



# POSTDATA: Poetry Standardization and Linked Open Data

## Deliverable N° (WP2.2) Domain Model for European Poetry

Deliverable Fact Sheet	
Deliverable Version:	1.0
Deliverable Nature:	R = Report
Dissemination Level:	RE = Restricted
Work Package:	WP n° 2
Organisation Responsible:	UNED
Document author/s	Helena Bermúdez Sabel

Actual Date of Delivery	
Audience	<input type="checkbox"/> Public <input checked="" type="checkbox"/> Restricted <input type="checkbox"/> Internal

Version number	
Status	<input type="checkbox"/> Draft <input checked="" type="checkbox"/> WP responsible/s accepted



<input checked="" type="checkbox"/> Project coordinator accepted
--

## Summary

This deliverable concerns the first version of a domain model for European Poetry (DM-EP). A domain model is a data model that captures the concepts of a context, its attributes and the relations between those concepts, and it is defined as an important milestone in the development of a metadata application profile (MAP), a construct of the semantic web. A MAP provides reference information about the global context of both the related data and its constraints. This domain model is the basis for a declaration of terms of Resource Description Framework (RDF) vocabularies that will serve the European community of poetry. A MAP is a construct that will provide means for this community to publish interoperable, linked open data. This report presents the methodology used to develop the DM-EP, referencing the methods and techniques employed as well as the organisation of the various sources of input of information during this process of development. The DM-EP was defined by studying twenty-three repertoires of poetry which represent different languages, periods of composition, and prosodic systems of the European Poetry.



## Contents

1. Introduction.....	4
1.1 Background.....	4
2. Methodology.....	5
2.1 Retrieval and analysis of informational needs.....	5
2.2 Building the Domain Model.....	5
2.2 Workflow.....	6
3. Results.....	8
References.....	9
Annexes.....	10
Annex 1 Acronyms.....	10



# 1. Introduction

As its name reveals, one of the main aims of Poetry Standardization and Linked Open Data project (POSTDATA) is to provide a means to publish European poetry (EP) data as Linked Open Data (LOD). Thus, developing a metadata application profile (MAP) as a common semantic model to be used by the EP community is a crucial step of this project. This MAP will enhance interoperability among the community members in particular, and among the EP community and other contexts in general (e.g. bibliographic records). This report presents the methodology followed in the process of defining the concepts of the domain model of this MAP, as well as some issues that arise when labelling philological terms.

## 1.1 Background

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.”

The starting point for the definition of the DM-EP is a set of twenty five repertoires, most of them available on the Web of Documents. These twenty five repertoires are served by databases, where the data models are relational or hierarchical. The repertoires represent different poetry traditions, languages and cultures. The criterion with most weight in the selection of repertoires was their availability, both in terms of having access to their internal structure and in terms of the ability of our research team to understand and analyse their contents. Nevertheless, great efforts were made in order to gather a representative sample for which the language, the period of composition, and the prosodic system (metre) were considered as defining criteria.

The map available in the link <https://goo.gl/9MCWrv> locates these repertoires in one of the countries where the poetic tradition at hand was originated, and they are grouped according to chronological criteria.



## 2. Methodology

### 2.1 Retrieval and analysis of informational needs

In order to develop a MAP for the European Poetry, the authors are following a systematic set of activities defined by Me4MAP (Curado Malta and Baptista n.d.).

The definition of the domain model, a common conceptual model that should represent the informational needs of the EP community of practice, integrates the data requirements that result from defining the functional requirements, together with the results of the following sub-activities:

- Analysis of the data model of a representative sample of EP databases.
- Analysis of a survey addressed to the final users of the repertoires in order to understand the data needs of the users of poetry databases.

### 2.2 Building of the Domain Model

During the process of analysis of a data model:

- every concept of the database, as well as the properties that characterize that concept, are identified
- the relationships between concepts are also defined.

In what follows, these conceptual elements will be referred to as “Concepts” regardless of whether they are concepts, properties or relationships between concepts.

As it has been mentioned, the analysis is very iterative, which means that similar Concepts that have been already identified in previous analyses are compared with the current analysis, and that those Concepts that are equivalent are given the same description and named in the same way.

Occasionally, the level of abstraction increases and names are changed retroactively; that is, previous analyses are re-evaluated. In the beginning of the process abstraction is low, but it increases with the number of analyses made. As a result, at the end of the procedure the level of abstraction is higher than it was at the start, which decreases the level of granularity of the final model. As in any process of semantic modelling, there is always some tension between interoperability and semantics. The level of semantics is related to the possibilities of data sharing, which means that the researchers look for the highest level of meaning in the definition of the Concepts without compromising interoperability. The same Concepts from different databases will contain data that can be shared, and different Concepts will contain data that cannot be shared. However, if a specific Concept is different but similar to other Concept that has been already identified while analysing other databases, semantics may be lost in favour of interoperability gain: a new broader Concept is 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.

Every Concept analysis integrates the actions presented in Figure 1. The process begins with the identification of a Concept in a data model analysis. Then a study of similar Concepts in the previous data model analyses is carried out in order to understand if the Concept at hand is new or if it has already been identified.

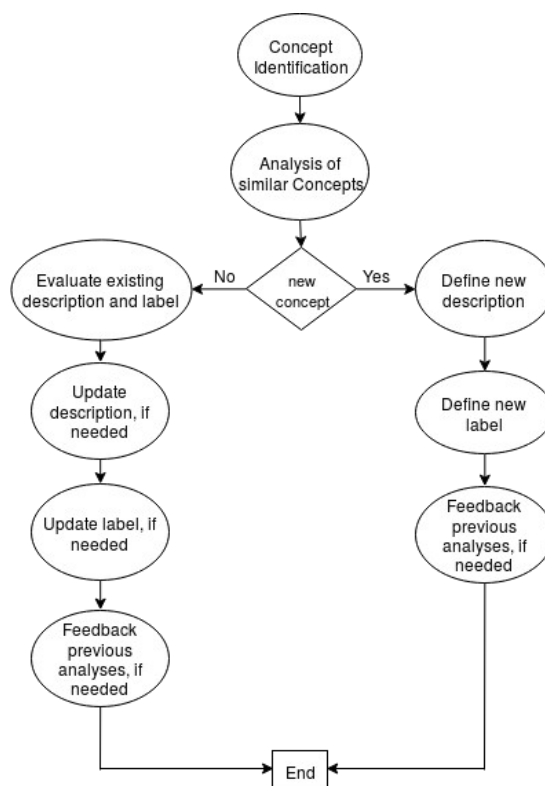


Figure 1. Diagram with the sequence of actions of a Concept analysis process

## 2.2 Workflow

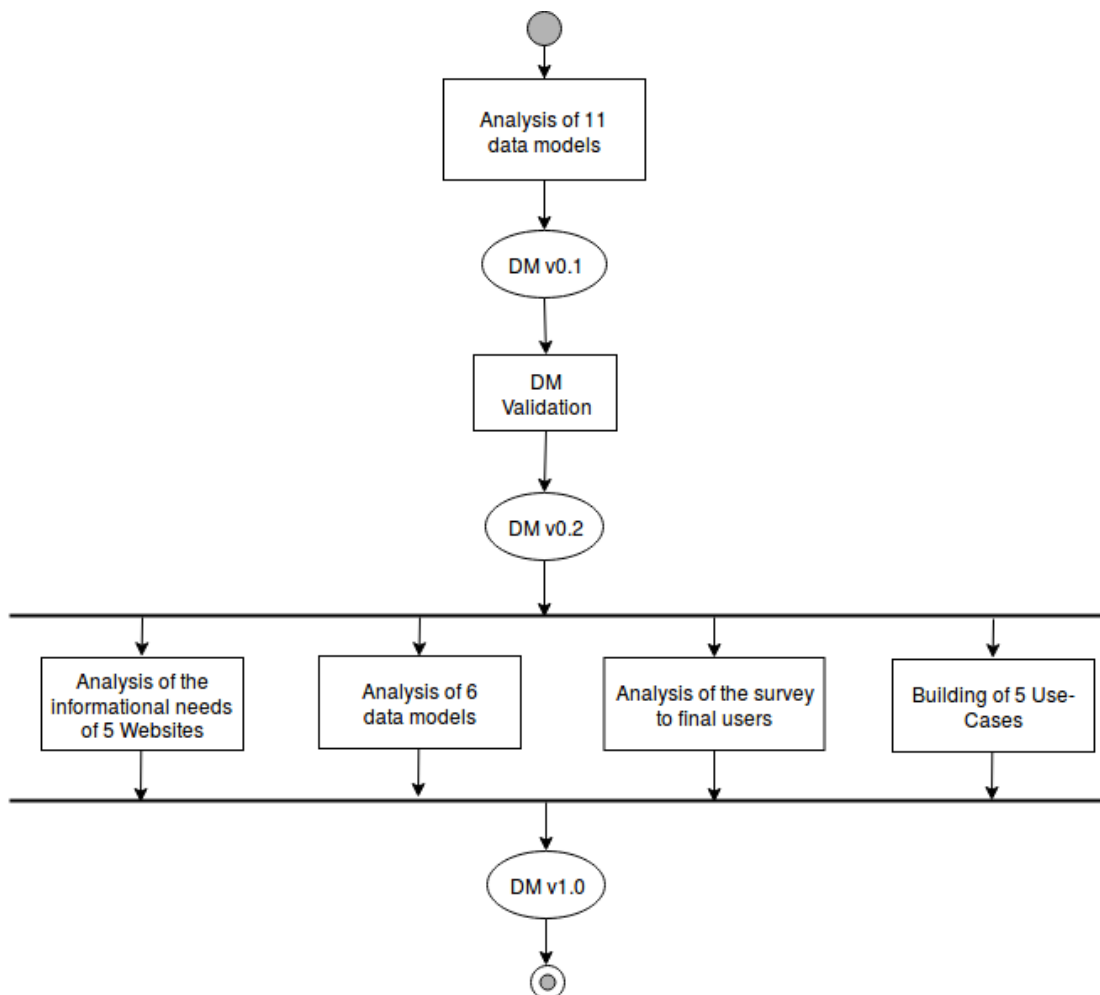
The repertoires mentioned in §1.1 were employed in different moments of the development process and the map available in the link <https://goo.gl/O0mqhI> shows what repertoires were used in which phases.

Figure 2 presents the development process of the DM-EP showing the different stages of work, analysis and validation. The DM-EP was developed in an iterative way and, over time, information from different sources was collected and included as explained as follows.

POSTDATA contacted delegates from the different repertoires, inviting them to participate as stakeholders of the project. The delegates were asked to send the structure of the databases and any additional documentation in order for the researchers to be able to analyse and study the data models. Thus, the starting point for the analysis was a set of MySQL dumps, XSD and XML files, Perl scripts, and spreadsheets. The workteam applied a

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.

reverse engineering approach using software engineering techniques. To extract and compare all the concepts in each data model and to construct a common model out of them, the work team decided to build conceptual models for each one of the data models analysed. This process is described in detail in Curado Malta, Centenera, and Gonzalez-Blanco (2017) and Bermúdez-Sabel, Curado Malta, and Gonzalez-Blanco (2017).



DM: Domain Model  
v: version

**Figure 2.** Diagram with the workflow of the Domain Model development process







## 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](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, and Ana Alice Baptista. n.d. *Me4MAP V1.0: A Method for the Development of Metadata Application Profiles*. Submitted for publication.
- 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>.
- Nilsson, Mikael, Thomas Baker, and Pete Johnston. 2009. 'Interoperability Levels for Dublin Core Metadata'. 2009. <http://dublincore.org/documents/interoperability-levels/>.



# Annexes

## Annex 1 Acronyms

DM = Domain Model

DM-EP = Domain Model for European Poetry

EP = European Poetry

LOD = Linked Open Data

MAP = Metadata Application Profile