TRABAJO DE FIN DE MÁSTER

PORTADA

VISTO BUENO DEL TUTOR Y DECLARACIÓN DE BUENA PRÁCTICA ACADÉMICA

SUMARIO O ÍNDICE

RESUMEN

INTRODUCCIÓN O PRÓLOGO

ESTADO DE LA CUESTIÓN

OBJETIVOS GENERALES

METODOLOGÍA

ESTRUCTURA

PRESENTACIÓN DE LOS RESULTADOS

CONCLUSIONES

REFERENCIAS

ANEXOS

1. Introducción (2 páginas)

El consiguiente trabajo de fin de Máster consiste en el desarrollo, diseño e implementación de un portal web destinado al Libro 1 de la obra Elegías del poeta griego Teognis de Megara.

BLA BLA BLA…

1. Contextualización. Estado de la cuestión (5 páginas)
   1. Contextualización del trabajo

Este trabajo se ha desarrollado como solución a algunas necesidades detectadas tras trabajar en el proyecto ERC POSTDATA (<http://postdata.linhd.es/>).

El proyecto, cuyo acrónimo significa “Poetry Standardization and Linked Open Data” nace en el seno de la UNED (Universidad Nacional de Educación a Distancia) y más concretamente en el LINHD (Laboratorio de Innovación de Humanidades Digitales).

Este tiene por objetivo crear una plataforma digital para la edición de poesía dirigida a diferentes tipos de usuarios: académicos que quieran trabajar en ediciones críticas; usuarios no experimentados que quieran leer, compartir y aprender más sobre tradiciones poéticas; y empresas que podrán utilizar este recurso para diferentes aplicaciones en campos como la educación, el turismo, la psicología u otros propósitos culturales: todo ello con el afán de construir puentes entre la filología y la informática, entre los usuarios y las metodologías.

Uno de los primeros pasos en este proyecto ha consistido en la búsqueda y análisis de las diferentes bases de datos de poesía en el continente europeo, gracias al cual se ha podido determinar el panorama actual en lo que se refiere a la situación de la poesía digitalizada que circula por la web.

* 1. Estado de la cuestión

Tras una búsqueda exhaustiva se ha realizado una clasificación de los proyectos de poesía digitalizada en Europa y que han tenido interés en unirse al proyecto. El conjunto seleccionado se ha dividido atendiendo a la temática de los corpus tratados en sus webs, quedando la clasificación de la siguiente manera:

<http://postdata.linhd.es/partners/>

Bases de datos de poesía de la Edad Media:

* Répertoire métrique de la poésie lyrique occitane des trobadours à leurs héritiers
* Analecta Hymnica Digitalia (<http://webserver.erwin-rauner.de/crophius/Analecta_conspectus.html>)
* ReMeTca: Repertorio Métrico Digital de la Poesía Medieval Castellana (<http://remetca.uned.es>)
* Bibliografia Elettronica dei Trovatori ([www.bedt.it](http://www.bedt.it))
* MedDB Base de datos da Lírica Profana Galego-Portuguesa ([www.cirp/gal.meddb](http://www.cirp/gal.meddb))
* The Oxford Cantigas de Santa Maria Database (<http://csm.mml.ox.ac.uk/>)
* Le Noveau Naetebus (<http://nouveaunaetebus.elte.hu>)
* Cantigas de Santa Maria for singers (<http://www.cantigasdesantamaria.com>)
* Répertoire de la poésie hongroise ancienne (<http://rpha.elte.hu>)
* Lyrik des hohen Mittelalters (<http://lhm-online.de>)
* Corpus Rhythmorum Musicum (<http://corimu.unisi.it>)
* Digital Edition of the Index of Middle English Verse (<http://dimev.net>)
* Skaldic Poetry of the Scandinavian Middle Ages (<http://skaldic.abdn.ac.uk/db.php>)
* Littera – Cantigas Medievais Galego – Portuguesas
* LIO: Lirica Italiana Degli Origini (<http://mirabileweb.it>)
* The last song of trobadours (<http://icalia.es/trobadours/ca>)
* Corpus des Trobadours (<http://trobadors.iec.cat>)

Bases de datos de poesía clásica:

* Pedecerto ([www.pedecerto.eu](http://www.pedecerto.eu))

Bases de datos de poesía moderna:

* Corpus of Spanish Golden-Age Sonnets (<https://github.com/bncolorado/CorpusSonetosSigloDeOro>)
* Sound and Metre in Italian Narrative Verse (<http://www.italianverse.reading.ac.uk>)
* Eighteenh-Century Poetry Archive (ECPA) (<http://eighteenthcenturypoetry.org>)
* Nederlandse Liederenbank ([www.liederenbank.l](http://www.liederenbank.l))
* Versologie ([www.versologie.cz](http://www.versologie.cz))
* Kalevala poetry (<http://dbgw.finlit.fi/skvr/>)
* Métrique en Ligne ([www.crisco.unicaen.fr/velaine](http://www.crisco.unicaen.fr/velaine))
* Russian Academy of Science (<http://ruscorporaru/>)
* Dainu Skapis (<http://dainuskapis.lv>)
* Eesti Regilaulude Andmebaas (<http://www.folklore.ee/regilaul/andmebaas>)

Si bien es cierto, todas estas bases de datos o portales web no son las únicas circulando por la red, sino simplemente aquellas que han mostrado interés en formar parte de un proyecto todavía mayor.

Si analizamos este conjunto de bases de datos, resalta el hecho de que solo una de ellas recoge un corpus de poesía clásica –latina en este caso-, lo cual deja ya entrever la escasez de portales web y bases de datos dedicados a este tipo de poesía.

En relación con la poesía clásica, de textos griegos y latinos, existen otras páginas web o bases de datos:

* PERSEUS (<http://www.perseus.tufts.edu/hopper/>) Es el portal web de textos clásicos más utilizado y completo entre los existentes. Además de recoger algunos corpus de poesía moderna, cuenta especialmente con un apartado destinado a obras griegas y latinas, con un corpus de lo más completo y exhaustivo. En lo que se refiere a estas obras, ofrece las mismas en versión original, así como una traducción al inglés. Cada una de las palabras de los textos originales constituye un hipervínculo a otra página interna de Perseus, donde se ofrece información gramatical de la misma en forma estadística atendiendo a las diferentes opiniones de los usuarios; así como con su localización en el texto.

El portal cuenta además con diferentes comentarios críticos de las obras.

* BIBLIOTHECA AUGUSTANA (<https://www.hs-augsburg.de/~harsch/augustana.html>). Es un portal web de textos clásicos muy completo entre los existentes. Cuenta con un corpus infinito de obras de la época clásica (griegas, latinas, hispánicas, gálicas, itálicas, etc). Se recogen las obras en su lengua original, pero no ofrece traducción de las mismas a una lengua moderna.
* THE LATIN LIBRARY (<http://www.thelatinlibrary.com/>) Es un portal web de textos clásicos latinos, cuyo corpus radica en el conjunto de obras latinas (la mayoría) sin traducción a ninguna lengua moderna.
* WIKISOURCE (<https://es.wikisource.org/>). En algunos casos encontramos aquí las obras completas de clásicos en lenguas modernas.
  1. Problemas detectados

Este análisis deja entrever la inexistencia de un portal web de obras clásicas con su traducción al español, lo cual deja aislada a la comunidad hispanohablante en este ámbito que debe recurrir siempre o casi siempre a trabajos o portales de origen anglosajón para sus investigaciones.

Asimismo y gracias a un análisis exhaustivo de cada una de las bases de datos participantes en el proyecto POSTDATA se detecta una inconsistencia en la conceptualización de los distintos términos en las bases de datos: esto es, los mismos términos tienen nombres diferentes en las distintas bases de datos, cuyos responsables no siguen un patrón a la hora de conceptualizar los elementos. Este análisis fue realizado de forma individual con especial atención a las bases de datos cuyo corpus recogía poesía de índole clásica; y con la ayuda de otros investigadores del proyecto (Helena Bermúdez, Mariana Curado Malta y Elena González-Blanco, investigadora principal del mismo).

Ver anexo (<https://zenodo.org/record/573767#.WYA0B9Pyi3U>)

1. Objetivos generales / Motivaciones (2 páginas)

En el contexto explicado previamente y en el del proyecto POSTDATA, surge la idea de la creación de un portal web dedicado a la obra de Teognis de Megara, poeta griego del siglo VI a.C, por su complejidad lingüística y por su corpus rico en topónimos y nombres propios, lo cual permite una mayor explotación del sitio creado.

Se realiza, por ello, un análisis de necesidades (ver anexo) que permite fijar los objetivos básicos del proyecto, así como el perfil del grupo meta al que va destinado el mismo.

Se establecen los siguientes objetivos:

* Facilitar a los lectores y estudiosos de la obra de Teognis de Megara la comprensión, interacción e investigación de la misma a través de un portal *user-friendly* con múltiples funcionalidades.
* Aportar a la comunidad académica hispanohablante un recurso más en español en relación con el mundo clásico.
* Aportar una edición única de la obra, con elementos inexistentes hasta ahora como: búsqueda de nombres propios que aparecen en la obra con descripción correspondiente, geolocalización actual de los topónimos de la obra, imagen visual de los tipos de verso atendiendo a la métrica de los mismos; así como otros elementos ya conocidos hasta entonces, como breve descripción de los personajes en el propio texto al pasar el ratón por encima de los mismos, o enlace de cada uno de los términos de la obra en griego con su entrada correspondiente en un diccionario de léxico griego antiguo
* Creación de un pequeño recurso poético en la web con una base de datos estándar desde el punto de vista tecnológico, que conceptualiza los términos seleccionados siguiendo el estándar de la Text Encoding Initiative (TEI)

1. METODOLOGÍA/DESARROLLO (15 páginas)

Para la realización de este trabajo se han seguido distintos pasos y utilizado diferentes tecnologías y lenguajes que han permitido el desarrollo e implementación del portal web como se presenta.

A continuación se mencionan de forma esquemática los distintos procedimientos llevados a cabo que, posteriormente, se desarrollarán en profundidad:

1. Proposición del corpus
2. Análisis de necesidades
3. Realización de una encuesta para el análisis de necesidades
4. Selección del corpus
5. Determinación de las funcionalidades de la página web
6. Comienzo del trabajo
7. Recopilación del texto griego digitalizado (del portal web de perseus)
8. Ante la inexistencia del texto español digitalizado + problemas para escaneo OCR + copia manual del texto en un editor de texto plano (SublimeText 3)
9. Creación de los documentos necesarios en XML
10. Etiquetado de los nombres de personas y topónimos siguiendo TEI (Text Encoding Initiative)
11. Búsqueda de los lugares aparentes en la obra en la actualidad y pequeña descripción
12. Geolocalización en Googlemaps de los topónimos (haciendo uso de la herramienta de GoogleMaps para la creación de mapas personalizados)
13. Búsqueda de información de cada uno de los personajes de la obra.
14. Elaboración de una tabla por cada personaje (nombre en griego, nombre en español, breve descripción)
15. Elaboración de un documento en PDF por cada personaje con la tabla anterior
16. Elaboración de documento HTML inicial del texto en griego y del texto en español a través de una hoja de transformaciones XSLT
17. Adaptación manual de los documentos HTML mencionados anteriormente
18. Comienzo de las funcionalidades:
19. Elaboración del primer buscador de nombres de personas en la obra (xPath)
20. Enlace del buscador con documento previo XML– (diferentes problemas)
21. Creación del buscador para los lugares en la obra / funcionamiento (xPath)
22. Creación de un programa en Phyton para enlace de cada término en griego con un diccionario léxico
23. Creación de los tooltip en el texto en español, con información breve y descriptiva de los personajes principales
24. Contextualización de la página web: vida y obra de Teognis, lengua y métrica. Extracción de la información de XXX
25. Elaboración del HTML con esta información
26. Redacción y elaboración de la guía de usuario para la navegación por la página web
27. Diseño de la página web / cambios en css / importación del framework de Bootstrap.

El proyecto comienza con la proposición del corpus seleccionado, así como con la verificación de que el mismo es interesante y de utilidad para la comunidad académica y para los futuros usuarios –en general- del portal web. Tras realizar un análisis de necesidades que sirvió de ayuda para determinar los objetivos del trabajo y con el fin de asegurar la necesidad y utilidad de la creación de un portal de esta índole se realiza una encuesta a alumnos del grado de Filología Clásica de la Universidad Complutense de Madrid, así como del Máster Interuniversitario en Filología Clásica de las Universidades Complutense y Autónoma de Madrid, cuyos perfiles encajan con el grupo meta del proyecto. Esta encuesta es completada por hasta dieciocho alumnos cuyas respuestas determinaron la necesidad de la creación del portal y de una traducción al español de la obra, así como la utilidad de las diferentes funcionalidades de la página que favorecen la interacción del usuario con la obra.

En esta encuesta se preguntó acerca de la necesidad y utilidad de crear un portal web sobre la obra de Teognis

Se proporciona a continuación un resumen con los gráficos de las preguntas y respuestas realizadas en esta primera encuesta:

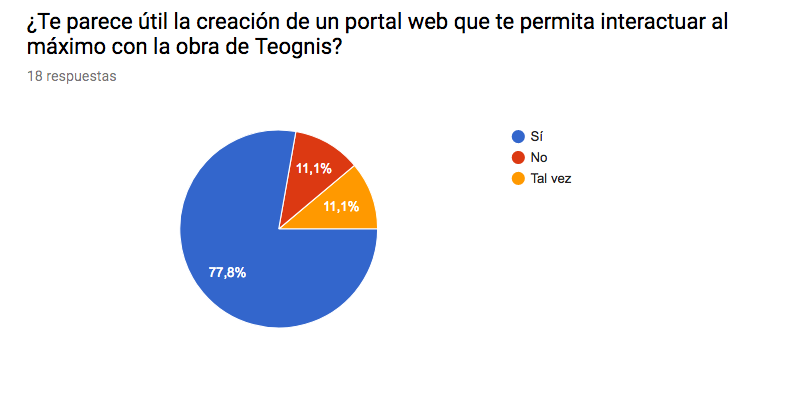


Figura 1. Encuesta AN



Figura 2. Encuesta AN

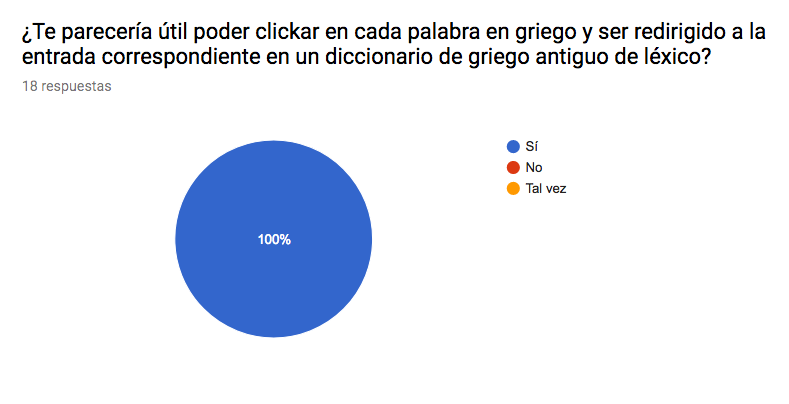


Figura 3. Encuesta AN

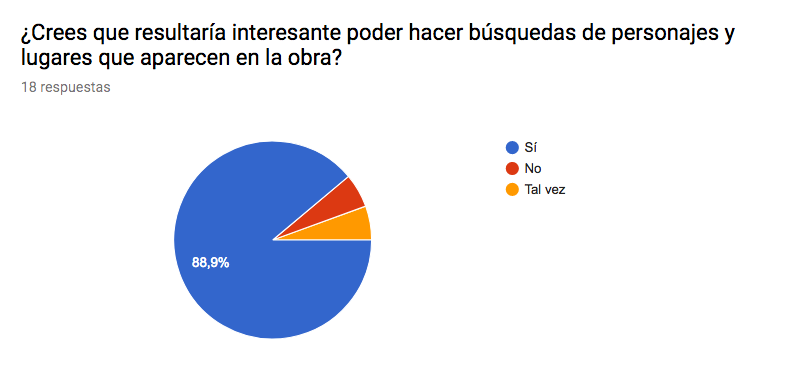


Figura 4. Encuesta AN

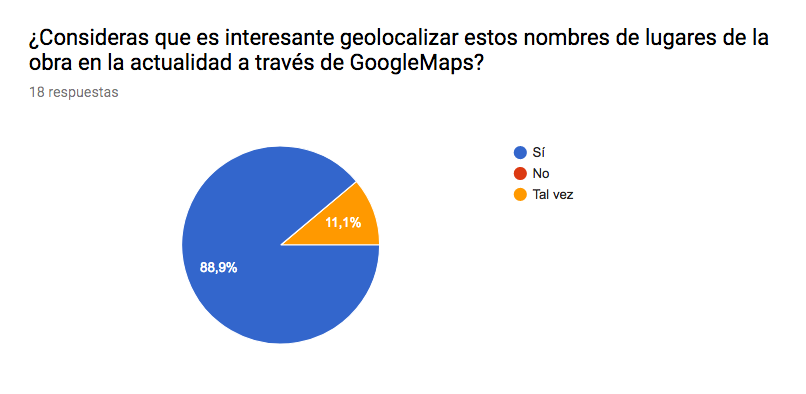


Figura 5. Encuesta AN

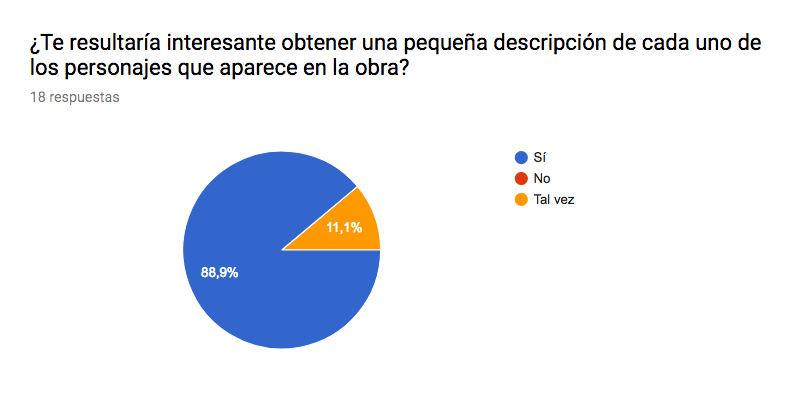


Figura 6. Encuesta AN

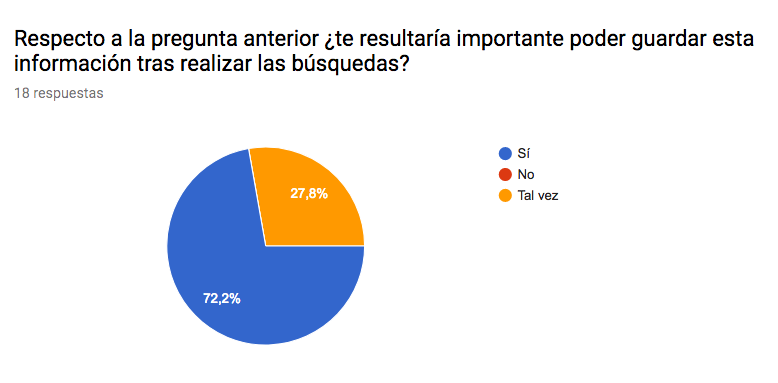
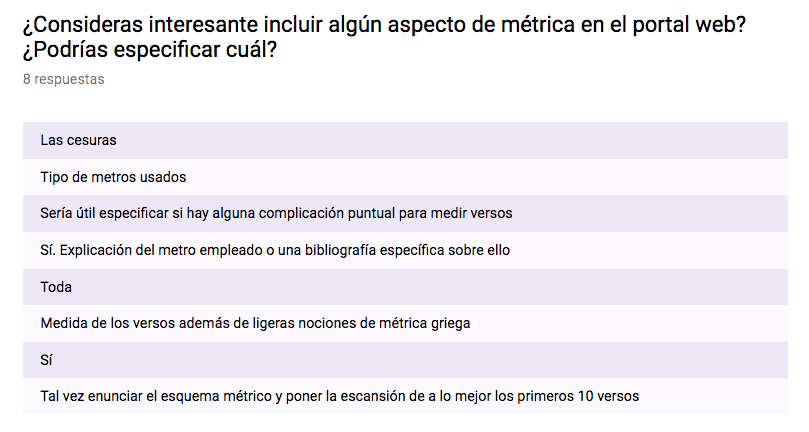
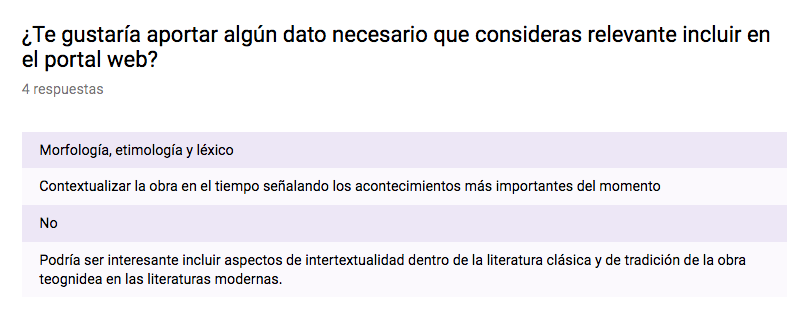


Figura 7. Encuesta AN



Figura 8. Encuesta AN





Tras el análisis de los resultados se determinó que el corpus seleccionado para la página web era el adecuado, por lo que se procedió a la elección de las funcionalidades que tendría la página web. De nuevo, atendiendo a los resultados de la encuesta se fijaron las distintas funcionalidades:

* Buscador de nombres de personajes del texto, con una pequeña descripción de cada uno de ellos.
* Buscador de nombres de lugares del texto, con geolocalización de los mismos a través del framework de GoogleMaps
* Marcador de versos atendiendo a tu tipología métrica (pentámetros, hexámetros o ambos)
* Enlace de cada término en griego con su entrada correspondiente en el diccionario de léxico griego LOGEION.
* Descripción breve de cada personaje al pasar el ratón por encima en el texto en español

Una vez fijado este marco teórico se procedió a comenzar el proyecto desde un punto de vista tecnológico.

En primer lugar se procedió a la obtención de los textos en formato digital: tanto el texto en griego como en castellano. Debido a la fragmentación del Libro II de las Elegías de Teognis, así como a sus fundadas dudas de autoría, se optó por acotar el corpus teognídeo del Libro I del mismo, de mucha mayor extensión, completitud y autoría prácticamente asegurada.

Para la obtención del texto en griego se hizo uso del encontrado en el portal web Perseus.

Por su parte, el texto en español no se encontraba digitalizado ni circulaba por la web en ningún caso. Por ello, se procedió a la transcripción manual del mismo a partir de la edición XXXX, debido a la imposibilidad de utilización de un escáner OCR. Estos textos se escribieron o copiaron en el editor de texto plano Sublime Text .

Una vez se tenían los textos en este editor y de forma digitalizada, se procedió a la elaboración un documento XML con el texto en español, ya que sería aquel en el que se realizarían las búsquedas pertinentes. Se utilizó la codificación UTF-8.

Como se ha comentado previamente, el proyecto cuenta con dos funcionalidades principales en lo que se refiere a búsquedas: un buscador de nombres de personajes y un buscador de topónimos en la obra. Es por esto que se procedió a realizar un etiquetado del XML siguiendo la codificación de la Text Encoding Initiative (TEI) de los nombres de personajes y de lugares. Para ello, se utilizaron las etiquetas persName y placeName el atributo @corresp que posteriormente enlazaría con los documentos en HTML.

REFERENCIAS

* [**DH Poetry Modelling: a Quest for Philological and Technical Standardization**](https://zenodo.org/record/60762)  
  *Rio Riande, Gimena del; Martínez Cantón, Clara Isabel; González-Blanco García, Elena*  
  *Jagiellonian University & Pedagogical University (Cracovia) 2016*  
  *doi:10.5281/zenodo.60765, doi:10.5281/zenodo.60762*
* [**Postdata: Towards publishing European Poetry as Linked Open Data**](http://e-spacio.uned.es/fez/view/bibliuned:363-Egonzalez9)  
  *Curado Malta, Mariana; Centenera, Paloma; González-Blanco García, Elena2016*
* [**POSTDATA – Towards publishing European poetry as linked open data**](http://hdl.handle.net/10400.22/8564)  
  *Curado Malta, Mariana; Gonzalez-Blanco, Elena*  
  *DCMI 2016*
* [**Un nuevo camino hacia las humanidades digitales: el Laboratorio de Innovación en Humanidades Digitales de la UNED (LINHD)**](http://e-spacio.uned.es/fez/view/bibliuned:352-Egonzalez21)  
  *González-Blanco García, Elena*  
  *Universidad Nacional de Educación a Distancia (España). Centro de Investigación de Semiótica Literaria, Teatral y Nuevas Tecnologías Signa: Revista de la Asociación Española de Semiótica, 25, 79-93 2016*
* [**Métrica, clasificación y tecnología en los repertorios métricos**](http://e-spacio.uned.es/fez/view/bibliuned:signa-2016-25-7050)  
  *Rodríguez Gómez, José Luis*  
  *Universidad Nacional de Educación a Distancia (España). Centro de Investigación de Semiótica Literaria, Teatral y Nuevas Tecnologías Signa: Revista de la Asociación Española de Semiótica, 25, 137-167 2016*
* [**TEIScribe: A graphic tool for composing and testing TEI documents in the context of the EVI-LINHD environment**](http://e-spacio.uned.es/fez/view/bibliuned:363-Egonzalez10)  
  *González-Blanco García, Elena2016*
* [**La codificación informática del sistema poético medieval castell ano, problemas y propuestas en la elaboración de un repertorio métrico digital: ReMetCa**](http://hdl.handle.net/10171/35718)  
  *González-Blanco-García, E. (Elena); Martínez-Cantón, C. I. (Clara Isabel); Martos-Pérez, M. D. (María Dolores); Río-Riande, M.G. (María Gimena) del*  
  *Servicio de Publicaciones de la Universidad de Navarra 2014*
* [**EVILINHD, a Virtual Research Environment open and collaborative for DH Scholars**](https://zenodo.org/record/574101)  
  *Elena González-Blanco2017*  
  *doi:10.5281/zenodo.574101*
* [**Researchers’ perceptions of DH trends and topics in the English and Spanish-speaking community. DayofDH data as a case study**](http://e-spacio.uned.es/fez/view/bibliuned:363-Egonzalez7)  
  *Robles-Gómez, Antonio; González-Blanco García, Elena; Ros, Salvador; Rio Riande, Gimena del; Hernández, Roberto; Tobarra, Llanos; Caminero, Agustín; Pastor, Rafael*  
  *Jagiellonian University & Pedagogical University (Cracovia) 2016*
* [**From syllables, lines and stanzas to linked open data: standardization, interoperability and multilingual challenges for digital humanities**](http://e-spacio.uned.es/fez/view/bibliuned:363-Egonzalez11)  
  *González-Blanco García, Elena; Manailescu, Mara; Ros, Salvador2016*
* [**La edición digital de textos literarios: planteamientos y perspectivas de futuro**](http://e-spacio.uned.es/fez/view/bibliuned:352-Egonzalez22)  
  *González-Blanco García, Elena*  
  *Universidad de Navarra RILCE: Revista de filología hispánica, 33 (1), 239-258 2017*
* [**Spanish CLARIN K-Centre**](http://hdl.handle.net/10045/57765)  
  *Bel, Núria; González-Blanco García, Elena; Iruskieta, Mikel*  
  *Sociedad Española para el Procesamiento del Lenguaje Natural 2016*
* [**Boundary Land: Diversity as a defining feature of the Digital Humanities**](http://e-spacio.uned.es/fez/view/bibliuned:363-Egonzalez5)  
  *O'Donnell, Daniel Paul; Bordalejo, Barbara; Murray Ray, Padmini; Rio Riande, Gimena del; González-Blanco García, Elena*  
  *Jagiellonian University & Pedagogical University (Cracovia) 2016*
* [**EVI-LINHD. A Virtual Research Environment for the Spanish-speaking Community**](http://e-spacio.uned.es/fez/view/bibliuned:363-Egonzalez6)  
  *Rio Riande, Gimena del; González-Blanco García, Elena; Martínez Cantón, Clara Isabel; Escribano, J*  
  *Jagiellonian University & Pedagogical University (Cracovia) 2016*
* [**Linked open data to represent multilingual poetry collections. A proposal to solve interoperability issues between poetic repertoires**](https://zenodo.org/record/60767)  
  *González-Blanco, Elena; Del Rio Riande, Gimena; Martinez Cantón,Clara2016*  
  *doi:10.5281/zenodo.60767*
* [**Digital repertoires of poetry metrics: towards a Linked Open Data ecosystem**](http://e-spacio.uned.es/fez/view/bibliuned:363-Egonzalez12)  
  *Curado Malta, Mariana; González-Blanco García, Elena; Martínez Cantón, Clara Isabel; Rio Riande, Gimena del2016*
* [**TEIScribe: A graphic tool for composing and testing TEI documents in the context of the EVI-LINHD environment**](https://zenodo.org/record/161379)  
  *González Blanco García, Elena; Robles-Gómez, Antonio; Ros, Salvador; Del Rio Riande, Gimena; Hernández, Roberto; Urizar, Miguel; Martínez Cantón, Clara; Pastor, Rafael; Cano, Jesús; Caminero, Agustín*  
  *Austrian Centre for Digital Humanities 2016*  
  *doi:10.5281/zenodo.161379*
* [**Analyzing poetry databases to develop a metadata application profile. Why each language uses a different way of modelling?**](https://zenodo.org/record/573767)  
  *Patricia Garrido2017*  
  *doi:10.5281/zenodo.573767*
* [**Linked open data to represent multilingual poetry collections. A proposal to solve interoperability issues between poetic repertoires**](https://zenodo.org/record/60768)  
  *González-Blanco, Elena; Del Rio Riande, Gimena; Martinez Cantón,Clara2016*  
  *doi:10.5281/zenodo.60768*
* [**Domain Model V0.1 of the metadata application profile for European Poetry**](https://zenodo.org/record/437827)  
  *Curado Malta, Mariana; Bermudez Sabel, Helena; Centenera, Paloma2017*  
  *doi:10.5281/zenodo.437827*