhythm and sound used in poetry, is very closely related to Linguistics but the concepts of both areas are not

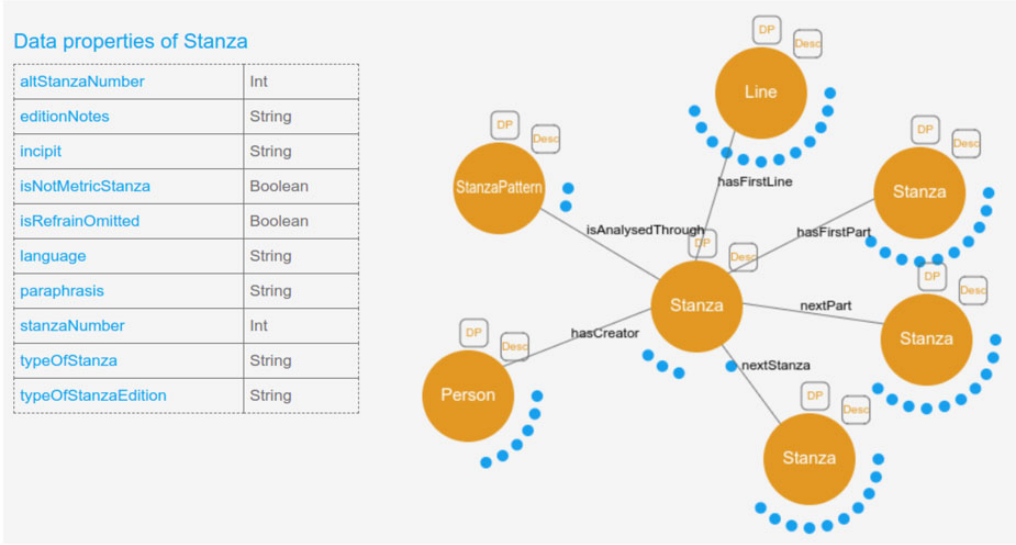

Fig. 1. The entity *Stanza* and some of its properties

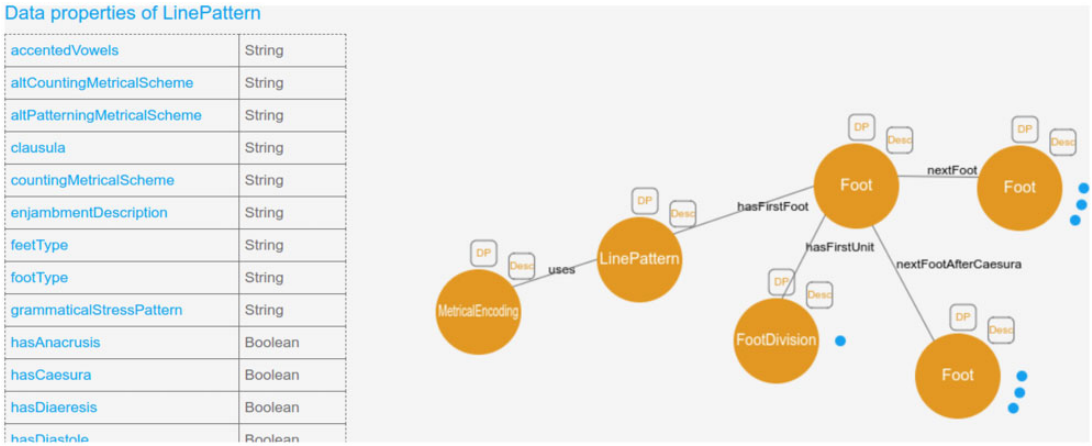

Fig. 2. The entity *LinePattern* and some its properties

always perfect synonyms. However, both approaches coexist in poetic analyses. For that reason, the DM-EP needs to be flexible while considering various theoretical backgrounds without ever prioritizing one over the others. A good example of this situation is the concept of *syllable*. A syllable in natural language does not always coincide with a prosodic syllable, especially in syllabic or accentual-syllabic verse. Due to poetic license, there are literary devices that a poet may explore to alter the number of syllables, so they fit the desired rhythm. Section 3.4 continues the discussion about the issues concerning the concept of syllable, but the point we want to make regarding polysemy is that the DM-EP needs to be unambiguous, be precise in the definitions, and employ terminology as neutral as possible.

3.3 Synonymy and Quasi-synonymy

Literature scholarship is a field with thousands of years, which means that some of the concepts we are analysing have been defined for many centuries and from different perspectives. However, no school of

Data properties of Syllable

coda	String
content	String
hasHiatus	Boolean
hasSynizesis	Boolean
isDialepha	Boolean
isStressed	Boolean
nucleusType	String
onset	String
positionInWord	String
weight	String

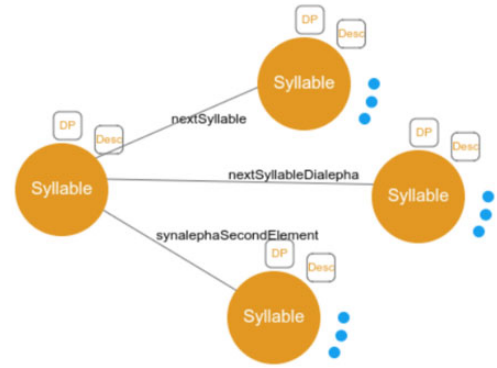

Fig. 3. The entity *Syllable* and its properties

thought or theory can be prioritized. We may find different terms for the same concept in which the use of one term over the other is related to philological schools (synonimy). In these cases, we need to evaluate which one is the less aligned (Bermúdez-Sabel *et al.*, 2017).

The cases of quasi-synonymy are slightly different. We may find two expressions that refer to two similar concepts. The semantic nuances that define this difference could be relevant or not. Thus, during the elicitation of the information requirements we need to evaluate if the two concepts can be merged into a more generic one or if the two items need to be formalized individually. Let us consider, for instance, the terms *syneresis* and *synizesis*. The first one refers to the contraction of two vowels (or two syllables) into one, most commonly to form a diphthong. On the other hand, the synizesis is the combination of two vowels that do not form a diphthong into one syllable. Therefore, the slight difference of meaning could be considered irrelevant in prosodic studies, so we decided to combine both concepts under the same property of the entity *Syllable*: *hasSynizesis* (see Fig. 3).[11]

3.4 Semantics versus Interoperability

Like in any other process of semantic modelling, there is some tension between interoperability and semantics.

For instance, poetry of the Western world is usually categorized as either qualitative or quantitative metre (Aroui and Arleo, 2009, pp. 11–2). Thus, metre may depend on the length of syllables and their distribution, or on the pattern created by stressed syllables coming at regular intervals. In the case of qualitative metre, instead of demanding a fixed pattern of all the stresses, some traditions only care about the position of a certain stressed syllable, like the last one.

However, some of the types of qualitative verse have many attributes that are interoperable with the quantitative ones. Therefore, we decided to make a conceptual division between metrical schemes that depend on patterns and those that are defined by 'counting' elements (such as counting how many syllables are there before the last stressed one). In this manner, little semantics are lost, because other properties make the distinction between qualitative and quantitative. However, with this conceptualization, we enable the comparison between types of metre that, even if they focus on different linguistic properties, have many things in common, such as the use of the foot as a structural unit.

When discussing polysemy, we mentioned that a syllable in natural language does not always coincide with a syllable from a prosodic point of view. Initially, we planned for the class *Syllable* to be valid for both linguistic and prosodic syllables. The user will do the scansion, that is, the division of the line in syllables, using the unit of reference of their preference, but the attributes and relations had to make it possible to inform of all possible phenomena either way. Later, it was decided to select one of the two

conceptualizations of syllable to enhance interoperability. It is more efficient to have only one way of modelling compatible concepts since we avoid the implementation of computational operations to create equivalences between the data available. The conceptualization of the class *Syllable* matches the linguistic definition of syllable. The reason we chose the linguistic definition over the prosodic one was due, again, to efficiency. In this manner, the user can benefit from the use of Natural Language Processing (NLP) methods to do the scansion. These technologies divide the line into linguistic syllables and then make the proper calculations to look for the literary devices that may affect them, following the same workflow as our ontology—for example, Navarro-Colorado (2017).

3.5 Granularity

Mereological relations proliferate in the domain under investigation, especially when it comes to the structure of texts. We observed that poetic texts could be represented by a set of verses or lines, or by a set of line groups that in turn could be formed by smaller line groups or by lines, and, in all cases, the order must be well defined (Keet *et al.*, 2012). That is to say, a poetic text is formed by an ordered list of lines or by an equally ordered list of stanzas. Likewise, a stanza can be formed by an ordered list of stanzas or by an ordered list of lines (see Fig. 1). At the same time, the line itself may be modelled as a set of tokens (words and punctuation marks) in which order is also important. The word can be broken down as well into ordered syllables.

The decomposition in all these units of analysis responds to a very granular description of a poetic work. However, not every online resource we analysed reached this level of detail. If we want to easily enable the endorsement of our ontology by existing resources, we need to contemplate that the analytical description might be attached to any structural element. This approach makes the model more complex because it multiplies the number of properties, most significantly, the number of object properties. However, we argue that it facilitates its implementation because existing resources are faced with a flexible model that allows them to easily map their logic models to our ontology.

The challenges and solutions presented in this section are implemented in the ontology through a combination of methods, namely through concept definition and hierarchization and the creation of controlled vocabularies.

4 Ontology Description

Section 3 describes the challenges we faced when creating the DM-EP. This model covers both the descriptive and bibliographic aspects of poetic works, including details about textual transmission, as well as aspects related to prosody, literary and rhetorical analyses, structural elements, and their relations with music. The result is a semantic model with 40 entities, 494 attributes, and 409 relationships. This domain model was the starting point for the development of the ontology.

Due to the complexity of the poetry domain, we decided to build a network of ontologies (POSTDATA Project, 2020a). This network consists of seven ontologies that cover the different dimensions related to poetic works (Fig. 4). A core ontology includes elements related to the description of the works, their creators, and related agents. Other ontologies are more specific and they conceptualize different aspects concerning the transmission of the work or other elements related to their written and musical manifestations. Outstanding ontologies are those designed to formalize the analytical dimension: structural analysis, prosodic analysis, and literary and rhetorical analysis. In addition, an ontology of dates has been developed to include the different formats of dating a literary work (Fig. 5).

The development process was carried out through an iterative-incremental model. Each ontology was built with the premise of reusing existing ontologies by aligning vocabularies and properties to facilitate its development, improving the semantic understanding of entities, and facilitating interoperability. In addition, thirty-three controlled vocabularies have been defined to represent the values of several data properties.

4.1 Reused Ontologies

We explored a set of high-level ontologies that could potentially include some of the classes and properties we needed to conceptualize. Even if we hardly ever encountered a perfect equivalent between these

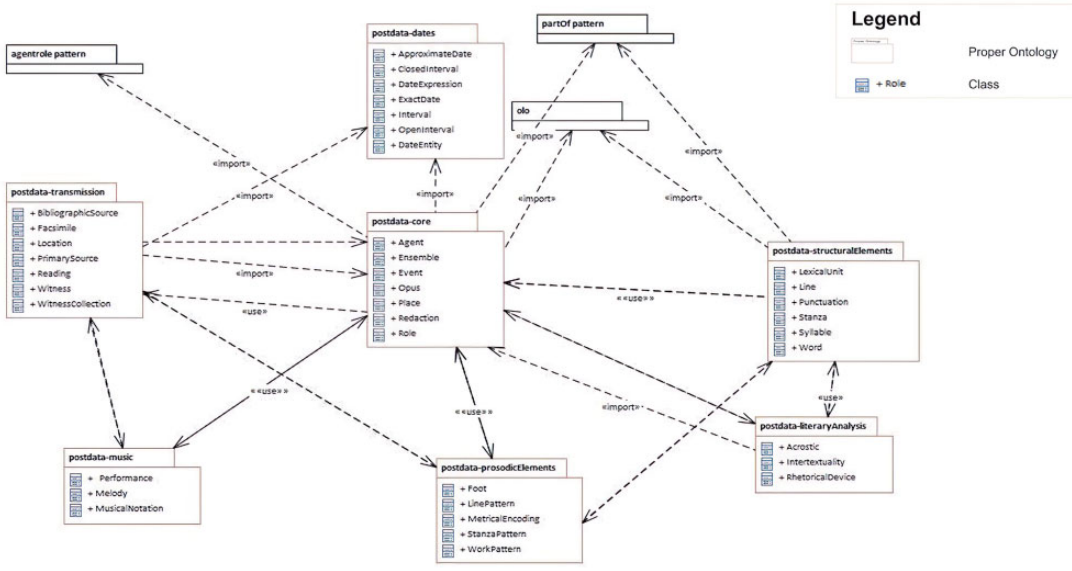

Fig. 4. Network of ontologies

ontologies and our target ontology, some classes of our ontology were defined as specializations of certain classes of those high-level ontologies. Among these ontologies, we used FOAF (Brickley and Miller, 2014), Dublin Core Vocabulary (DCMI Usage Board, 2020), and Schema.[12] The classes of our ontology were also aligned with the Europeana one (Isaac, 2013). This provides the means to facilitate data collection initiatives implemented by Europeana. Regarding domain ontologies, the one reused the most was the FRBRoo ontology[13] because it represents several of the concepts we need to bibliographically describe a poetic work and its manifestations. The FRBRoo classes were thus specialized to express the desired semantics.

4.2 Some Design Challenges

Mereological relations are very common in the Literature domain, in particular when defining the structural elements of the texts. In Section 3.5, we briefly discussed the variability that we might find when breaking down a poem in smaller units. To model this, we used the Ordered List Ontology (Abdallah and Ferris, 2010), because it provides the necessary entities and properties to describe ordered lists as a semantic graph. Therefore, stanzas, lines, words, and syllables have been modelled as subclasses of the *olo: Slot* class since its inclusion of an index enables the representation of the elements in an orderly way.

For cases in which order was not relevant (like when a particular poem belongs to an unordered collection), we use the ontological design pattern *PartOf*.[14]

Another ontological design pattern, *AgentRole*,[15] was implemented to represent the different agents that participate in the creation of a poetic work and its transmission. A very important role is of course the creator of a work, but many other agents might be involved, such as editors, copyists, or composers of the musical arrangement. By implementing the pattern *AgentRole* and extending its classes, we can model phenomena like wrong authorship attributions or heteronyms. For example, a text can mention a specific person as the author, but the real creator is somebody else. Our ontology enables the modelling of these complex phenomena that could be hardly represented with a single property—like for instance, the property *creator*, from Dublin Core Metadata Terms (DCMI Usage Board, 2020).

It is common for the dating of a poetic work or its manifestations to be uncertain or to be represented as

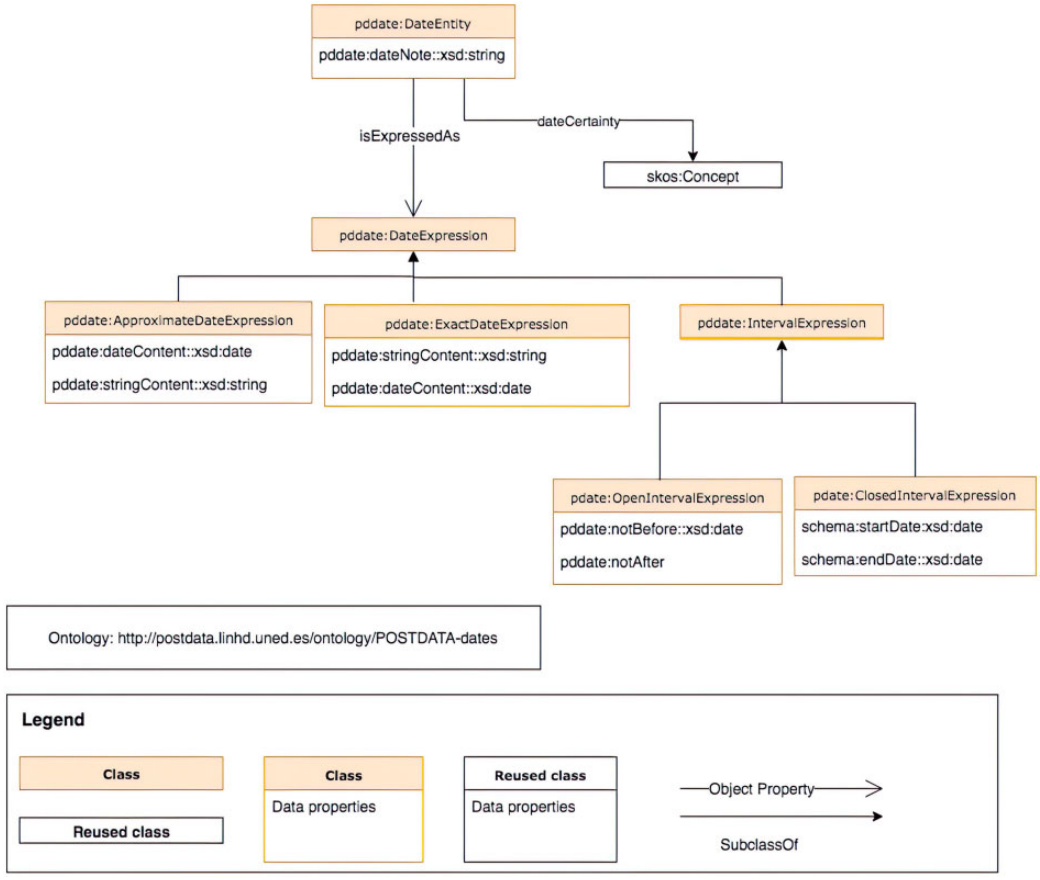

Fig. 5. *Postdata-dates* ontology

a range (with more or less certainty). For this reason, we created an independent ontology (POSTDATA Project, 2020c) to be reused by other domains that deal with similar dating problems (see Fig. 5).

The existence of a large number of terms to express analytical categories related to the poetic domain requires their organization as a finite list of concepts and normalized terms. Thus, we have created a number of controlled vocabularies to define the range of several properties of the ontology (POSTDATA Project, 2019). These controlled vocabularies were implemented with SKOS (Simple knowledge organization system) a common data model for sharing and linking knowledge organization systems via the Web (Zaytseva and Ďurčo, 2020). The implementation of linguistic-specific vocabularies like OntoLex-Lemon would enhance the usability of these controlled-vocabularies by establishing semantic relations between terms in different languages (Declerck *et al.*, 2010).

4.3 The Ontologies

4.3.1 *Postdata-core* ontology

This ontology describes the main metadata of a poetic representation (see Fig. 6). It provides information about a poetic work and its manifestations. A poem may have different versions; therefore, we separate the abstract concept (*pdcore: poeticWork*) from its materializations (*pdcore: Redaction*). In addition, a work can belong to a collection (*pdcore: Ensemble*). These classes have been defined as the specialization of FRBRoo classes. The complete ontology consists of 44 classes, 158 data and object properties (POSTDATA Project, 2020b).

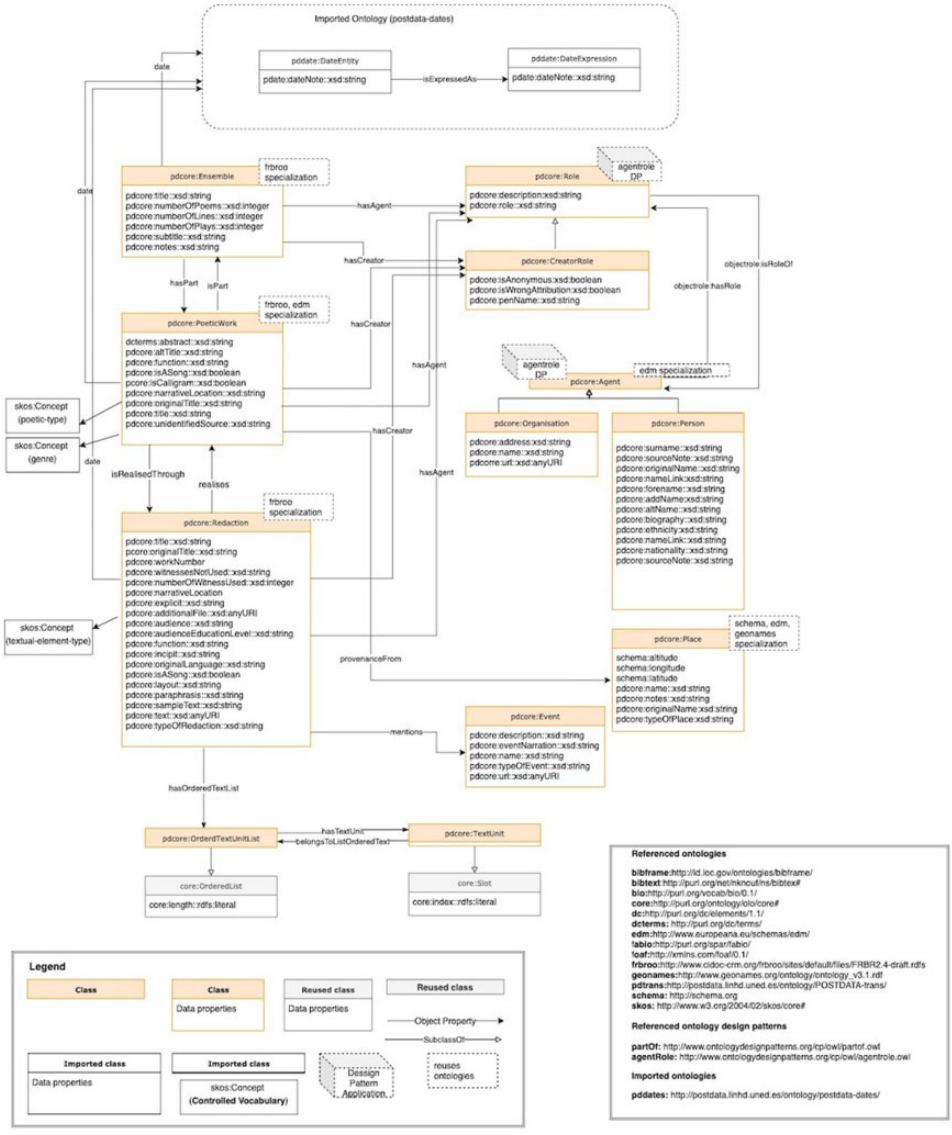

Fig. 6. *Postdata-core* ontology

4.3.2 *Postdata-structuralElements* ontology

The *postdata-structuralElements* ontology contains all the information related to the structural elements of a *pdcore: Redaction*. This ontology illustrates the challenges described in Section 4.2 in regards to the mereological relations. The complete ontology contains eight new classes, twenty-one data properties, and forty object properties (POSTDATA Project, 2020f).

4.3.3 *Postdata-prosodicElements* ontology and *postdata-literaryAnalysis*

These two ontologies contain the necessary elements to carry out a prosodic and a rhetorical analysis of a poetic work.

The prosodic analysis features the description of the metrical patterns of a poem which can be defined at three different levels: poem, stanza, or verse line. Thus, we have created three different classes:

pdprosodic: LinePattern, *pdprosodic: StanzaPattern*, and *pdprosodic: WorkPattern*.

The ontology *postdata-prosodicElements* needs to enable the representation of the poetic licenses that affect the number of syllables, such as diaeresis, synaeresis, synalepha, or hiatus. These figures involve the alteration of the pronunciation (sometimes also the writing) of a word without changing its meaning. These devices are modelled thanks to the class *pdprosodic: Metaplasm*. In order to identify the type of metaplasm, the range of the property *pdprosodic: typeOfMetaplasm* was defined as a controlled vocabulary.

The *postdata-prosodicElements* contains ten classes, fifty-two data properties, and forty object properties (POSTDATA Project, 2020e), while the *postdata-literaryAnalysis* consists of five classes, seven data properties, and twenty object properties (POSTDATA Project, 2020d).

5 Conclusions and Future Work

The conversion of literary collections into interoperable machine-readable repositories will open the door to pose new research questions and to perform comparative philological analysis between heterogeneous poetic corpora. However, the development of a data model that expects to serve the community of practice of EP in the LOD ecosystem entails a great complexity. On the one hand, there is no uniform academic approach to analyse and classify poetic manifestations, and the divergence of theories is even bigger when comparing poetry schools from different languages and periods. Therefore, a thorough but complex philological standardization is needed. On the other hand, the type of final user that will consume that data is very diverse. Moreover, the applications that might be built with these data are many and very heterogeneous. These factors complicate the elicitation of the functional and no functional requirements, thus arising very interesting issues during the modelling process.

In this paper, we presented some of the problems we faced while building the common semantic data model for EP that was used as the conceptual source for the construction of a network of ontologies. This domain model contains descriptive features of the poetic work, including details about textual transmission and other bibliographic references, as well as aspects related to the prosodic and literary analysis together with its relationships with the musical domain. Thus, it is not a model that focuses on the definition of poetic concepts, but it covers every aspect needed to do a holistic description of any poetic resource. The result is a data model with 40 entities, 494 attributes, and 409 relationships that was then implemented as a network of seven ontologies.

This network of ontologies has not been completely published and the paper succinctly describes the ones that are available. Besides finishing the publication of all the elements of this network, some enhancements of this network could be envisioned. For example, a multilingual lexicographical ontology of poetic terms could be developed to further increase the usability and interoperability of the POSTDATA ontology.

Funding

This project was supported by the European Research Council [ERC-2015-STG-679528].

References

Abdallah, S. and Ferris, B. (2010). *The Ordered List Ontology Specification*. http://smiy.sourceforge.net/olo/spec/orderedlistontology.html (accessed 10 September 2021).

Abgaz, Y. (2020). Using OntoLex-Lemon for representing and interlinking lexicographic collections of Bavarian dialects. In *Proceedings of the 7th Workshop on Linked Data in Linguistics (LDL-2020)*. Marseille, France: European Language Resources Association, pp. 61–9. https://www.aclweb.org/anthology/2020.ldl-1.9 (accessed 22 November 2020).

American Heritage Dictionaries. (2015a). *Dieresis*. American Heritage Dictionary of the English Language.

American Heritage Dictionaries. (2015b). *Verse*. American Heritage Dictionary of the English Language.

Aroui, J.-L. and Arleo, A. (2009). *Towards a Typology of Poetic Forms: From Language to Metrics and Beyond*. John Benjamins Publishing.

Bartalesi, V. and Meghini, C. (2017). Using an ontology for representing the knowledge on literary texts: The Dante Alighieri case study. *Semantic Web. 10.3233/SW-***150198**.

Bermúdez Sabel, H. and González-Blanco, E. (2019). Prefacio. *Revista de Poética Medieval*, **33**: 11–23. doi: 10.37536/RPM.2019.33.0.79936.

Bermúdez-Sabel, H., Curado Malta, M., and Gonzalez-Blanco, E. (2017). Towards interoperability in the European Poetry community: The standardization of philological concepts. In Gracia, J., McCrae, J. P., Buitelaar, P., and Bond, F. (eds), *Language, Data, and Knowledge: First International Conference, LDK 2017, Galway, Ireland, June 19–20, 2017, Proceedings*. Cham: Springer International Publishing, pp. 156–65. doi: 10.1007/978-3-319-59888-8_14.

Bosque-Gil, J., Gracia, J., Montiel-Ponsoda E., and Gómez-Pérez A. (2018). Models to represent linguistic linked data. *Natural Language Engineering*, **24**(6): 811–59. doi:10.1017/S1351324918000347.

Brickley, D. and Miller, L. (2014). *FOAF Vocabulary Specification 0.99*. http://xmlns.com/foaf/spec/20140114.html (accessed 1 September 2021).

Cimiano, P., Chiarcos, C., McCrae, J. P., and Gracia, J. (2020a). Linguistic linked data in digital humanities. In Cimiano, P., McCrae, J. P., Chiarcos, C., and Gracia, J. (eds), *Linguistic Linked Data: Representation, Generation and Applications*. Cham: Springer International Publishing, pp. 229–62. doi:10.1007/978-3-030-30225-2_13.

Cimiano, P, Chiarcos, C., McCrae, J. P., and Gracia, J. (2020b). *Linguistic Linked Data: Representation, Generation and Applications*. 1st edn. Springer Nature Switzerland AG.

Ciotti, F., Peroni, S., Tomasi, F., and Vitali, F. (2016). An OWL 2 formal ontology for the text encoding initiative. In Eder, M. and Rybicki, J. (eds), *Digital Humanities 2016: Conference Abstracts. Kraków*, pp. 151–3.

Curado Malta, M., Bermúdez-Sabel, H., Baptista, A. A., and Gonzalez-Blanco, E. (2018). Validation of a metadata application profile domain model. In *International Conference on Dublin Core and Metadata Applications*. pp. 65–75. http://dcevents.dublincore.org/IntConf/dc-2018/paper/view/555/675.

Curado Malta, M. and Bermúdez-Sabel, H. (2018). *Domain Model for European Poetry v 1.0*. Zenodo. doi: 10.5281/zenodo.1164876.

Daquino, M., Giovannetti, F., and Tomasi, F. (2019). Linked data per le edizioni scientifiche digitali. Il Workflow di pubblicazione dell'edizione semantica del quaderno di appunti di Paolo Bufalini. *Umanistica Digitale*, **7**. doi: 10.6092/issn.2532-8816/9091.

Davies, J., Fensel, D., and van Harmelen, F. (eds) (2003). *Towards the Semantic Web: Ontology-Driven Knowledge Management. 1st edn*. Chichester, UK: Wiley.

DCMI Usage Board. (2020). *DCMI Metadata Terms*. https://www.dublincore.org/specifications/dublin-core/dcmi-terms (accessed 1 September 2021).

Declerck, T., Buitelaar, P., Wunner, T., McCrae, J. P., Montiel-Ponsoda, E., and Cea, G. A. (2010). *Lemon: An Ontology-Lexicon model for the Multilingual Semantic Web*. https://www.w3.org/International/multilingualweb/madrid/slides/declerck.pdf (accessed 1 September 2021).

Declerck, T., Siegel, M., and Racioppa, S. (2019). Using OntoLex-Lemon for representing and interlinking german multiword expressions in OdeNet and MMORPH. In *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*. Florence, Italy: Association for Computational Linguistics, pp. 22–9. doi: 10.18653/v1/W19-5104.

Even-Zohar, I. (1978). *Papers in Historical Poetics*. Tel Aviv: Porter Institute for Poetics and Semiotics.

Gerber, A. and Hunter, J. (2009). (2) A compound object authoring and publishing tool for literary scholars based on the IFLA-FRBR. *International Journal of Digital Curation*, **4**(2): 28–42. doi: 10.2218/ijdc.v4i2.91.

Giovannetti, F. (2020). *LIFT—Linked Data from TEI*. https://github.com/fgiovannetti/lift (accessed 27 December 2020).

Gonzalez-Blanco, E., Ros, S., Ruiz, P., et al. (2018). *Poetry and Digital Humanities Making Interoperability Possible in a Divided World of Digital Poetry: POSTDATA Project*. https://zenodo.org/record/2203807#.XA_RYWhKjIU (accessed 26 April 2019).

González-Blanco, E. and Selaf, L. (2014). 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*. Lausanne, Switzerland: Peter Lang, pp. 321–32.

IFLA. (2009). *Functional Requirements for Bibliographic Records*. http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm (accessed 22 February 2017).

Isaac, A. (ed.) (2013). *Europeana Data Model Primer*. https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf (accessed 1 September 2021).

Jewell, M. O. (2010). Semantic screeplays: Preparing TEI for linked data. In *Digital Humanities*. http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/html/ab-878.html.

Keet, C. M., Fernández-Reyes, F. C., and Morales-González, A. (2012). Representing mereotopological relations in OWL ontologies with OntoPartS. In Aroyo, L., Traverso, P., Ciravegna, F., et al. (eds), *The Semantic Web: Research and Applications. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, pp. 240–54. doi: 10.1007/978-3-642-30284-8_23.

Larousse. (s.d.). *Diérèse*. Dictionnaire de Français, Dictionnaire En Ligne. http://www.larousse.fr/dictionnaires/francais (accessed 2 January 2020).

McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. (2017). The OntoLex-Lemon model: Development and applications. In *Electronic Lexicography in the 21st Century: Proceedings of ELex 2017 Conference. Lexical Computing*, pp. 587–97.

Miles, A. and Bechhofer, S. (eds) (2009). *SKOS Simple Knowledge Organization System Reference*. https://www.w3.org/TR/skos-reference/ (accessed 20 February 2020).

Navarro-Colorado, B. (2017). A metrical scansion system for fixed-metre Spanish poetry. *Digital Scholarship in the Humanities*, **33**(2). doi: 10.1093/llc/fqx009.

Picca, D., Gliozzo, A. M., and Gangemi, A. (2008). LMM: An OWL-DL MetaModel to represent heterogeneous lexical knowledge. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. LREC 2008. Marrakech, Morocco: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/pdf/608_paper.pdf (accessed 20 December 2020).

POSTDATA Project. (2019). *Skos-codelists*. https://github.com/linhd-postdata/skos-codelists (accessed 20 September 2021).

POSTDATA Project. (2020a). *Network of Ontologies for European Poetry. POSTDATA*. http://postdata.linhd.uned.es/results/network-of-ontologies/ (accessed 1 December 2020).

POSTDATA Project. (2020b). *Postdata-core ontology*. https://postdata.linhd.uned.es/ontology/postdata-core/ (accessed 1 September 2021).

POSTDATA Project. (2020c). *Postdata-dates ontology*. https://postdata.linhd.uned.es/ontology/postdata-dates/ (accessed 1 September 2021)

POSTDATA Project. (2020d). *Postdata-literaryAnalysis Elements ontology*. https://postdata.linhd.uned.es/ontology/postdata-literaryAnalysisElements/ (accessed 1 September 2021).

POSTDATA Project. (2020e). *Postdata-prosodicElements ontology*. https://postdata.linhd.uned.es/ontology/postdata-prosodicElements/ (accessed 1 September 2021).

POSTDATA Project. (2020f). *Postdata-structural ontology*. https://postdata.linhd.uned.es/ontology/postdata-structuralElements/ (accessed 1 September 2020).

Rubiera, E., Polo, L., Berrueta, D., and El Ghali, A. (2012). TELIX: An RDF-based model for linguistic annotation. In Aroyo, L., Traverso, P., Ciravegna, F., et al. (eds), *The Semantic Web: Research and Applications. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, pp. 195–209. doi: 10.1007/978-3-642-30284-8_20.

Ruiz Fabo, P., Bermudez-Sabel, H., Martinez, C. and Gonzalez-Blanco Garcia, E. (2020). The diachronic Spanish sonnet corpus: TEI and linked open data encoding, data distribution, and metrical findings. *Digital Scholarship in the Humanities (fqaa035). doi:* 10.1093/llc/fqaa035.

Zaytseva, K. and Ďurčo, M. (2020). *Controlled Vocabularies and SKOS*. Version 1.1.0. DARIAH-Campus [Training module]. https://campus.dariah.eu/id/D8d6OrLdpLlGRqBSQDVN0 (accessed 9 October 2021).

Notes

1 Please visit the project's website for more details: <http://postdata.linhd.uned.es/> (accessed 20 February 2020).

2 For a comprehensive overview of the representation of linguistics data in the linked data cloud, see Cimiano *et al.* (2020b). Chapter 13 of this monograph focuses on the implementation of linguistic linked data in Digital Humanities projects, and literary resources are specifically addressed (Cimiano *et al.*, 2020a).

3 In order to build a common model, we have used a reverse engineering approach to extract and to compare all the concepts present in the different resources analysed. For a detailed exposition about how these informational needs were elicited and other methodological aspects, see Bermúdez-Sabel *et al.* (2017).

4 See the map available at <http://postdata.linhd.uned.es/partners/> to see the projects that have collaborated with us (accessed 2 November 2020). In Curado Malta *et al.* (2018), there is more information about all the resources that were analysed and what type of study was done of each one of them.

5 The perfect team would have an expert on every poetic tradition, that is, a scholar for every European language

and literary period. Regretfully, it is hard to find a project in Humanities with that type of resources, therefore, collaborative work is essential.

6 This means that common mapping assertions using SKOS 'skos: closeMatch' or 'skos: exactMatch' would not be appropriate (Miles and Bechhofer, 2009).

7 Other models exist, like the OntoLex-Lemon one; a model whose extended use to describe lexical resources is proof of its usability—see the list available at McCrae *et al.* (2017, p. 591) and more recent applications like Declerck *et al.* (2019) or Abgaz (2020).

8 A boolean attribute of the entity *Stanza*, *isNotMetricStanza*, formalizes the difference between a line group that corresponds to a metric unit and the one that does not.

9 The core of the model is formed by two entities that represent the poetic work. On the one hand, there is an entity that encompasses the poetic creation in an abstract conceptualization, *PoeticWork*. On the other hand, we have any of its 'physical' manifestations, that is, the entity *Redaction*. The *PoeticWork* can be compared with the entity 'Work' in the Functional Requirements for Bibliographic Records model (FRBR) (IFLA, 2009). Depending on the context, *Redaction* may be equivalent to the FRBR entities 'Expression', 'Manifestation', or 'Item'.

10 The reasoning behind prioritizing the term that has a clearer equivalent in Linguistics was to promote the relations between our ontology and linguistic data models. However, as discussed in Bosque-Gil *et al.* (2018), 'Phonetics and phonology remain two areas with relatively low coverage in the LLOD cloud' (Bosque-Gil *et al.*, 2018, p. 17). Indeed, the authors have not found any instatization of a phonetic phenomenon like 'hiatus', although it could, of course, be formalized, with models designed for linguistic annotation like TELIX (Rubiera *et al.*, 2012) since such a phenomenon could be described using feature structures.

11 The property is described as 'It presents the union in pronunciation of two adjacent vowels into one syllable without forming a diphthong'. Therefore, our definition does not make any references to the concept of diphthong.

12 See <https://schema.org/> (accessed 1 October 2021).

13 See <https://www.iflastandards.info/fr/frbr/frbroo.html> (accessed 1 October 2021).

14 See <http://www.ontologydesignpatterns.org/cp/owl/partof.owl> (accessed 1 October 2021).

15 See <http://www.ontologydesignpatterns.org/cp/owl/agentrole.owl> (accessed 1 October 2021).