

La codificación XML en el ámbito de las Ciencias Historiográficas. Algunas propuestas para su uso y desarrollo

XML Encoding in Historiographical Sciences. Some Proposals for its Use and Development

Dirección

Clara Martínez
Cantón
Gimena del Río
Riande
Ernesto Priani

Secretaría

Romina De León

Víctor CABALLERO GÓMEZ

Universidad Complutense de Madrid

victorcaballero@ucm.es

ABSTRACT

In this paper, we will focus on several projects that linked the eXtensible Markup Language (XML) to Historiographical Sciences, particularly on the scope of text encoding. After this, we will review three of the most frequently used standards for text encoding and we will decide which method is the most accurate for our disciplines. To this end, it is important to bear in mind some issues related to the adaption of the encoding to Paleography and Diplomatics. In addition, when this happens, we will attempt to reach an acceptable solution, as is the case for information related to diplomatics analysis. Finally, we will examine what results can be expected from the transformation of the XML documents, a fact that can serve to convince of the benefits of the text encoding in opposition to other editions that lack this resource.

KEYWORDS

Paleography, Diplomatics, Digital Edition, eXtensible Markup Language (XML), Text Encoding Initiative (TEI).

RESUMEN

En el presente trabajo intentaremos realizar un acercamiento a los distintos proyectos que han vinculado el metalenguaje de marcado XML con las Ciencias Historiográficas en lo relativo a la codificación textual. Revisaremos, posteriormente, tres de los estándares más empleados para la codificación de textos y valoraremos cuál es el más adecuado a nuestras disciplinas, poniendo especial énfasis en los aspectos que susciten una mayor problemática en orden a su adaptación. Para los casos en los que ello suceda, como en la información referida al análisis diplomático, trataremos de encontrar una posible solución. Por último, expondremos cuáles son algunos de los resultados que podemos esperar a partir de la transformación de los documentos XML, un paso que podría ayudar a convencer de los beneficios de la codificación respecto de ediciones que no incluyen este recurso.

PALABRAS CLAVE

Paleografía, Diplomática, edición digital, eXtensible Markup Language (XML), Text Encoding Initiative (TEI).

1. INTRODUCCIÓN

Aunque la tendencia parezca corregirse, siguen siendo pocos los trabajos que, en España y más allá del ámbito filológico, han empleado o se han referido a las diferentes herramientas digitales puestas a disposición de las Humanidades. Los que sí lo han hecho coinciden en concluir la escasa implantación y los limitados rendimientos que se han obtenido de las Humanidades Digitales en áreas como la historia (Ramírez Sánchez, 2016; Bueno Sánchez y Perreux, 2017), tesis que no podemos dejar de suscribir.

Buen ejemplo de ello es el caso que nos ocupa, el de la codificación con diferentes formatos del eXtensible Markup Language (XML), entendido como el modo de “digitally encoding or enriching digital objects with metadata and markup, including categorization of words or information about the objects described” (Hughes, Constantopoulos y Dallas, 2016, p. 161)¹. Este ámbito, como destaca Nyhan (2012), ha sido una pieza clave desde los inicios de la investigación relacionada con las Humanidades Digitales. Ya no es necesario, por su extraordinaria difusión, convencer de las grandes posibilidades que ofrece trabajar en este ámbito. No obstante, sí creemos conveniente, debido a la limitada influencia que ha tenido en los trabajos españoles relacionados con las Ciencias Historiográficas, mostrar qué se ha hecho, cómo se puede avanzar y, muy especialmente, qué podemos conseguir a partir del etiquetado XML en ciencias como la Paleografía o la Diplomática.

En el presente trabajo intentaremos realizar un acercamiento a los distintos proyectos que han vinculado el metalenguaje de marcado XML con las Ciencias Historiográficas en lo relativo a la codificación textual. Presentaremos, posteriormente, tres de los estándares más empleados para la codificación de textos y valoraremos cuál es el más adecuado a nuestras disciplinas, poniendo especial énfasis en los aspectos que susciten una mayor problemática en lo relativo a su adaptación. Para los casos en los que ello suceda, además, trataremos de ofrecer una posible solución. Por último, expondremos cuáles son algunos de los resultados que podemos esperar a partir de la transformación de los documentos XML, un paso que debería resultar especialmente atractivo y que podría ayudar a convencer de los grandes beneficios del marcado XML.

Ofrecemos, por lo tanto, un primer acercamiento al tema y una serie de soluciones a varios problemas que surgen, de una manera inevitable, al hilo de la puesta en práctica de unos recursos de cierta complejidad. También sería deseable que este trabajo invitara a la reflexión y aportara nuevas oportunidades para la investigación en el ámbito de la paleografía, la epigrafía, la diplomática, la numismática e, incluso, de la historia en general. Nuestros objetivos

¹ Una buena descripción de las principales características de XML y TEI puede encontrarse en el trabajo de Alcaraz Martínez y Vázquez Puig (2016).

se verán cumplidos si, además de lo anterior, conseguimos iniciar al lector en un campo tan provechoso como el que nos ocupa, llevándole a otras lecturas, seguramente más provechosas que esta.

2. ALGUNOS PROYECTOS SOBRE CODIFICACIÓN CON XML

De los diferentes proyectos relacionados con la codificación XML, no podemos dejar de referirnos a varios de ellos por su vinculación, directa o indirecta, con las Ciencias Historiográficas. Así, aunque algunos de ellos hayan sido desarrollados en el ámbito de la Filología o de la Archivística, pueden servir igualmente a la investigación en ciencias como la Paleografía, la Diplomática e incluso la Numismática.

Hace ya algunos años, en un trabajo elaborado en el marco del proyecto de investigación *Edición digital del patrimonio documental de Asturias*, se planteaba la dificultad que entrañaba el análisis masivo de los datos contenidos en los repertorios documentales impresos y se consideraba “la posibilidad de que los lenguajes informáticos pudiesen contribuir a un procesamiento detallado de esa información desde las perspectivas de investigación de las Ciencias y Técnicas Historiográficas” (Sanz Fuentes y Calleja Puerta, 2011, p. 379). Apuntaba, finalmente, a la conveniencia del uso del metalenguaje de marcado XML, concretamente a los estándares de la Text Encoding Initiative (TEI) y de la Charters Encoding Initiative (CEI). Tres años antes se había llegado a la misma conclusión en el Archivo Histórico Eclesiástico de Bizkaia, institución desde la que se apostó por la Encoded Archival Description (EAD) para la descripción archivística y por TEI para la codificación de la información paleográfica y diplomática de los documentos transcritos, prueba de ello es la propuesta de Romero Andonegi (2008). A pesar de que los resultados de dichos proyectos ya no son accesibles, sirven para darnos cuenta de la ineludible necesidad de adoptar unos estándares internacionales que nos asistan en la edición y el tratamiento de las fuentes escritas, pero en función de los intereses concretos de las Ciencias Historiográficas.

En ese sentido, aunque ya fuera del ámbito estatal, debemos detenernos en dos grandes plataformas online que deberían servir como punto de partida para el desarrollo de cualquier trabajo en el asunto que nos ocupa. Nos referimos al proyecto *Monasterium*² y al *Codice diplomatico della Lombardia medievale, secoli VIII-XII*³. Más allá de la enorme cantidad de fuentes primarias a las que dan acceso, son importantes aquí por el modo en el que tratan la información. El archivo virtual *Monasterium* (MOM) permite, a través de su editor integrado EditMOM, la codificación XML/CEI de sus documentos (Heinz, 2010) de una manera muy intuitiva. Por su parte, el *Codice diplomatico della Lombardia medievale* permite acceder a la

² Accesible desde: <http://monasterium.net:8181/mom/home>.

³ Accesible desde: <http://www.lombardiabeniculturali.it/cdlm>.

codificación XML/TEI de los documentos que han sido editados en el seno del proyecto (Ansani, 2002). Ambos están ideados desde el ámbito de la diplomática, por lo que han tenido muy en cuenta aspectos clave, sobre todo, para el análisis diplomático de las fuentes, algo que todavía no ha sido considerado de un modo suficiente por el consorcio TEI (2019) en sus *Guidelines for Electronic Text Encoding and Interchange*, a pesar de contemplar un módulo específico para la descripción de manuscritos.

Desarrollada por el proyecto *Recognition and Enrichment of Archival Documents* (READ) de la Universitat Politècnica de València y orientada al reconocimiento de texto manuscrito, la plataforma Transkribus⁴ permite el etiquetado de la información relativa a los manuscritos que conviene a la Paleografía y a la Diplomática, siendo su uso extensible a muchas otras disciplinas. Su gran ventaja es la posibilidad de trabajar con cualquier documento y la vinculación línea a línea entre la imagen digital importada y el panel de transcripción. El resultado de esa edición se puede exportar a varios formatos, entre los que se encuentra el XML/TEI.

Debemos fijar nuestra atención, igualmente, en algunos proyectos españoles cuyos criterios y objetivos son de naturaleza filológica. Es el caso de los diferentes corpus online que han empleado XML para su codificación. Sus aplicaciones, además de en la Filología o la Paleografía, caben rastrear incluso en trabajos históricos, cuyas hipótesis pueden “apoyarse sobre bases más empíricas cuando se dispone de esta masa textual en un solo corpus, y de las aplicaciones informáticas adecuadas para realizar un análisis más complejo” (Carrasco Manchado, 2011, p. 345). Por su gran difusión y por la gran cantidad de documentos que contiene destacamos, en primer lugar, el *Corpus Hispánico y Americano en la Red: Textos Antiguos* (CHARTA)⁵, compuesto por una serie de subcorpus como el *Corpus de Documentos Españoles Anteriores a 1.700* (CODEA) o el *Corpus Histórico del Español Norteño* (CORHEN).

El motor de búsqueda de CHARTA permite recuperar determinada información en función del etiquetado que se le ha aplicado a los textos, lo que evita buena parte del ruido documental. Así, con una rápida búsqueda, obtenemos un total de 518 documentos en los que hay algún tipo de elemento al margen de la caja de escritura. La principal carencia, de cara a las Ciencias Historiográficas, es la falta de etiquetas relacionadas con el análisis diplomático. De este modo, entre las marcas con contenido diplomático solo encontramos algunas relacionadas con la invocación como crismón o cruz y otras relativas a la validación como firma, rúbrica, rueda, sello y suscripción. Su gran beneficio, en cambio, es el elevado nivel de acceso a la documentación que permite al usuario, especialmente en comparación con el *Corpus*

⁴ Accesible desde: <https://transkribus.eu/Transkribus>.

⁵ Accesible desde: <http://www.corpuscharta.es/consultas.html>.

*Documentale Latinum Cataloniae*⁶ o con el *Corpus Diacrónico del Español*⁷, inconveniente ya señalado, entre otros, por Vaamonde (2015) o Rodríguez Puente (2018)⁸.

Todos ellos, así como muchos otros, tienen un innegable valor para los previsibles futuros trabajos relacionados con la edición digital en el ámbito de las Ciencias Historiográficas, ya sean como herramienta, como modelo o como fuente para el acceso a una muy valiosa información. Su uso y consulta son altamente recomendables para los investigadores de nuestras disciplinas, pues ofrecen una serie de posibilidades incomparables, no solo en términos cuantitativos, sino sobre todo en términos cualitativos, algo derivado inequívocamente del proceso de etiquetado de los textos.

3. ALGUNAS CONSIDERACIONES SOBRE EL ETIQUETADO TEI

A pesar de que exista un estándar específico para la codificación de textos en el ámbito de la Diplomática, es evidente la ventaja que ha obtenido a lo largo de los años el lenguaje TEI. Son muchos, además, los proyectos que incluso orientados a estudios de carácter diplomático no recurren a CEI, sino a las especificaciones de la TEI. Por eso creemos conveniente empezar ocupándonos de él. Como ya advertimos previamente, no está diseñado específicamente para el ámbito de las Ciencias Historiográficas, por lo que se han desarrollado diversas adaptaciones para aquellos elementos cuyo etiquetado planteaba dificultades. Analizaremos aquí cuáles son los principales problemas y cuáles, a nuestro juicio, las mejores soluciones⁹.

El primero del que debemos ocuparnos, puesto que ya nos hemos referido a él, es el relativo al etiquetado de los diferentes elementos del análisis diplomático. En el *Codice diplomatico della Lombardia medievale*, atendiendo a las indicaciones de Ansani (2000-2019), se introducen etiquetas ajenas a las *Guidelines* de la TEI (2019) para dicha tarea. Así, encontramos tags como <PROTOCOLLO> o <INVOCATIO>. Siguiendo este modelo, Romero Andonegi (2008) propone etiquetas tales como <texto> y <notificatio>. Otra solución arbitrada es recurrir a la etiqueta <div> (text division), incluida en el estándar TEI, e introducir el nombre del elemento como un atributo @type. Así lo prefiere Vogeler (2005) y de ese modo podemos codificar la información relativa al análisis diplomático con las etiquetas que ofrece por defecto la aplicación Transkribus, a la que nos referimos anteriormente. De acuerdo con su definición,

⁶ Accesible desde: <http://gmlc.imf.csic.es/codolcat>.

⁷ Accesible desde: <http://corpus.rae.es/cordenet.html>.

⁸ Asimilables al Corpus CHARTA, tanto en el uso de XML/TEI como en sus ventajas e inconvenientes, encontramos los recursos del Hispanic Seminary of Medieval Studies como el *Old Spanish Textual Archive* (Gago Jover y Pueyo Mena, 2018). Igualmente, aunque ya no dedicados a los textos hispánicos, debemos destacar plataformas como el *Oxford Text Archive* (<http://ota.ox.ac.uk/>) o el *Medieval Nordic Text Archive* (https://menota.org/EN_forside.xhtml).

⁹ Un desarrollo más exhaustivo, en el que se recojan todas las especificaciones necesarias para el proceso de etiquetado, debe esperar a un proyecto de mayor envergadura, dentro del cual se opte por unos criterios u otros en función de los objetivos de las ediciones digitales en el ámbito de las ciencias historiográficas.

puede parecer más adecuado, sin embargo, emplear la etiqueta `<seg>` (arbitrary segment) y el mismo atributo `@type`. Es la propuesta de Burkard, Vogeler y Gruner (2007) y lo que parecen apuntar desde CHARTA (Isasi Martínez, Lobo Puga, Martín Aizpuru, Pérez Isasi, Pierazzo, y Spence, 2014). En el documento del Archivo Histórico Nacional, Diversos-Comunidades, car. 67, n. 215, que utilizaremos, a partir de aquí, para ilustrar las diferentes opciones elegidas, resultaría:

```
<seg type="protocol">
<seg type="invocatio"> Yhesus. </seg>
</seg>
```

Especial cuidado debe tenerse en la delimitación de líneas, puesto que, si siguiéramos el modelo que se suele emplear, por ejemplo, en el etiquetado de poemas, generaríamos un solapamiento de las etiquetas, debido a la superposición con el elemento `<seg>`. Por ello, sea cual sea, debe optarse por una etiqueta vacía. De ese modo, estaríamos marcando el inicio de la línea, por lo que `<lb>` (line beginning) nos parece la etiqueta más adecuada. Es la solución a la que llega, por ejemplo, Ansani (2000-2019). Idéntico caso se produciría con las divisiones en folios, por lo que proponemos, para dicho supuesto, la etiqueta `<pb>` (page beginning). Se deberían complementar con un atributo `@n` que indicara el número de línea o de folio. El resultado sería:

```
<lb n="1"/>
<seg type="notificatio"> Sia a todos maniffiesto que </seg>
```

Respecto de las abreviaturas y su tratamiento, nos alejamos de las soluciones que se siguen habitualmente para atender, estrictamente, al estándar TEI, que contempla la etiqueta `<choice>`, con la que podemos agrupar varias codificaciones alternativas para el mismo punto del texto. Dentro de `<choice>` queda espacio para las etiquetas `<abbr>` (abbreviation) y `<expan>` (expansion):

```
<choice>
<abbr> q </abbr>
<expan> que </expan>
</choice>
```

El último de los aspectos en el que nos queremos detener es el relativo a los nombres. Es habitual etiquetar los nombres de personas y de lugares, pero a nuestro juicio es conveniente incluir el de instituciones y el de monedas, debido esto último a la importancia que tienen, como señala Santiago Fernández (2012), las fuentes documentales para la mayoría de las líneas de investigación relacionadas con la historia de la moneda. Unificando los criterios y adecuándolos del mayor modo posible a las *Guidelines* de la TEI, creemos conveniente emplear la etiqueta `<name>` (*name*, *proper noun*), completada con el atributo `@type`, que especifique si se trata de una persona (*person*), de un lugar (*place*), de una organización o una institución (*org*) o si se trata

de una moneda (*coin*). Este criterio unificador facilitaría la generación de índices y la posibilidad de añadir nuevos tipos para la etiqueta:

```
<name type="coin">
  <choice>
    <abbr> sols </abbr>
    <expan> soldos </expan>
  </choice>
</name>
```

La elección de un criterio u otro, en última instancia, siempre debe quedar condicionada por las necesidades de la edición que se quiera realizar, pero atendiendo, hasta donde sea posible, a las *Guidelines*, criterio que hemos respetado en estas consideraciones. Para el resto de los elementos a etiquetar, que previsiblemente presentan menos complicaciones, cabe recurrir a dicho documento.

4. OTRAS SOLUCIONES: LOS LENGUAJES CEI Y EPIDOC

Basados en el metalenguaje XML, según hemos apuntado previamente, cabe recurrir a otros lenguajes de marcado distintos de la TEI, desarrollados para casos específicos de nuestras disciplinas, como CEI para la Diplomática o EpiDoc para la Epigrafía. Igualmente, y aunque no nos ocupemos de él, no podemos dejar de recordar el estándar EAD, por su gran relevancia en el ámbito de la archivística.

CEI¹⁰ tiene una menor difusión respecto del lenguaje TEI, por lo que, si se apuesta por él, es conveniente respetar, en todo lo posible, el estándar TEI, lo cual facilitaría el intercambio de datos. La gran ventaja que posee es la menor dificultad en el proceso de etiquetado, ya que contempla elementos¹¹, sobre todo relacionados con el análisis diplomático, de los que no disponemos en TEI (Burkard et al., 2007). Así, contamos con tags comunes con TEI y que ya hemos visto aquí, como <abbr> o <expan>, pero también con algunas etiquetas específicas, como <protocol> o <invocatio>, que resuelven algunos de los problemas a los que nos enfrentamos en el apartado anterior. Además de todo esto, su adopción por parte del proyecto Monasterium, uno de los más relevantes a nivel europeo en el terreno de la diplomática, lo acredita como una opción a tener muy en cuenta.

Equivalentes problemas encontramos a la hora de realizar una codificación con TEI en el ámbito de la epigrafía. EpiDoc es un lenguaje de marcado basado en TEI y que ha generado sus propias *Guidelines*¹². Así, entre otras muchas, encontramos una solución para un hipotético texto borrado y restaurado con posterioridad gracias a la etiqueta <supplied>, que se podría complementar con el atributo @reason, cuyo valor sería *lost*. EpiDoc, no obstante, mantiene

¹⁰ Accesible desde: <https://www.cei.lmu.de/index.php>.

¹¹ La lista completa de etiquetas puede consultarse en: <https://www.cei.lmu.de/taglib>.

¹² Accesible desde: <http://www.stoa.org/epidoc/gl/latest/>.

todas las etiquetas propias de TEI que sirven para la codificación en epigrafía, por lo que se facilita la labor de intercambio de datos.

5. OBTENCIÓN DE RESULTADOS

Probablemente, el aspecto que menos se haya tratado, en el ámbito de las Ciencias Historiográficas, sea aquel relacionado con la generación de resultados visualmente atractivos a partir de los documentos a los que se les ha aplicado algún tipo de etiquetado, ya sea en lo relativo a la presentación de los contenidos del texto, en lo relacionado con la exploración de los metadatos o en lo referente a la visualización de aspectos analíticos y cuantitativos. Es este un campo que requiere mayores conocimientos técnicos y, casi sin remedio, la asistencia de un ingeniero informático, pero que puede bastar para convencer de la conveniencia de emplear un lenguaje de marcado en las ediciones digitales. Para ello, podemos recurrir, entre otras opciones, al lenguaje de transformación eXtensible Stylesheet Language (XSL).

Gracias al diseño de hojas de estilo XSL somos capaces de realizar, de un modo automático, una gran variedad de transformaciones a partir de los datos que hemos introducido en nuestra codificación. Estos resultados, además, pueden tener diferentes formatos, como TeX, PDF o DOCX. Para el estándar TEI encontramos algunas hojas de estilo alojadas en la plataforma GitHub¹³. En el caso de EpiDoc es el propio consorcio el que ha desarrollado una serie de hojas de estilo para la transformación de los documentos¹⁴.

El paso más elemental, en este sentido, es el de obtener una forma de presentación adecuada del documento, de acuerdo a los criterios de edición deseados y exportable a cualquier plataforma para su inclusión en una página web o en un trabajo académico. Para proyectos más amplios puede ser interesante la creación de un corpus informático (que puede, si se prefiere, no estar disponible online), al cual se implemente un motor de búsqueda al estilo del que se incluye en el Corpus CHARTA y que permita, de ese modo, la búsqueda sencilla o la búsqueda por etiquetas determinadas, como aquellas relativas al análisis diplomático a las que nos hemos venido refiriendo. Ello, ya en lo relacionado con la exploración de los metadatos, simplificaría trabajos relacionados con las tipologías documentales o con un elemento determinado de la estructura diplomática, como el que se ocupa de la intitulatio de la documentación regia en los reinados de Pelayo a Alfonso VI (Cabezas Fontanilla y Ávila Seoane, 2010). En esa misma línea y tendente a la visualización de los aspectos analíticos, podemos generar, con una hoja de estilo XSL, un HTML con el esquema diplomático. En el caso del documento que hemos etiquetado previamente de acuerdo al estándar TEI, el resultado podría ser el que se muestra en la tabla 1.

¹³ Accesible desde: <https://github.com/TEIC/Stylesheets>.

¹⁴ Accesible desde: <https://sourceforge.net/p/epidoc/wiki/Stylesheets/>.

AHN, Diversos, Comunidades, Car. 67, N. 215
Esquema diplomático

Protocolo	Invocación Notificación Intitulación
Texto	Dispositivo Dirección (inserta) Exposición Cláusula corroborativa
Escatocolo	Data tópica Data cronológica Testigos Suscripción

Tabla 1. Esquema diplomático generado a partir de una hoja de estilo XSL.

Otro elemento que ha suscitado nuestro interés ha sido el de las abreviaturas, debido fundamentalmente a la relevancia que estas adquieren en el ámbito de la paleografía, hecho que se pone de manifiesto en trabajos de carácter braquigráfico como los de Galende Díaz (2014) o Cuenca Muñoz (2017). De este modo, podemos generar listas con las abreviaturas que se emplean en los documentos que hayamos etiquetado, indicando cuáles son las letras que aparecen en el manuscrito y cuál es su desarrollo, así como la línea en la que estas se encuentran. El ejemplo con el documento que hemos etiquetado a lo largo de estas páginas puede verse en la tabla 2.

AHN, Diversos, Comunidades, Car. 67, N. 215
Lista de abreviaturas

Abreviatura	Desarrollo	Línea
albara	albarán	8
Arago	Aragón	15
Comnidat	Comunidat	2 y 5
e	en	7
fuero	fueron	10
genal	general	5
Joha	Johan	4
mosse	mossén	11
not	notario	1, 4 y 14
porq	porque	7
psent	present	6 y 8
psentes	presentes	11 y 16
q	que	1
sols	soldos	6
Yhs	Yhesus	0

Tabla 2. Lista de abreviaturas generada a partir de una hoja de estilo XSL.

A pesar de que no nos hayamos ocupado de ello en este trabajo, debido al extenso tratamiento que ofrece el consorcio TEI (2019) en sus *Guidelines*, también pueden ser relevantes

para el ámbito paleográfico el etiquetado y las transformaciones realizadas en relación con la identificación de características gráficas que ayuden a la distinción de las diversas manos que elaboran un código. Igualmente, puede ser adecuada la aplicación del *geoparsing*¹⁵, de cuya descripción y usos da buena cuenta García Gómez (2018), al estudio relativo, entre otros, a la introducción de los diferentes ciclos escriturarios en determinados territorios o, en el ámbito numismático, a la circulación monetaria.

6. CONCLUSIONES

Como hemos podido apreciar a lo largo de este trabajo, la codificación XML de los textos y las ediciones digitales surgidas a partir de esta han conseguido abrirse paso, aunque quizá no con toda la importancia que cabría esperar, en el ámbito de las Ciencias Historiográficas. Su presencia y consolidación todavía son asignaturas pendientes en el ámbito académico español y la apuesta por su desarrollo puede aportar resultados muy positivos, tanto cuantitativamente, en el sentido de que se puede procesar mucha más información, como cualitativamente, en cuanto que la información procesada puede relacionarse de un modo mucho más efectivo.

Creemos cumplidos los objetivos principales que nos proponíamos al comenzar: detectar algunos problemas y proponer algunas soluciones relativas al etiquetado en Ciencias Historiográficas, así como ofrecer, de un modo visualmente atractivo, una serie de posibilidades de obtención de resultados a partir de la codificación XML. Sin embargo, aún queda mucho por hacer en este campo y solo será posible a través del trabajo conjunto y multidisciplinar, como viene siendo la tónica habitual en los proyectos relacionados con las Humanidades Digitales. Queda, por lo tanto, el fin último de este trabajo, condicionado a la futura apuesta por los diferentes campos abiertos aquí.

REFERENCIAS BIBLIOGRÁFICAS

- Alcaraz Martínez, R. y Vázquez Puig, E. (2016). TEI: un estándar para codificar textos en el ámbito de las Humanidades Digitales. *BiD: textos universitaris de biblioteconomia i documentació*, 37. doi:[10.1344/BiD2016.37.25](https://doi.org/10.1344/BiD2016.37.25).
- Ansani, M. (s.f.). *Procedure di lavoro: la codifica XML dei testi*. Recuperado de <http://www.lombardiabeniculturali.it/cdlm/progetto/codifica-xml> el 20/08/2019.

¹⁵ Técnica que consiste en la identificación de los topónimos que aparecen en el texto y en la asignación de las coordenadas correspondientes a cada topónimo, en orden al procesamiento automático de la información geográfica.

- _____ (2002). Il Codice diplomatico digitale della Lombardia medievale: note di lavoro. En D. Puncuh, *Comuni e memoria storica: alle origini del comune di Genova. Atti del Convegno di studi* (pp. 23-49). Genova: Società Ligure di Storia Patria.
- Bueno Sánchez, M. y Perreux, N. (2017). La doble ruptura: problemas epistemológicos a partir de la construcción de una base de datos histórica relativa a la conversión forzada. *Revista de Humanidades Digitales*, 1, 259-286. doi:[10.5944/rhd.vol.1.2017.16765](https://doi.org/10.5944/rhd.vol.1.2017.16765)
- Burkard, B., Vogeler, G. y Gruner, S. (2007). Informatics for Historians: Tools for Medieval Document XML Markup, and their Impact on the History-Sciences. *Journal of Universal Computer Science*, 14(2), 193-210. Recuperado de <http://hdl.handle.net/2263/9963> el 18/08/2019.
- Cabezas Fontanilla, S. y Ávila Seoane, N. (2010). La intitulación en la documentación regia: de Pelayo a Alfonso VI. En A. García Leal, R. Gutiérrez González y C. E. Prieto Entrialgo (Coords.), *MC aniversario de la muerte de Alfonso III y de la tripartición del territorio del Reino de Asturias*, 1 (pp. 263-312). Oviedo: Universidad de Oviedo.
- Carrasco Manchado, A. I. (2011). Nuevas herramientas para la historia de la Edad Media hispánica: los corpus textuales informatizados. En *la España medieval*, 34, 343-372. doi:[10.5209/rev_ELEM.2011.v34.36304](https://doi.org/10.5209/rev_ELEM.2011.v34.36304)
- Cuenca Muñoz, P. (2017). Las abreviaturas en la escritura latina. En J. de Santiago Fernández y J. M. de Francisco Olmos (Eds.), *Homenaje a María Ruiz Trapero* (pp. 311-333). Madrid: Museo Casa de la Moneda.
- Gago Jover, F. y Pueyo Mena, F. J. (2018). El Old Spanish Textual Archive, diseño y desarrollo de un corpus de textos medievales: lematización y etiquetado gramatical. *Scriptorium Digital*, 7, 25-35. Recuperado de <https://bit.ly/2z8XRKx> el 20/08/2019.
- Galende Díaz, J. C. (2014). *Diccionario de abreviaturas en español*. Madrid: Síntesis.
- García Gómez, S. (2018). Del texto literario al mapa digital. Las posibilidades de la georreferenciación en los estudios literarios. *Revista de Humanidades Digitales*, 2, pp. 120-141. doi:[10.5944/rhd.vol.2.2018.22141](https://doi.org/10.5944/rhd.vol.2.2018.22141).
- Heinz, K. (2010). Monasterium.net – Auf dem Weg zu einem europäischen Urkundenportal. En T. Kölzer, W. Rosner y R. Zehetmayer (Eds.), *Regionale Urkundenbücher. Die Vorträge der 12. Tagung der Commission Internationale de Diplomatique* (pp. 139-145). St. Pölten: NÖ Landesarchiv. Recuperado de <https://bit.ly/2KKKIDf> el 20/08/2019.
- Hughes, L., Constantopoulos, P. y Dallas, C. (2016). Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines. En S. Schreibman, R. Siemens y J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 150-170). Chichester/Malden: Wiley-Blackwell.

- Isasi Martínez, C., Lobo Puga, A., Martín Aizpuru, L., Pérez Isasi, S., Pierazzo, E. y Spence, P. (Coords.) (2014). *Guía para editar textos CHARTA según el estándar TEI: una propuesta*. Recuperado de <https://bit.ly/30jCvpL> el 20/08/2019.
- Nyhan, J. (2012). Text Encoding and Scholarly Digital Editions. En C. Warwick, M. Terras y J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 117-137). Londres: Facet Publishing, UCL Centre for Digital Humanities.
- Ramírez Sánchez, M. (2016). Las ciencias y técnicas historiográficas en el contexto de las Humanidades Digitales : oportunidades para su desarrollo. En A. Marchant Rivera y L. Barco Cebrián (Eds.), *“Dicebamus hesterna die...”*. Estudios en homenaje a los profesores Pedro J. Arroyal Espigares y M^a Teresa Martín Palma (pp. 365-392). Málaga: Encasa. Recuperado de <https://riuma.uma.es/xmlui/handle/10630/11055> el 20/08/2019.
- Rodríguez Puente, P. (2018). En busca de lo hablado en lo escrito en los corpus diacrónicos del español: una comparativa con los corpus anglosajones. *E-Scripta Romanica*, 5, 89-127. Recuperado de <https://bit.ly/2KZDWTE> el 20/08/2019.
- Romero Andonegi, A. (2008). Nuevas herramientas para el estudio de la diplomática: los lenguajes de marcado TEI y EAD. *Documenta & Instrumenta*, 6, 91-118. Recuperado de <https://bit.ly/2P4UMGZ> el 20/08/2019.
- Santiago Fernández, J. de. (2012). Reflexiones sobre la investigación y estudio de la moneda en la Edad Moderna. En M. T. Muñoz Serrulla (Coord. y Ed.), *La Moneda: Investigación numismática y fuentes archivísticas* (pp. 97-115). Madrid: Asociación de Amigos del Archivo Histórico Nacional-UCM. Recuperado de <https://bit.ly/2zeS2ex> el 20/08/2019.
- Sanz Fuentes, M. J. y Calleja Puerta, M. (2011). Edición digital del patrimonio documental de Asturias (I). Diseño de la aplicación y digitalización de los documentos medievales publicados. En E. Cantarell Barella y M. Comas Via (Coords.), *La escritura de la memoria: los registros* (pp. 379-385). Barcelona: PPU.
- TEI Consortium. (2019). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Recuperado de <https://tei-c.org/Guidelines/P5/> el 20/08/2019.
- Vaamonde, G. (2015). Limitaciones en el uso de corpus diacrónicos del español. Nuevas aportaciones desde el proyecto de investigación Post Scriptum. *E-Aesla*, 1. Recuperado de <https://bit.ly/33IEar3> el 20/08/2019.
- Vogeler, G. (2005). Towards a Standard of Encoding Medieval Charters with XML. *Literary and linguistic computing*, 20(3), 269-280. doi:[10.1093/lc/fqi031](https://doi.org/10.1093/lc/fqi031).