POSTDATA: Poetry Standardization and Linked Open Data

Deliverable Nº (WP2.3)

Form to validate the Domain Model for European Poetry

Deliverable Fact Sheet	
Deliverable Version:	1.0
Deliverable Nature:	R = Report
Dissemination Level:	PU
Work Package:	WP nº 2
Organisation Responsible:	UNED
Document author/s	Helena Bermúdez Sabel María Luisa Díez Platas

Actual Date of Delivery	20/07/2018
Audience	☐ Public ☒ Restricted ☐ Internal

Version number	01.00.00

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

Status	☐ Draft ☒ WP responsible/s accepted ☐ Project coordinator accepted

Summary

This deliverable concerns the validation of the first stable version of the domain model for European Poetry (DM-EP). This questionnaire is built in such a manner that each of the questions has an equivalent to an entity, property or relation of the DM-EP. Thus, by following the questionnaire, the scholar creates as many instances of the class as they need and they establish the proper relations between those instances. In addition, the attributes are populated as well with the values filled-in by the scholar. And the end of the questionnaire, the scholar will have the opportunity to provide feedback regarding any missing concepts. Besides acting as a validation step, this form can be used to perform the refinement of the model that facilitates the task of generating a standard linked open data model and the creation of datasets modelled accordingly to the DM-EP.

.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

Contents

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

1. Introduction

The first stable version of the Domain Model will be entirely validated by the community. Differently to the other validation processes (see paper DCMI), this one does not require to know the data model itself. Through a user friendly web form, scholars will be able to analyse any poetic resource by answering the questions presented in this questionnaire. The targeted audience are scholars who had analysed poetic resources of European traditions and there is an obvious interest in getting to participate as many representatives as possible of the traditions that were not studied during the definition of the Domain Model.

This questionnaire is built in such a manner that each of the questions has an equivalent to an entity, property or relation of the DM-EP. Thus, by following the questionnaire, the scholar creates as many instances of the class as they need and they establish the proper relations between those instances. In addition, the attributes are populated as well with the values filled-in by the scholar. And the end of the questionnaire, the scholar will have the opportunity to provide feedback regarding any missing concepts or attributes or those that are considered unnecessary or not correctly applied. Besides acting as a validation step, this form can be used for the creation of datasets modelled accordingly to the DM-EP.

1.1 Background

The research community of poetry works with digital repertoires of poetry. A repertoire is a catalogue that gives account of the metrical and rhythmical schemes of either a poetical tradition, a period or school, gathering a corpus of poems that are defined and classified by their main characteristics. These kind of repertoires may sometimes contain the text of the poem and information related to authors, manuscripts, editions, music, and other features, all of them related to the poems (Curado Malta et al. 2016).

These repertoires exist on the Web but are not interoperable González-Blanco and Seláf (2014). They have real data from research projects on poetry and this data has been structured by information modellers that have built these systems without concern with the possibility of interoperability. Since their interest laid in answering the particular research questions of their project, their goal is to just serve the specific needs of the local community. The poetry scientists want now to explore new possibilities; they want to cross or compare data from different traditions that is stored in different silos of information. Also, the possibility to link the data of those silos with other resources present in the LOD ecosystem is seen as a huge opportunity to enrich the data that already exists.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

1.2 Developing the Domain Model

Publishing data as LOD in the Web of Data is a process that must start with a good data modelling. Linked data must endorse a semantic model before being published. Since this data comes from different sources that incorporate multiple contexts within various cultures and languages, this process of modelling becomes very complex. According to Nilsson, Baker, and Johnston (2009), metadata must be modelled as a metadata application profile (MAP) in order to become interoperable. Coyle and Baker (2009) define a MAP as "a generic construct for designing metadata records."

A very important of the MAP is the definition of the Domain Model (DM) a common conceptual model that should represent the informational needs of the European Poetry (EP) community of practice.

The development process of defining the Domain Model was made of two well-defined moments of construction and two well-defined moments of validation (see Figure 1). Nevertheless, there were certainly less distinct tasks of validation and construction since there were informal moments of discussion with poetry scientists during local presentations in the laboratory with visitors or in meetings with all the laboratory colleagues.

The process was iterative since we defined Version 0.1 (DM v0.1[1] in Figure 1) and validate it. Out of this first validation we issued Version 0.2[2] (DM v0.2 in Figure 1). Then, in a new period of construction, we defined Version 0.3[3] (DM v0.3 in Figure 1), finally this version was validated and we issued the first stable version of the Domain Model (DM v1.0 in Figure 1 –version submitted to a scientific journal, waiting for editorial decision). This process is described in detail in Curado Malta, Centenera, and Gonzalez-Blanco (2017) and Bermúdez-Sabel, Curado Malta, and González-Blanco (2017).

1.3 Rationale

The process followed during the building of the Domain Model for European poetry reveals the importance of validation, hence the upgrade of version that each one of them causes as it is shown in Figure 1.

1 Available at https://doi.org/10.5281/zenodo.832885 –accessed in April 21, 2018.
2 Available at https://doi.org/10.5281/zenodo.832906 –accessed in April 21, 2018.
3 Available at http://doi.org/10.5281/zenodo.1164193 –accessed in April 21, 2018.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

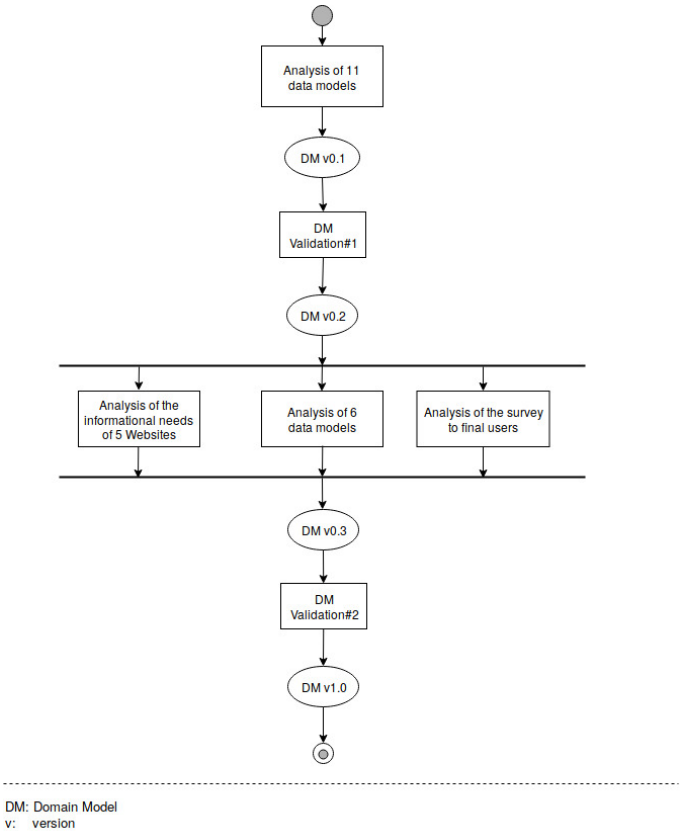

Figure 0. The process of development of the Domain Model of the MAP-EP

The DM-EP must cover any relevant concepts to carry out research about European poetry. This is, without doubt, a very ambitious goal, thus the importance of the collaboration of the community.

We might create a rigorous model from a semantic and philological point of view, but lack the acceptance of the community (Davis 1985). We need to be mindful that the conceptualization of certain elements might be perceived as "intrusive" by the experts. This is why we need to evaluate the input of the targeted community of users in order to make decisions regarding the relevance of the concepts, the quality of their definition (and even their denomination).

Taking all this into consideration, our work should always be open to be reviewed, which means that we need to design the manner in which the community of practice can provide an evaluation. With this goal in mind, we have designed the current deliverable: a validation form for the DM-EP.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

2. Methodology

2.1 Conceptualisation

The goal of the form is to validate the DM-EP by experts of the domain (not experts on the LOD ecosystem). Considering the dimensions of the DM-EP and its complexity, the major challenge was to create an user-friendly method that would cover the validation of each entity, attribute and relation of the DM-EP.

The team considered that a form through which the scholar is able to depict the analysis of a poetic resource will be the most advantageous method due to the following reasons. On the one hand, by working on a specific use-case, the scholar has a better control at expressing his or her informational needs. On the other hand, we have the opportunity to provide something back for their effort: an RDF dataset of their analyses.

2.2 Programming of the form

The core of validation of the DM-EP is the proof of its concepts and relations. Its development consists on the creation of a group of HTML5 and PHP forms. The results will be capture by using JavaScript and JQuery libraries.

In order to enable the independent sending and progressive storage of data, we implemented the Ajax technique of asynchronous requests. This technique makes it easier to support the workflow of the forms and increases both the usability and the track of the process in regards to the users, since the request are sent in a transparent way and the user doesn't need to reload the pages.

The group of forms was designed following these criteria:

1. The concepts are classified in order to generate both descriptive and structural metadata. We apply a technique in line with the principle of divide and rule which eliminates the complexity when designing forms.

2. The development flow has been designed so that, in the first place, the classes with less data and, above all, with fewer properties of objects are presented.

3. The form has been developed by means of a flow designed with the objective of presenting it to the user so that it can carry out its filling in the simplest and most guided way possible. To this end, it has been considered to request first the introduction of concepts with more data and that appear in more relationships due to its widespread use as people, organizations and places. In this way, it facilitates the definition of the own and specific elements of poetry and the link with the entities already created.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

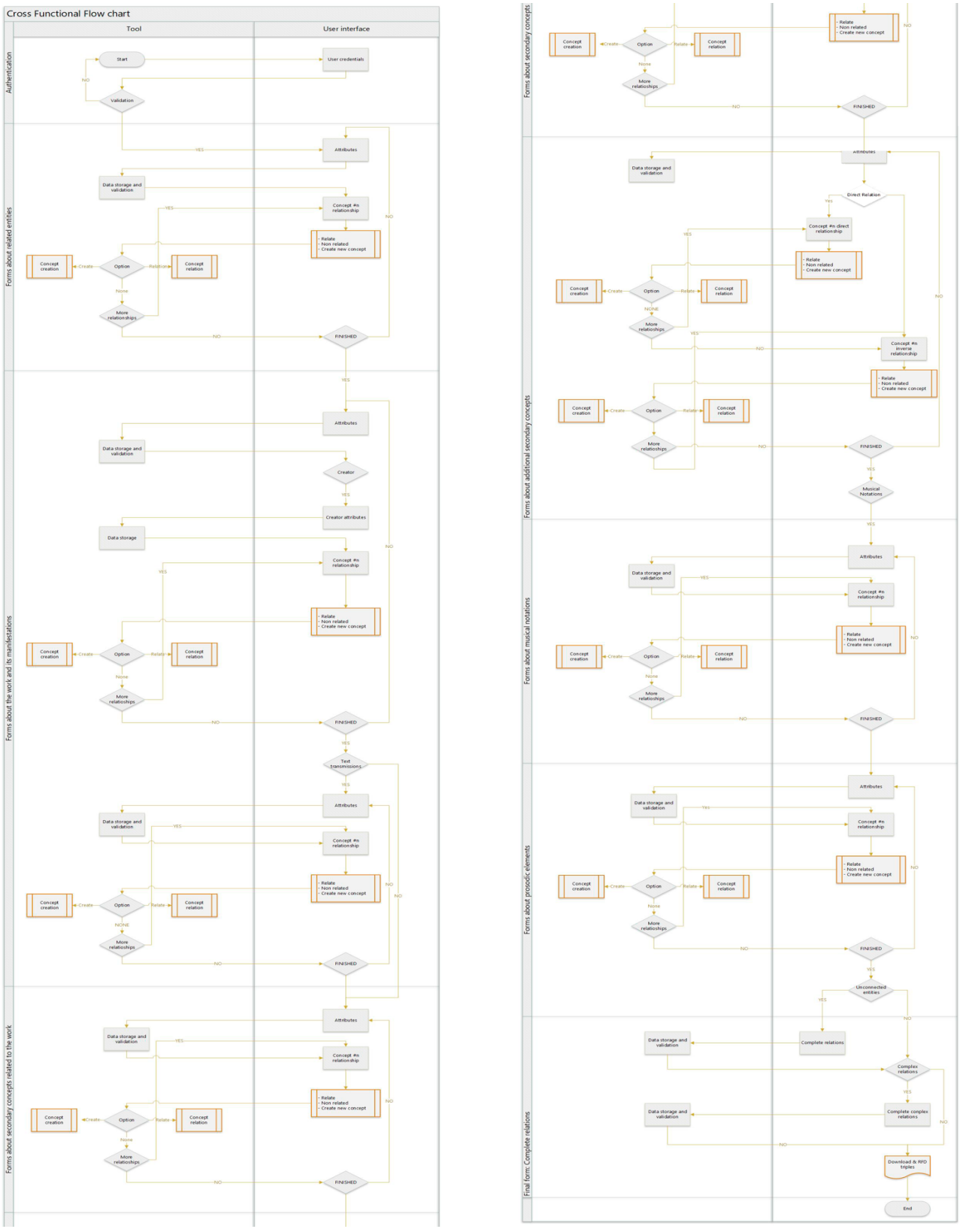

Figure 1 Internal flow of forms and interaction

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

4. The storage of the information has been conceived with a future orientation. The information is stored in triple that are based on RDF, although the properties and concepts are not yet assigned with the ontologies and the data models. Therefore, the validation form provides the improvement value of the transition to what will be the future virtual environment in which the corresponding data sets will be generated.

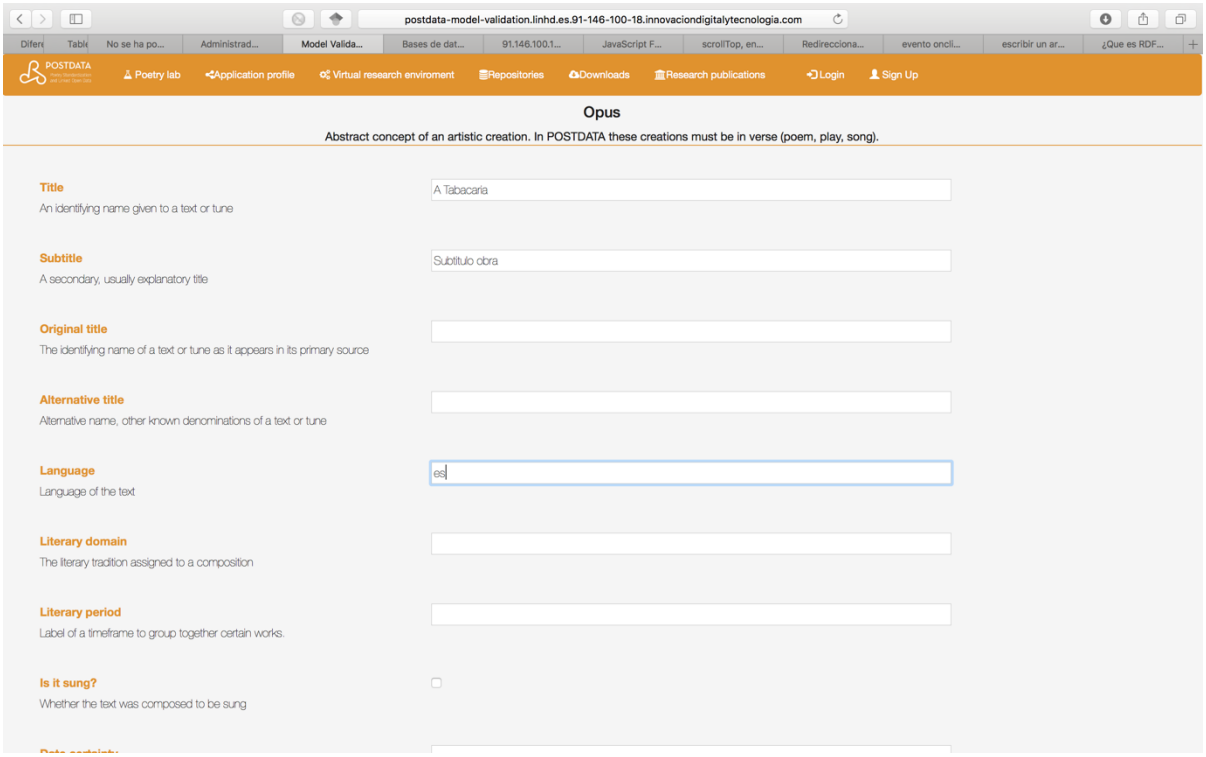

Figure 0. Example: form with the properties of Opus

As shown in Figure 3, for each concept, and in a dynamic way, we load the generic template with their data properties. Before sending the data, the available buttons give the option to create more instances of that same concept or to advance to the next screen. The form does not validate if all the fields are empty: this prevents the user for creating an instance with no information.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

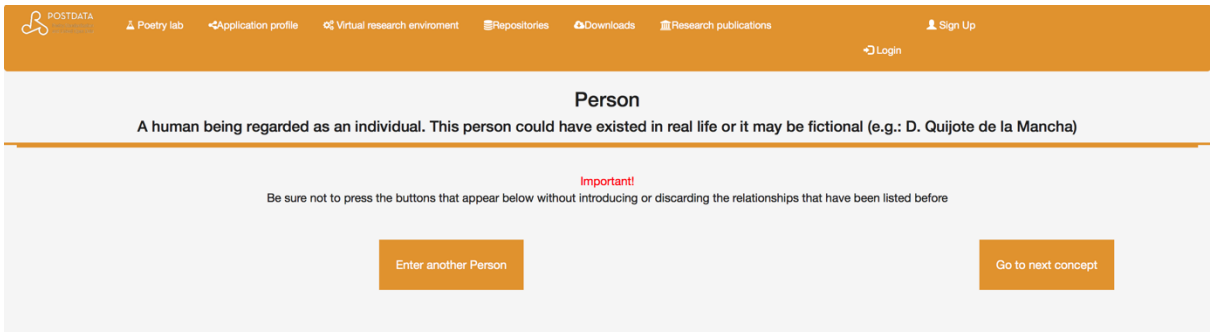

Figure 3. Example: possibilities before sending the data

2.3 Storing and transformation of data

As a result of this form, the user creates a data set with the information provided, according to the MAP of POSTDATA.

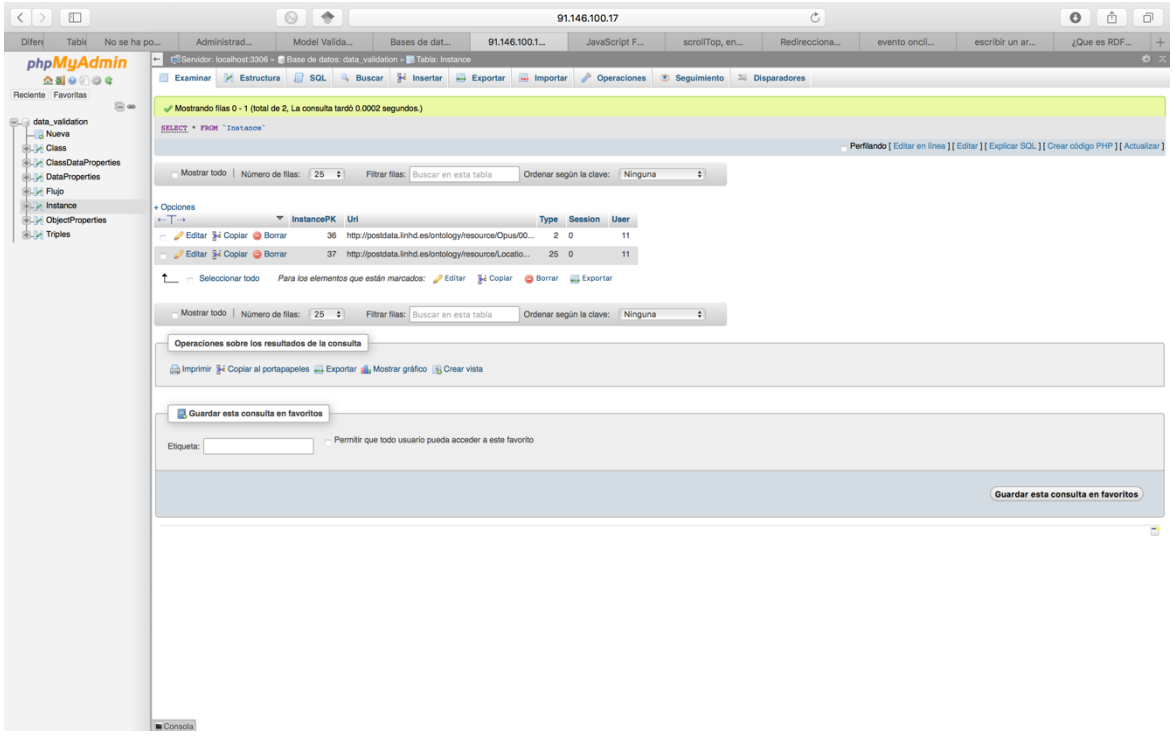

Figure 0. Example: generated instances with its type and the automatically created URI

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

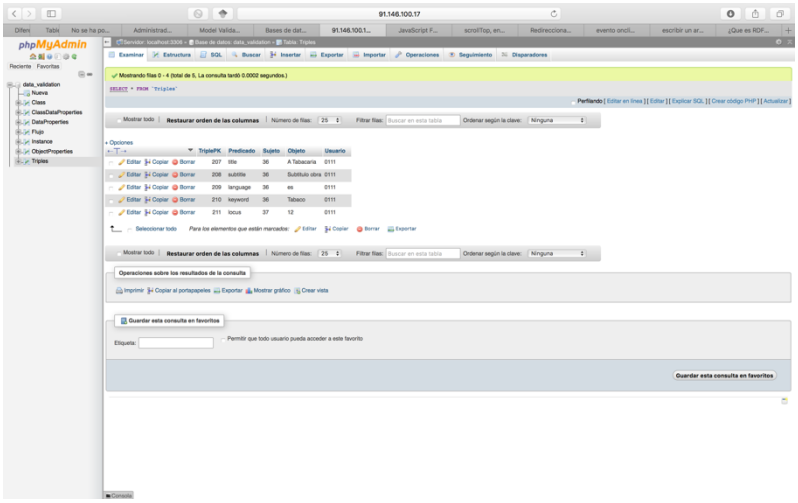

Figure 5. Example: generated triples

```
< rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#' xmlns:ns='http://www.example.org/#' >
    < rdf:Description rdf:about="http://postdata.linhd.es/ontology/resource/Opus/00024" >
        < rdf:type rdf:resource="http://postdata.linhd.es/ontology/Opus" />
        < title >A Tabacaria< /title >
        < subtitle >Subtitulo obra< /subtitle >
        < language >es< /language >
        < keyword >Tabaco< /keyword >
    < / rdf:Description >

    < rdf:Description rdf:about="http://postdata.linhd.es/ontology/resource/Location/00012" >
        < rdf:type rdf:resource="http://postdata.linhd.es/ontology/Location" />
        < locus >12< /locus >
    < / rdf:Description >

< /rdf:RDF>
```

Figure 0. Example: resulted RDF/XML

3. Dissemination

In order to create the DM-EP, the team studied multiple databases of EP. However, it was impossible to tackle all European literary traditions. Thus, the most important goal of the validation form was to engage scholars of the traditions that were not studied during the development process.

We created a list of contacts as follows:

- After listing all European languages, we looked for experts in poetry for each one of them.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

- We included the contact of Literature departments of major European Universities such as Oxford, UCL or Sapienza - Università di Roma.
- We looked for research centres whose main field of study was the so called Digital Humanities.
- We also included mail lists related to the Humanities in general and Digital Humanities in particular.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

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. https://doi.org/10.1007/978-3-319-59888-8_14.

Coyle, Karen, and Thomas Baker. 2009. 'Guidelines for Dublin Core Application Profiles'. 2009. http://dublincore.org/documents/profile-guidelines/.

Curado Malta, Mariana, Paloma Centenera, and Elena Gonzalez-Blanco. 2017. 'Using Reverse Engineering to Define a Domain Model: The Case of the Development of a Metadata Application Profile for European Poetry'. In *Developing Metadata Application Profiles*, by Mariana Curado Malta, Ana Alice Baptista, and Paul Walk, 146–80. IGI Global. https://doi.org/10.4018/978-1-5225-2221-8.

Davis, F. D. 1985. 'A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results'. Massachusetts Institute of Technology.

González-Blanco, E, and Levente Seláf. 2014. 'Megarep: A Comprehensive Research Tool in Medieval and Renaissance Poetic and Metrical Repertoires'. Edited by L. Soriano, M. Coderch, G. Sabaté, and Espluga. *Humanitats a La Xarxa: Món Medieval/Humanities on the Web: The Medieval World*, 321–332.

Nilsson, Mikael, Thomas Baker, and Pete Johnston. 2009. 'Interoperability Levels for Dublin Core Metadata'. 2009. http://dublincore.org/documents/interoperability-levels/.

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

Annexes

Annex X

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.

Annex X Acronyms

DM = Domain Model
DM-EP = Domain Model for European Poetry
EP = European Poetry
LOD = Linked Open Data
MAP = Metadata Application Profile

POSTDATA has received funding from by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 679528.