



European  
Research  
Council



# PoetryLab

## An Open Source Toolkit for the Analysis of Spanish Poetry Corpora

Javier de la Rosa ([versae@linhd.uned.es](mailto:versae@linhd.uned.es))

Álvaro Pérez ([aperez@linhd.uned.es](mailto:aperez@linhd.uned.es))

Salvador Ros ([sros@scc.uned.es](mailto:sros@scc.uned.es))

Elena González Blanco ([egonzalezblanco@faculty.ie.edu](mailto:egonzalezblanco@faculty.ie.edu))

Laboratorio de Innovación en Humanidades Digitales (LINHD), UNED, Spain

Postdoctoral Researcher at ERC POSTDATA Project (Poetry Standardization and Linked Open Data)



# POSTDATA Project

- ERC-funded POSTDATA Project network of ontologies
- Interchange of data using Linked Open Data
- Varied research interests result in different corpora formats

**PoetryLab** leverages NLP to foster the completeness of the interchanged corpora

## Articulated as:

- Independent Python and Java libraries
- Command Line Interface (CLI) applications
- Unified Web UI and REST API

## Meter (Rantanplan)

- Lexicon
- Prosody
- Rhyme
- Stanzas
- Scansion

- Usage example

```
from rantanplan import get_scansion

poem = """Me gustas cuando callas porque estás como ausente,
y me oyes desde lejos, y mi voz no te toca.
Parece que los ojos se te hubieran volado
y parece que un beso te cerrara la boca.

Como todas las cosas están llenas de mi alma
emerges de las cosas, llena del alma mía.
Mariposa de sueño, te pareces a mi alma,
y te pareces a la palabra melancolía."""

get_scansion(poem)
```

- Usage example

```
[{'tokens': [{'word': [{'syllable': 'Me',  
  'is_stressed': False,  
  'is_word_end': True}],  
  'stress_position': 0},  
{'word': [{'syllable': 'gus', 'is_stressed': True},  
  {'syllable': 'tas', 'is_stressed': False, 'is_word_end': True}],  
  'stress_position': -2},  
{'word': [{'syllable': 'cuan', 'is_stressed': False},  
  {'syllable': 'do', 'is_stressed': False, 'is_word_end': True}],  
  'stress_position': 0},  
...],  
'phonological_groups': [{'syllable': 'Me',  
  'is_stressed': False,  
  'is_word_end': True},  
{'syllable': 'gus', 'is_stressed': True},  
{'syllable': 'tas', 'is_stressed': False, 'is_word_end': True},  
{'syllable': 'cuan', 'is_stressed': False},  
{'syllable': 'do', 'is_stressed': False, 'is_word_end': True},  
...],  
'rhythm': {'stress': '-+----+----+---+', 'type': 'pattern', 'length': 14}},  
...
```

# PoetryLab: Features

## Style

- Syntax (JollyJumper)
  - Rhetorical figures (syntactic)
- Genre (in progress)

## Semantics

- Topics (Daltons)
- Named Entity Recognition (HisMeTag)
- Sentiments (in progress)
- Rhetorical figures (in progress)

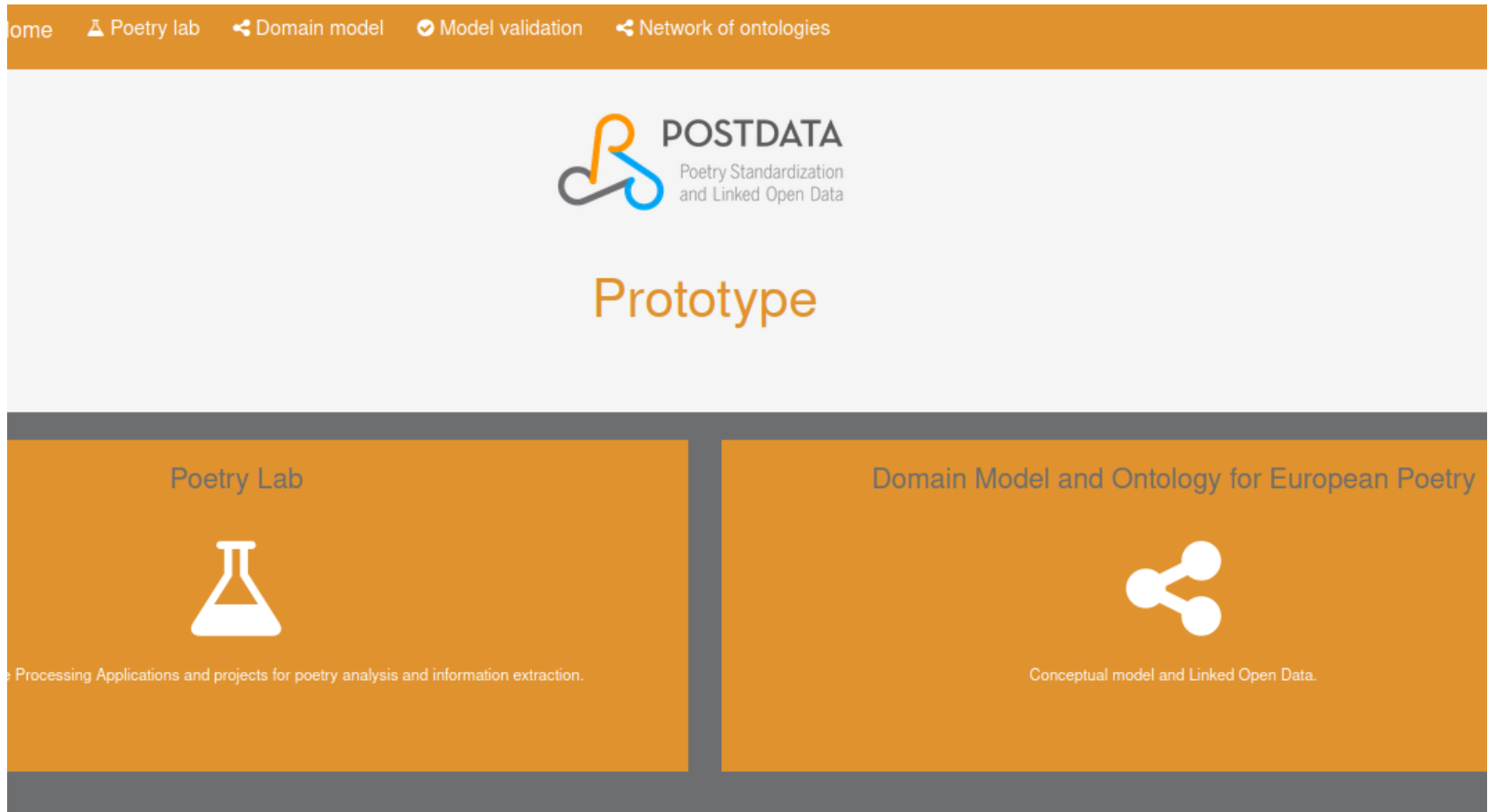


# PoetryLab: Features

## Intertextuality

- Recommender systems
- Relations to image and music

# PoetryLab: Availability



The screenshot shows the PoetryLab website interface. At the top, there is a navigation bar with the following items: Home, Poetry lab, Domain model, Model validation, and Network of ontologies. Below the navigation bar is the POSTDATA logo, which consists of a stylized 'P' icon and the text 'POSTDATA Poetry Standardization and Linked Open Data'. The main content area is divided into two columns. The left column is titled 'Poetry Lab' and features a flask icon. Below the icon, the text reads: 'Processing Applications and projects for poetry analysis and information extraction.' The right column is titled 'Domain Model and Ontology for European Poetry' and features a share icon. Below the icon, the text reads: 'Conceptual model and Linked Open Data.'

<http://postdata.linhd.uned.es/>  
<https://github.com/linhd-postdata/>

# PoetryLab: Challenges

## Interoperability

- Connect disparate packages
- Common API
- Public front-end
- RDF output
- Integration with POSTDATA

## Analysis

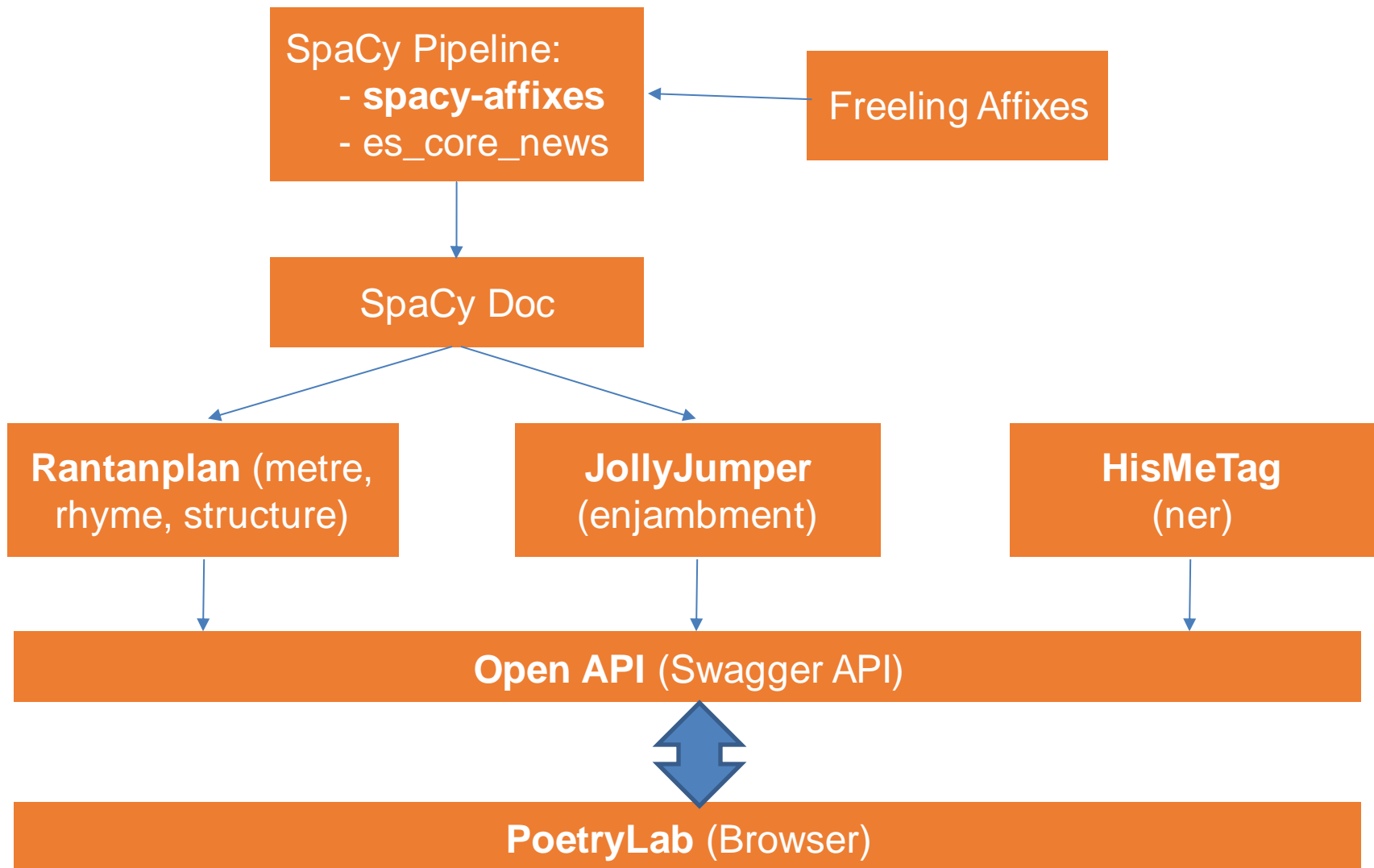
- Syllabification
  - «al-buhe-ra», «cel-tohis-pa-na», «ca-ñiher-la», «de-sahi-jar», «an-ti-hiá-ti-co».
- Proper PoS
- Handling of clitics
- Ambiguity
  - al-du-ro-hie-rro-y-al-cruen-too-fi-cio

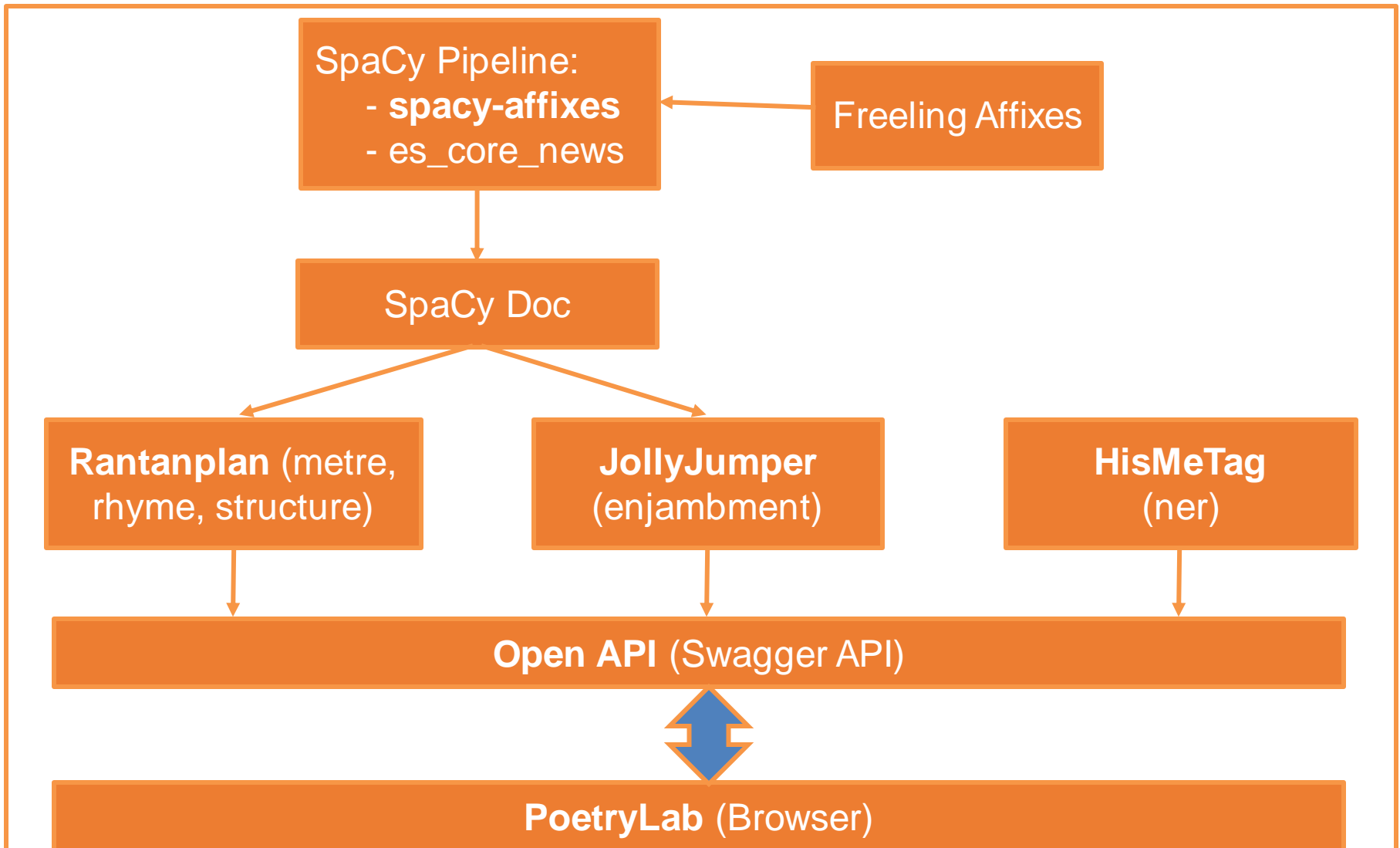
# PoetryLab: Architecture



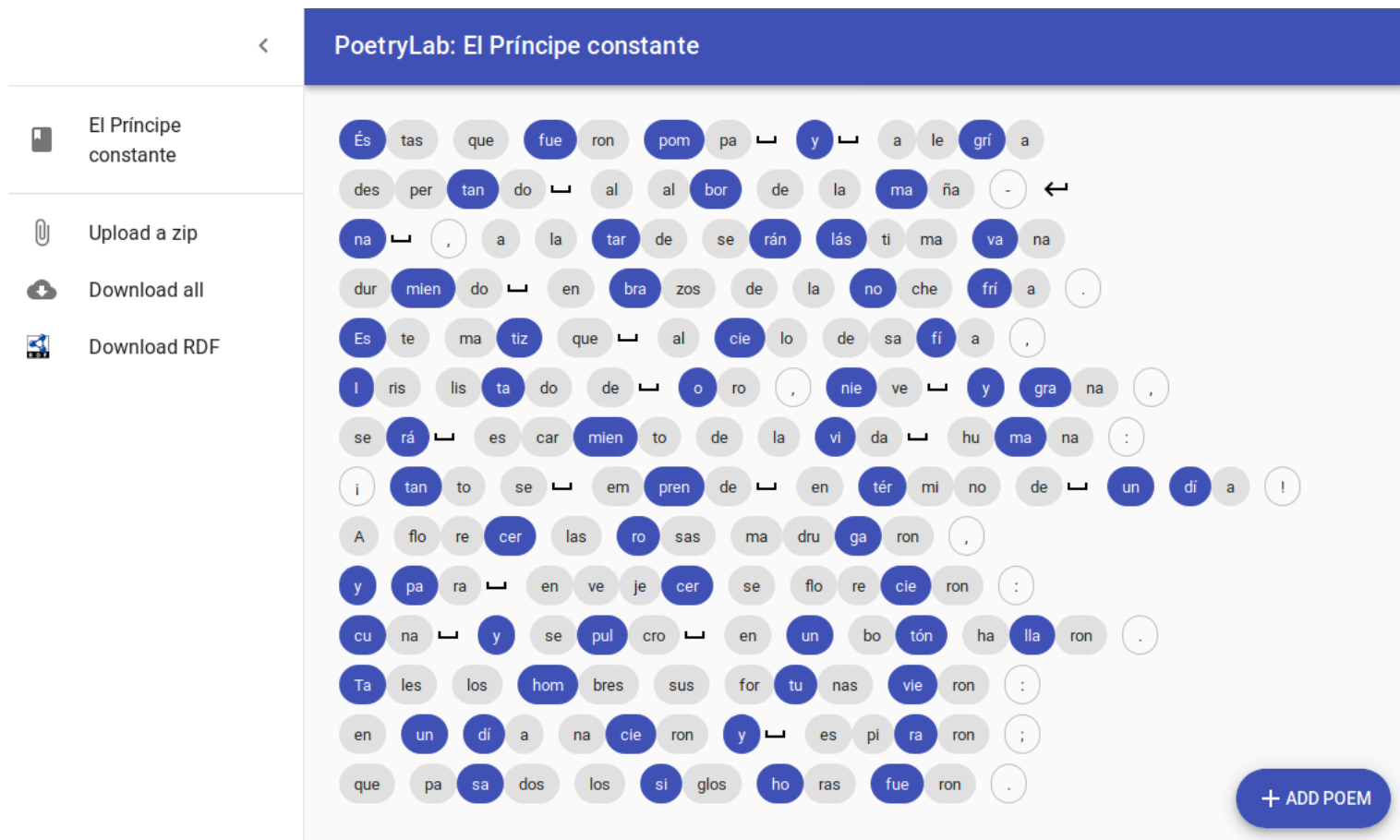
POSTDATA

Poetry Standardization  
and Linked Open Data





- Usage example with the UI



The screenshot displays the PoetryLab interface for the poem 'El Príncipe constante'. The main area shows the text of the poem with individual words and punctuation marks highlighted in blue. A sidebar on the left provides navigation options: 'El Príncipe constante', 'Upload a zip', 'Download all', and 'Download RDF'. A blue button labeled '+ ADD POEM' is located in the bottom right corner.

PoetryLab: El Príncipe constante

És tas que fue ron pom pa y a le gri a  
des per tan do al al bor de la ma ña -  
na , a la tar de se rán lás ti ma va na  
dur mien do en bra zos de la no che fri a .  
Es te ma tiz que al cie lo de sa fi a ,  
I ris lis ta do de o ro , nie ve y gra na ,  
se rá es car mien to de la vi da hu ma na :  
i tan to se em pren de en tér mi no de un dí a !  
A flo re cer las ro sas ma dru ga ron ,  
y pa ra en ve je cer se flo re cie ron :  
cu na y se pul cro en un bo tón ha lla ron .  
Ta les los hom bres sus for tu nas vie ron :  
en un dí a na cie ron y es pi ra ron ;  
que pa sa dos los si glos ho ras fue ron .

+ ADD POEM

# PoetryLab: Benchmarking



POSTDATA

Poetry Standardization  
and Linked Open Data

Task	Tool	Success (%)
Syllabification (EDFU)	<b>PoetryLab (rantanplan)</b>	<b>99.99</b>
	Navarro-Colorado	98.74
Metrical patterns (Navarro-Colorado)	<b>PoetryLab (rantanplan)</b>	<b>96.22</b>
	Navarro-Colorado	94.87
	Gervás	88.73
	Agirrezabal	90.84



# PoetryLab: Lifecycle

- Rantanplan & JollyJumper
  - Continuous Integration
    - Rantanplan: <https://travis-ci.org/linhd-postdata/rantanplan>
    - JollyJumper: <https://travis-ci.org/linhd-postdata/jollyjumper>
  - Stable master
  - Manual release
- API & Frontend
  - CI: <https://travis-ci.org/linhd-postdata/poetrylab-api/>
  - Automatic Docker push:  
<https://hub.docker.com/r/linhdpostdata/poetrylab-api>
  - Automatic Frontend deploy to GitHub pages

# PoetryLab: Demo

- API:
  - <http://postdata.uned.es:5000/openapi.json>
  - <http://postdata.uned.es:5000/ui>
- Frontend
  - <http://postdata.uned.es/poetrylab/>



POSTDATA

Poetry Standardization  
and Linked Open Data

[postdata.linhd.uned.es](http://postdata.linhd.uned.es)  
[@linhduned](https://twitter.com/linhduned)



**POSTDATA**

Poetry Standardization  
and Linked Open Data