POSTDATA

Poetry Standardization and Linked Open Data

Poetry and Digital Humanities making interoperability possible in a divided world of digital poetry: POSTDATA project

Salvador Ros
UNED-LINHD
Elena González-Blanco García
COVERWALLET – LINHD

Galway, 9 December 2018

UNED

LINHD
LABORATORIO DE INNOVACIÓN EN HUMANIDADES DIGITALES

erc
European Research Council
Established by the European Commission
Supporting top researchers from anywhere in the world

The “digital gap”

POSTDATA Vision is to ease humanists research through its digital tranformation

POSTDATA
Poetry Standardization and Linked Open Data

Traditional Research		Digital Humanities
Counting syllables with fingers	→	Automatic syllable counter
Comparing 2-3 poems at the same time	→	Comparing thousands of poems in parallel
Looking for similar evidences by reading all papers and books	→	Finding multiple evidences through one search engine
Individual research	→	Collaborative, interdisciplinary research

The three POSTDATA Axes

POSTDATA
Poetry Standardization and Linked Open Data

Semantic Infrastructure

Linked open data

Virtual Research Environment

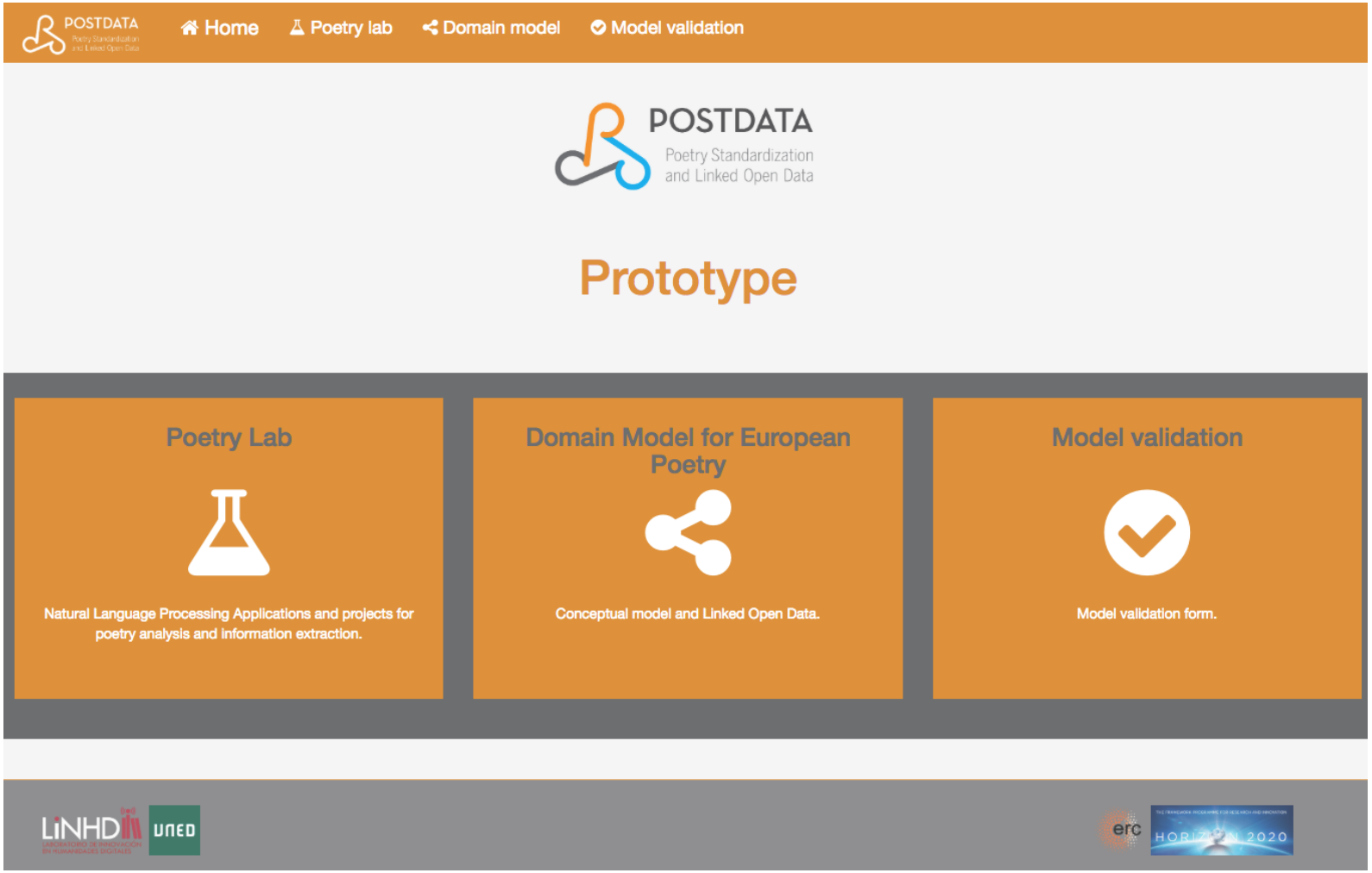
Digital Scholarly Editing

Poetry Lab

Natural language processing

Postdata Prototype

POSTDATA
Poetry Standardization
and Linked Open Data



DOMAIN MODEL FOR EUROPEAN POETRY

POSTDATA
Poetry Standardization and Linked Open Data

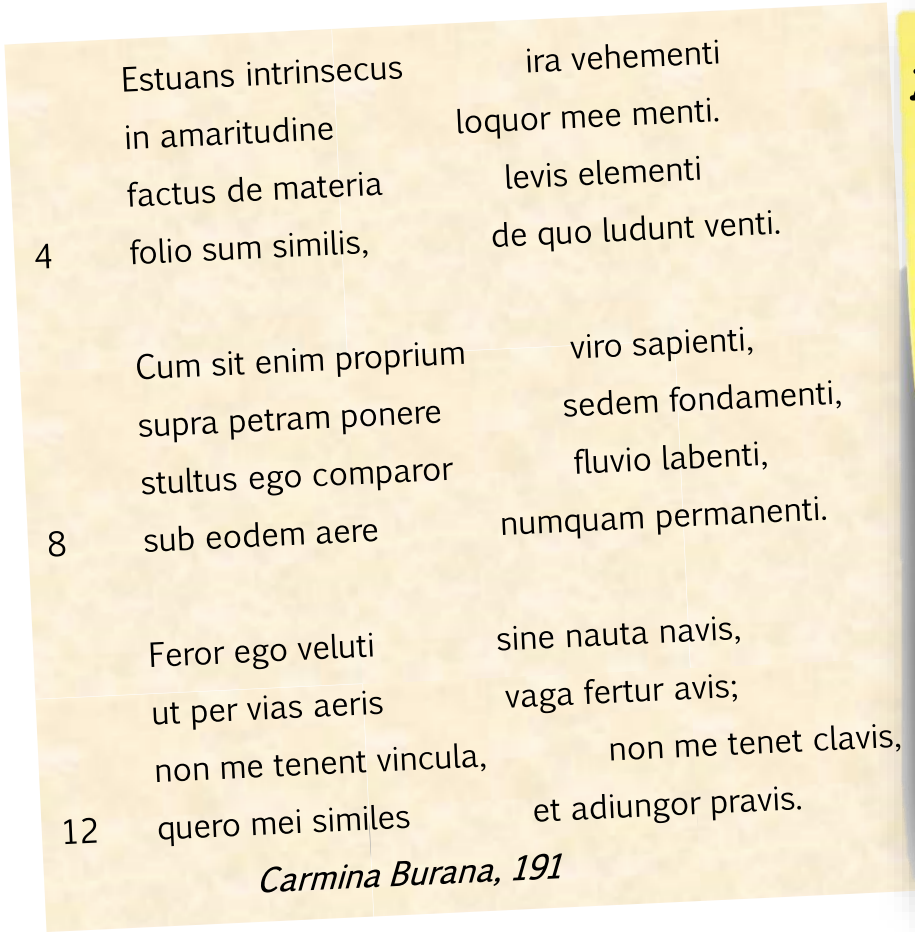
Western Literature

- There is not an ontological model for poetry
- It is necessary to standardize specific fields (rhyme, metrical scheme)
- It is necessary to map controlled vocabularies across different languages

DOMAIN MODEL FOR EUROPEAN POETRY

POSTDATA
Poetry Standardization and Linked Open Data

But, what does define a POEM?



Estuans intrinsecus ira vehementi
in amaritudine loquor mee menti.
factus de materia levis elementi
4 folio sum similis, de quo ludunt venti.

Cum sit enim proprium viro sapienti,
supra petram ponere sedem fondamenti,
stultus ego comparor fluvio labenti,
8 sub eodem aere numquam permanenti.

Feror ego veluti sine nauta navis,
ut per vias aeris vaga fertur avis;
non me tenent vincula, non me tenet clavis,
12 quero mei similes et adiungor pravis.

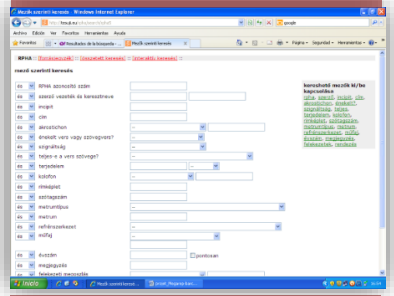
Carmina Burana, 191

- Author
- Title
- Incipit
- Manuscript
- Post quem
- Ante quem
- Language
- Topics
- Edition
- Online edition
- Work
- Isometrism
- Isostrophism
- Metrical scheme
- Rhyme scheme
- Rhyme
- Musical notation
- Number of stanzas
- Number of lines
- Poetic form

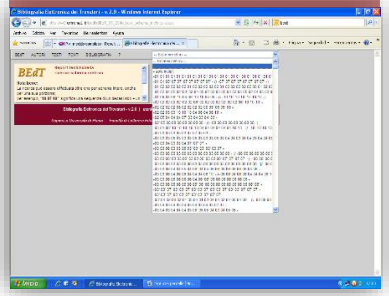
DOMAIN MODEL FOR EUROPEAN POETRY

POSTDATA
Poetry Standardization and Linked Open Data

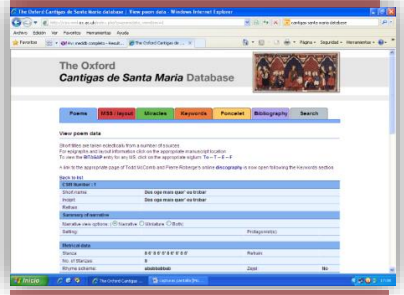
RPHA - Hungary



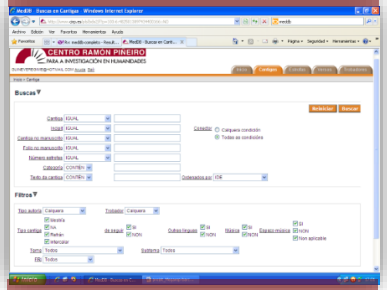
BedT - Italy



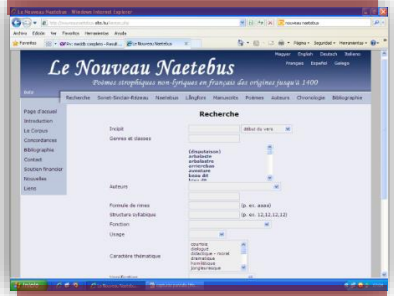
Cantigas Oxford



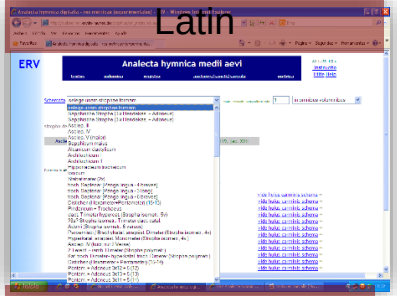
MedDB2 – Galicia



Naetebus - France



Analecta Hymnica Latin



ReMetCa - Spain

Skaldic - Scandinavia

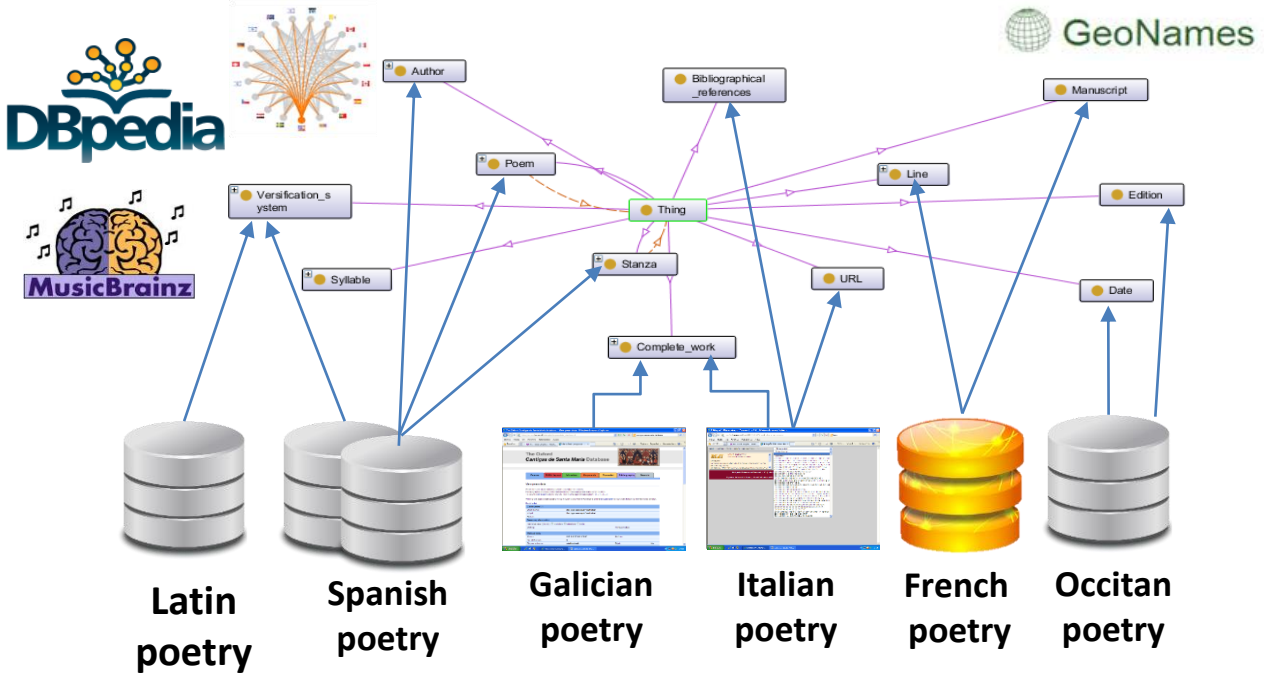
Mirabile - Italy

The puzzle of digital poetic repertoires

DOMAIN MODEL FOR EUROPEAN POETRY

POSTDATA
Poetry Standardization and Linked Open Data

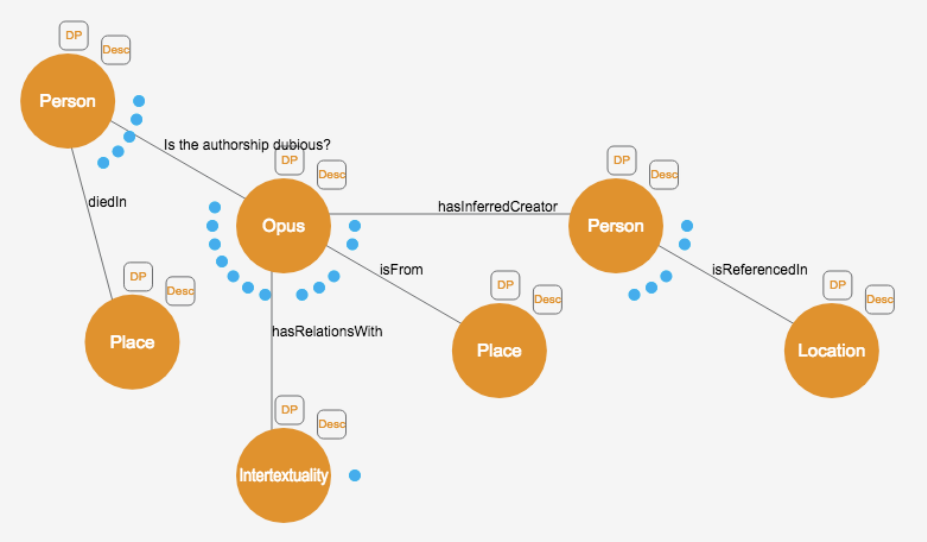
POSTDATA looks for an abstract model with common elements across the different poetic systems



DOMAIN MODEL FOR EUROPEAN POETRY

POSTDATA
Poetry Standardization and Linked Open Data

- DOMAIN MODEL FOR EUROPEAN POETRY
- key concepts:
  - standardisation and
  - interoperability
- The most advantageous paradigm
  - Linked Open Data



Poetry Lab : NLP tools

POSTDATA
Poetry Standardization and Linked Open Data

- Tools for (automatic) poetry analysis
- Largely based on Natural Language Processing
- Insofar as literary phenomena in domain model rely on linguistic traits, NLP can help automate their annotation

Poetry Lab: Work done

POSTDATA
Poetry Standardization and Linked Open Data

Content analysis:

- Medieval Spanish Named Entity Recognition (Díez Platas et al., 2017)

Phonetics and metrics:

- Lexical Syllabification (prereq for scansion)

Other stylistic traits:

- Enjambment detection (Ruiz, Martínez et al., 2017)

Poetry Lab: Syllabification

POSTDATA
Poetry Standardization and Linked Open Data

- • A prerequisite for scansion
- • Syllable boundary detection + pos-tagging to determine stressability (weak vs. strong categories)

#	Text	1	2	3	4	5	6	7	8	9	10	11
1	Ya se doran las hojas, ya las tardes	**Ya**	se	**do**	ran	las	**ho**	jas	**ya**	las	**tar**	des
2	Son lo mejor del día	**Son**	lo	me	**jor**	del	**dí**	a				
3	Bajo su luz madura de sosiego,	Ba	jo	su	**luz**	ma	**du**	ra	de	so	**sie**	go

Poetry Lab: Enjambment

POSTDATA
Poetry Standardization and Linked Open Data

- Mismatch between metrical and syntactic units
- Defined in syntactic terms
- Can be detected based on pos-tagging / parsing
- Definition/detection not trivial
- Previously unanalyzed with automatic means

Poetry Lab: Enjambment

POSTDATA
Poetry Standardization and Linked Open Data

- Enjambment takes place when particularly cohesive syntactic units are broken up by a metrical pause (across lines or hemistichs)

- Example

1. Prefiero
2. quedarme desnuda //
3. de amigo insincero

1. I'd rather
2. be naked //
3. of fake friend

INTERPRETATIONS

- a) naked (no clothes)
- b) naked of fake friend (without such a person)

Poetry Lab: NERC

POSTDATA
Poetry Standardization and Linked Open Data

- Hispanic Medieval Tagger (Díez Platas et al., 2017)
- Medieval Spanish NERC
- Person, location, organization, role names
- Normalization of historical spellings
- Geolocalization
- TEI output

Poetry Lab: NERC

POSTDATA
Poetry Standardization
and Linked Open Data

sros@scc.uned.es

http://postdata.linhd.es

POSTDATA

Poetry Standardization
and Linked Open Data