Towards interoperability in the European poetry community

Helena Bermúdez Sabel
helena.bermudez@linhd.uned.es
Laboratorio de Innovación en Humanidades Digitales, UNED, Spain
Member of the project Poetry Standardization and Linked Open Data - POSTDATA
Outline

• Presentation
• Rationale
• Methodology
• Concepts standardisation
• Conclusion and future work
Outline

• Presentation
• Rationale
• Methodology
• Concepts standardisation
• Conclusion and future work
POSTDATA axes

Semantic Web Infrastructure
Linked Open Data (LOD)

Virtual Research Environment
Digital Scholarly Editing

Poetry Lab
Natural Language Processing
Outline

• Presentation
• **Rationale**
• Methodology
• Concepts standardisation
• Conclusion and future work
Western literature
And yet...

- Diversity of paradigms
- Fragmentary access to poetic resources
And yet...

- Diversity of paradigms
- Fragmentary access to poetic resources
- Partiality of methods (and results)
A workaround

• The key concepts: standardisation and interoperability
• The most advantegous paradigm: Linked Open Data
LOD cloud

http://lod-cloud.net/
LOD

- A huge database (worldwide) on the Web of Data
- Since the data is open, the community can rely also on other sources of data
- Big advantage over other approaches
Goal

• Enhance interoperability between existent poetic resources

• Facilitate the creation of new resources
Outline

• Presentation
• Rationale
• **Methodology**
• Concepts standardisation
• Conclusion and future work
Main task

• Defining a Metadata Application Profile (MAP) for European Poetry
Main task

- Main task: Defining a Metadata Application Profile (MAP) for European Poetry

(MAP: a semantic model in the context of LOD)
Main task

• Main task: Defining a Metadata Application Profile (MAP) for European Poetry
• Milestone: to define a Domain Model
Main task

• Main task: Defining a Metadata Application Profile (MAP) for European Poetry
• Milestone: to define a Domain Model

(Domain Model: a model that presents the concepts of the context and its relations)
Starting point

- Repertoires of European Poetry and Poetry Metrics
  - Data models
  - Functionalities of the Websites
- Report of data needs of the community
Starting point
Workflow

Analysis of 11 data models

DM v0.1

DM Validation

DM v0.2

Analysis of the informational needs of 5 Websites

Analysis of 6 data models

Analysis of the survey to final users

DM v0.3

DM Validation

DM v1.0
Process

• 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.
Analysis of the data models

- Analysis of the file sent by each partner with the structure of the database
- Study of the structure of the database
- Identification of every concept of the database, as well as the properties that characterize that concept
- Identification of the relationships between concepts
Concept analysis process

1. Concept identification
2. Evaluate existing description and label
   - No: new concept?
     - Yes: Define new description
     - No: Update description, if needed
   - Yes: Define new label
3. Update label, if needed
4. Reevaluate and revise previous analyses, if needed
5. End
Types of modeling

- **egoistic modeling**: to express specific research ideas in cases where data is being created to support the creator’s own research needs

- **altruistic modeling**: to serve as an interchange format for some types of users and user communities where data is typically being created and modeled with someone else’s needs in mind

(Jannidis & Flanders 2013)
Importance of data modeling

- (Effective) semantic interoperability
- Data needs of the community are met
Looking for balance

The tension between semantics and interoperability

- Level of abstraction tends to increase (and names are changed retroactively)
- Semantics may be lost in favor of interoperability gain
Outline

• Presentation
• Rationale
• Methodology
• Concepts standardisation
• Conclusion and future work
Issues

• Multi-linguistic data
  Eg. “poetic line”

• Diachronic corpora

• Different philological traditions
  Eg. “dieresis”

Same term in different technical vocabularies

• Different terms depending on the theoretical background
  for similar concepts
  Eg. “apparatus”
Issues

• Multi-linguistic data
Issues

- Multi-linguistic data
- Eg. “poetic line”
Issues

- Multi-linguistic data
- Diachronic corpora
Issues

• Multi-linguistic data
• Diachronic corpora
  • Eg. “line group as a metrical unit”
Issues

- Multi-linguistic data
- Diachronic corpora
- Different philological traditions
  - Same term in different technical vocabularies
    - Eg. “diaeresis”
Issues

- Multi-linguistic data
- Diachronic corpora
- Different philological traditions
  - Same term in different technical vocabularies
  - Different terms depending on the theoretical background for similar concepts
    - Eg. “apparatus”
Semantic efficiency

- Evaluation of which concepts should be integrated
- Evaluation of which concepts should be merged
- Evaluation of which concepts should be distinguished
Depending on the prosodic tradition:

• we count syllables
• we discern the distribution of stresses
• we analyse the length of the syllables
Metrical scheme

- "syllabicMetricalScheme"
- "accentualMetricalScheme"
Validation

• We might create a rigorous model from a semantic and philological point of view, but lack the acceptance of the community

• The conceptualization of certain elements might be perceived as “intrusive”
Outline

• Presentation
• Rationale
• Methodology
• Concepts standardisation

• **Conclusion and future work**
Research possibilities

• Questions about the contents
• Questions about the structure
Research possibilities

• Questions about the contents
• Questions about the structure
• Conclusion: a broader contextualization and a more accurate picture
Work in progress

- Vocabulary alignment: match each concept with a term of a RDF vocabulary
- Encode the semantic model
- Report on how to enrich the repertoires’ data with links to existent resources
- Develop documentation: Manuals, HowTos, Use case examples
helena.bermudez@linhd.uned.es

http://postdata.linho.uned.es/
http://linhd.uned.es
@linhdununed