

Introducción a la Text Encoding Initiative. Definición, aplicaciones prácticas y recursos*

Susanna Allés Torrent

2019

Esta lección ofrece una primera aproximación a la Text Encoding Initiative, un marco de trabajo, utilizado internacionalmente y por una gran comunidad de usuarios, concebido especialmente para la codificación, la explotación informática, la publicación en línea y la preservación de textos en ciencias sociales y humanidades. Para ello, se tratan conceptos generales como la codificación de los textos, su utilidad o los campos de aplicación.

I. Codificación de los textos y TEI

En el ámbito de las humanidades y las ciencias sociales, el texto representa, en la mayoría de los casos, el punto de partida de muchas investigaciones. La introducción de métodos y herramientas propios de la ciencia informática llevaron hace ya más de veinte años a replantearse la importancia del texto y su procesamiento informático. La gran cantidad de textos a nuestra disposición conlleva la adopción de nuevas estrategias para la creación, la explotación y su conservación.

La iniciativa de marcado de textos nace pues de la necesidad de procesar los textos desde un punto de vista informático, para que, en definitiva, no sea susceptible solo de una lectura “plana”, sino para que sea procesable a través de múltiples herramientas como pueden ser programas de concordancias, análisis estilométricos, gráficos de frecuencias, segmentación gramatical o, sobretodo, su publicación web y la posibilidad de realizar búsquedas inteligentes.



Figura 1: Logo de la Text Encoding Initiative

*Estos materiales fueron creados en el marco del certificado de [Experto Universitario en Humanidades Digitales](#), ofrecido desde 2014 por el Laboratorio de Innovación en Humanidades Digitales de la Universidad Nacional de Educación a Distancia.

La *Text Encoding Initiative* surgió a finales de los años ochenta y fue la primera iniciativa científica destinada a la codificación informática de textos, actualmente es una de las más utilizadas y uno de los temas centrales en la disciplina de las humanidades digitales.

¿Qué es TEI?

La Text Encoding Initiative se ampara en primera instancia en el *Consortio TEI*, cuya estructura veremos más abajo. Este consorcio compuesto por personas e instituciones es el responsable de desarrollar y mantener actualizadas las *Guías directrices* para el marcado de los textos en formato digital. Estas guías son una especie de manual de codificación o de recomendaciones de buenas prácticas que especifican una metodología concreta de codificación de los textos de manera que sean leídos y procesados por las computadoras.

TEI posee una plataforma web donde se publican tanto las *Guías directrices*, como una gran cantidad de herramientas e informaciones sobre su comunidad. A continuación, podemos leer una primera definición que nos será útil para empezar:

[English] The *TEI Guidelines for Electronic Text Encoding and Interchange* define and document a markup language for representing the structural, renditional, and conceptual features of texts. They focus (though not exclusively) on the encoding of documents in the humanities and social sciences, and in particular on the representation of primary source materials for research and analysis. These guidelines are expressed as a modular, extensible XML schema, accompanied by detailed documentation, and are published under an open-source license. The Guidelines are maintained and developed by the TEI Consortium, through its Technical Council, with the support and participation of the TEI community.

[Español] “Las *Guías Directrices para la codificación y el intercambio de textos electrónicos* definen y documentan un lenguaje de marcado para la representación de las características estructurales, físicas y conceptuales de los textos. Su foco de atención es el marcado de documentos en Humanidades y Ciencias Sociales (aunque no exclusivamente), y en particular la representación de fuentes primarias para su explotación y análisis. Estas líneas directrices se expresan de forma modular, a través de esquemas XML extensibles, están acompañadas de una documentación detallada, y publicadas bajo una licencia en acceso abierto. Las *Guías Directrices* son mantenidas y desarrolladas por el Conaoexio TEI, a través de su Comité Técnico, con el respaldo y la participación de la comunidad TEI”.
<http://www.tei-c.org/Guidelines/>

Veamos paso por paso los diferentes puntos básicos de esta definición:

- El núcleo central de la TEI son las *Guías Directrices* que constituyen el manual de uso y la recomendación de buenas prácticas sobre cómo proceder. Su misión principal es proponer unas pautas claras para el marcado y el intercambio electrónico de los textos y ofrecer una documentación general y explicativa.
- Las *Guías Directrices* proponen un tipo concreto de lenguaje de marcado que consiste

en aislar a través de “marcas” o “etiquetas” las características textuales, ya sean estructurales (como capítulos de libro, secciones, apartados, párrafos, estrofas, versos, etc.), como físicas (distribución de la página en columnas en un manuscrito), como conceptuales o semánticas (nombres de personas, de lugares, palabras clave, etc.).

- TEI es especialmente usada en las disciplinas humanísticas, ciencias sociales o lingüística, y en la representación de las fuentes primarias, tales como manuscritos (de hecho, TEI es también utilizada en algunas bibliotecas para la descripción y catalogación de sus fondos).
- El sistema de marcado propuesto por TEI tiene una naturaleza modular que agrupa fenomenología diversa en un mismo “módulo”. Tiene cuatro obligatorios, que corresponden a elementos compartidos por todos y cada uno de los documentos TEI (como puede ser el elemento raíz, o el elemento párrafo). Esta característica modular permite customizar o personalizar el modelo de marcado que cada proyecto necesita.
- TEI se expresa a través del lenguaje estándar web XML, Extensible Markup Language o Lenguaje de Marcas Extensible, que se ve en detalle en la lección “[El lenguaje XML \(eXtensible Markup Language\) y conceptos generales](#)”. Además, por lo general, los esquemas o modelos de marcado también utilizan este mismo lenguaje.
- TEI ofrece una gran cantidad de documentación detallada sobre cada uno de los elementos, módulos y en general cualquier aspecto relacionado con su uso y buenas prácticas.
- Todo este material es de acceso abierto y puede consultarse libremente online, descargarse, reutilizarse e, incluso, ser susceptible de mejoras a través de la retroalimentación de los usuarios. Todo el material está disponible en [GitHub](#).

En lo que concierne a los objetivos de TEI, podríamos resumirlos de la siguiente manera:

- Ofrecer unas guías directrices para la creación y la manipulación de todo tipo de datos textuales. Los datos pueden proceder de textos de cualquier naturaleza, manuscritos, documentos de archivos, inscripciones, correspondencia, novelas, obras dramáticas, etc.
- Promover la creación, el intercambio y la integración de los datos textuales informatizados.
- Contemplar textos de cualquier tipología, género y disciplina (Ciencias Humanas y Sociales, Humanidades, Lingüística), en cualquier lengua y de cualquier periodo cronológico.
- Ser accesible tanto a un público novel, sin conocimientos previos en informática, que quiera codificar un material textual, como a un público técnico, capaz de buscar y proponer nuevas soluciones informáticas.

Las ventajas que ofrece el uso de TEI son múltiples y entre ellas podemos señalar:

- XML-TEI no depende de ningún software o programa informático y por tanto es gratuito

e independiente. Un documento XML será siempre el mismo en cualquier sistema operativo o aplicación. Este hecho evita que se puedan dar situaciones catastróficas, en que, por ejemplo, un programa con un formato propietario ya no se mantenga o no se utilice y los documentos queden obsoletos.

- XML-TEI ha sido diseñado por y para la comunidad científica que es la encargada en última instancia de promover y mejorar las guías de marcado. Cuantos más usuarios y más gente implicada haya utilizando el mismo sistema de marcado, más se avanzará hacia una propuesta sólida de codificación.
- Un marcado a través de un estándar web, independiente de cualquier software o plataforma web, como es el caso de en XML-TEI, permite y facilita la reutilización del mismo material: en diferentes formatos, en contextos diferentes, por diferentes usuarios. Ello permite que proyectos posteriores o contemporáneos puedan establecer un vínculo y un uso diferente del material publicado, evitando el tener que empezar de cero y avanzando en el conocimiento.
- XML-TEI se centra en el significado y el contenido del texto y no en su apariencia o en su presentación final (es decir, no es tan relevante el señalar un título en negrita, como definir que efectivamente se trata de un “título”).

¿Qué entendemos por texto?

Antes de adentrarnos en ejemplos concretos, conviene que reflexionemos por un momento sobre qué es un texto. Una pregunta tan sencilla y obvia esconde en realidad una respuesta compleja y no siempre fácil de definir.

El texto es una idea abstracta que presupone la existencia de una secuencia de símbolos lingüísticos. ¿Pero donde se encuentra exactamente un texto? ¿Lo debemos buscar en el original que escribió el autor o en todas las otras copias y ediciones posteriores? ¿Lo debemos extraer de la historia que se explica o en las intenciones últimas del autor? ¿O es quizás en la presentación y distribución de la página?

Centrémonos por un momento en los siguientes ejemplos e intentemos reflexionar sobre dónde está el texto:



Figura 2: Primera Parte del Ingenioso Hidalgo, 1605

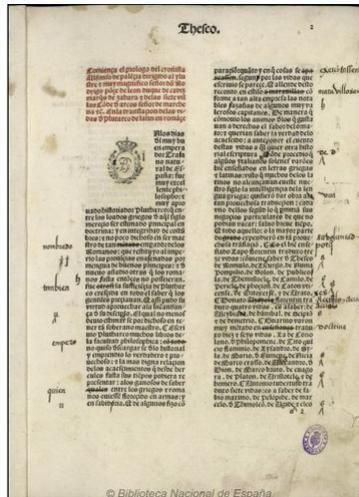


Figura 3: Alfonso de Palencia, *Parallelae*, Sevilla 1491



Figura 4: Hunain Ibn Ishaq, *Libro de los buenos proverbios*, s. IX



Figura 5: Ludovico Ariosto, *Orlando Furioso*, Venezia, 1546



Figura 6: *Dictionnaire latin-français*, Paris: Hachette, 1934

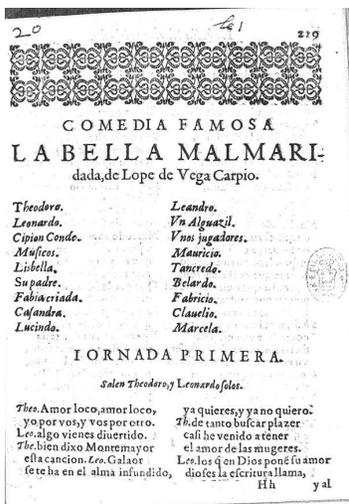


Figura 7: Lope de Vega, *La Bella Malmaridada*, Madrid, Alonso Martín, 1610

Cada una de estas tipologías contiene aspectos diferentes y todos en realidad tienen un texto. El diccionario contiene voces y significados; la obra teatral tiene escenas y personajes; una postal, además del mensaje, contiene información sobre quien y para quien se escribió; el manuscrito antiguo puede contener diversas obras, la edición del siglo XVI contiene un rico frontispicio con informaciones sobre la impresión, etc. Además hay otros aspectos que en realidad también forman parte del texto aunque sea desde un punto de vista físico: la estructura textual (parágrafos, listas, tablas. . .), en el caso del manuscrito vemos letras dañadas, caracteres especiales, líneas del folio, abreviaciones, anotaciones manuscritas, correcciones de autor, errores, entre otra casuística.

El texto, pues, no es solo el contenido textual sino que hay muchas otras variantes y componentes que debemos tener en cuenta a la hora de su codificación. Es importante que cuando afrontamos cualquier proyecto de marcado tengamos en cuenta todas y cada una de estas características que componen el “texto” en un sentido amplio. Por ello, será útil establecer una diferencia entre los conceptos de “texto” y “documento”. Mientras que el documento corresponde a un objeto en el mundo real que podemos escanear o digitalizar, como podría ser el caso de estas imágenes. El texto, en cambio, corresponde a la idea abstracta del contenido de ese documento, creado por y para una comunidad de lectores. Es esta última acepción de “texto” con la que en realidad trabajamos la mayoría de las veces al editar o codificar y requiere un análisis previo así como un cierto nivel de abstracción.

La codificación de un texto

Todas las informaciones que queramos localizar y reutilizar deben ser marcadas de una manera explícita y a través de unas marcas informáticas. Sólo las informaciones que son explícitas podrán ser buscadas a posteriori, procesadas y presentadas.

Un marcado electrónico es un valor añadido al texto pues puede proporcionar múltiples anotaciones, difícilmente señalables en una edición en papel, como por ejemplo en qué secuencia un autor canceló y añadió ciertas correcciones.

Es importante antes de empezar a codificar saber qué elementos queremos aislar y cuáles son útiles para nuestro proyecto:

- Divisiones estructurales dentro del texto: título de la página, capítulo, escena, poema, línea, párrafo. . .
- Elementos tipográficos puntuales: cambios de letra, caracteres especiales. . .
- Informaciones semánticas: personas, lugares, eventos, fechas. . .
- Otros elementos: estructuras sintácticas, formas gramaticales, localización de ilustraciones, gráficos, tablas, imágenes. . .

La decisión, una vez más, dependerá de las informaciones que queramos recuperar *a posteriori* y procesar informáticamente.

Formato vs Contenido

Conviene aclarar antes de entrar en materia que existen diferentes tipos de lenguajes de marcado. Para simplificarlo, -y para lo que aquí nos conviene- podemos establecer dos tipos básicos de lenguajes informáticos, por un lado, están los que se ocupan de la presentación (en inglés, “Presentational markup”), es decir, de cuestiones como los diferentes tipos de fuentes, el diseño y a la distribución en la página web, etc.; y por otro lado, los lenguajes descriptivos (“Descriptive markup”) que se ocupan de señalar lo que son las cosas sin preocuparse de cómo aparecerán en la página web, sea cual sea su formato de salida. De hecho, en la mayoría de los casos su transformación y presentación se relega a una etapa posterior.

Un lenguaje de presentación es claramente el lenguaje HTML que se ocupa de presentar los contenidos en la web; por el contrario, el lenguaje XML es descriptivo pues se centra en el contenido semántico. Volvamos al ejemplo anterior: imaginemos que queremos marcar el título principal de una obra; pues bien, HTML lo marcaría de esta manera: `Titulo`, donde `` está por “bold” o negrita; mientras que XML lo marcaría con otra etiqueta que indicara lo que es, no cómo debe aparecer: `<title type="principal">Titulo</title>`. El lenguaje XML-TEI, al centrarse en el contenido semántico y no en su presentación, no señala si una palabra irá o no en negrita, sino la naturaleza de ese elemento, en este caso un “título”. De esta manera, el contenido nunca está limitado a la perspectiva de un único editor, el texto existe y la presentación que se obtenga al final es relativamente irrelevante. Además, un documento XML puede ser transformado con facilidad a múltiples formatos, como .doc, epub, html, pdf, pero no siempre al revés.

En definitiva, el hecho de separar la forma o su presentación de su contenido facilita la reutilización de un documento pues no contiene todavía los corsés del formato, de manera que su difusión y reciclaje es mucho más flexible.

II. La Text Encoding Initiative

Un poco de historia...

Hacia los años ochenta, diferentes tecnologías ya habían entrado a formar parte de ciertos campos de las humanidades y las ciencias sociales; recordemos por ejemplo que la lingüística de corpus ya había visto aparecer diferentes iniciativas de proyectos lexicográficos a gran escala, como el *Index Thomisticus* del padre [Roberto Busa](#) que contó con la colaboración de IBM.

El reto que se planteó entonces fue el de desarrollar, mantener y publicar un método de codificación para los datos en humanidades en un formato electrónico que fuera independiente tanto de *hardware* y sistemas operativos como de programas informáticos. Ante el crecimiento de sistemas diferentes de representación de materiales textuales, incluidos procesadores de texto, urgía la necesidad de ofrecer a la comunidad científica un método sólido para alargar la vida de sus datos y poder reutilizarlos en el futuro con objetivos diferentes. El hecho de que cada compañía ofreciera un programa y formatos propietarios entorpecía el trabajo académico y hacía difícil la planificación de proyectos a gran escala y de la larga duración.

Así pues, en 1987 tuvo lugar la primera reunión en el Vassar College (Poughkeepsie, NY), financiada por la *Association for Computers in the Humanities* y el *National Endowment for the Humanities*, con el fin de crear un sistema para la codificación de textos electrónicos. El texto que de ahí salió se puede consultar todavía en *Design Principles for Text Encoding Guidelines*.

En los años sucesivos, un comité, con la participación de más de 50 colaboradores, empezó la redacción de lo que acabaría siendo la primera propuesta de las *Guías Directrices*, conocida como TEI P1, publicada en junio de 1990. Se inició, posteriormente, una segunda fase con la creación de los hoy habituales “working groups” o grupos de trabajo por disciplina, con

más de 100 colaboradores que, tras un trabajo de revisión y mejoras, publicaron una segunda propuesta. Pocos años después, en mayo de 1994, salieron las primeras *Guías directrices* oficiales (P3), para las cuales se había contado ya con más de 200 colaboradores.

Empezó entonces un enorme trabajo de difusión del modelo, a través de seminarios y talleres, pero sobretodo con muchos test y la retroalimentación de los colaboradores, que es lo que todavía hoy ayuda a identificar problemas, señalar necesidades concretas y proponer soluciones; para tal fin, TEI utiliza la plataforma GitHub donde se pueden solicitar nuevas funcionalidades TEI (“[Feature Requests](#)”) o llamar la atención sobre errores existentes (“[Bug](#)”).

En el año 2000 se creó, en fin, el Consorcio TEI que aspiraba a constituirse como una organización sin ánimo de lucro, permanente, académica y económicamente independiente. Una de las prioridades más urgentes que se planteó en ese momento fue la conversión de las *Guías Directrices* al lenguaje XML, pues las primeras se habían creado a partir del lenguaje SGML, un sistema más complejo, pesado y no tan ágil como XML. Así, solo dos años después, en 2002, apareció la P4, que por primera vez proponía un marco de trabajo basado ya en XML.

Finalmente, en noviembre de 2007, apareció la P5, que corresponde a la quinta propuesta todavía hoy en uso; periódicamente se realizan mejoras y se aumenta el número de su versión. El [historial de actualizaciones](#) puede encontrarse en su sitio web.

Estructura interna de la organización TEI

La Text Encoding Initiative es una organización internacional, fundada en el año 1987, encargada de desarrollar, mantener y publicar unas guías directrices para la codificación electrónica de textos en humanidades y ciencias sociales.

Desde el año 2000 su estructura organizativa se oficializó con la creación del Consorcio TEI que tiene una serie de [objetivos](#):

- Desarrollo de las *Guías Directrices* TEI.
- Difusión y promoción de las *Guías Directrices* TEI.
- Formación y divulgación
- Promoción de una comunidad de investigación TEI.

Esta organización, sin ánimo de lucro, es autofinanciada por sus miembros y subscriptores que pueden ser instituciones (bibliotecas, universidades, proyectos académicos, unidades de investigación) y personas.

Tiene dos niveles de gestión:

- *TEI Board of Directors*, encargado de la gestión del Consorcio
- *TEI Technical Council*: 12 personas, encargadas de la elaboración y el desarrollo de las *Guías Directrices*

Además, TEI cuenta con otros dos niveles de participación y organización que consisten en grupos de trabajo especializados en una materia o disciplina concretas:

- *TEI Workgroups*: equipos de investigadores y profesionales que tratan cuestiones generales relativas al marcado TEI, como la gestión de la bibliografía o de informaciones biográficas, estructuras de rasgos, descripción de manuscritos, migraciones de sistemas de marcado o de versiones TEI diferentes, etc. Su trabajo se refleja de una manera concreta en las *Guías Directrices*.
- *Special Interest Groups*: grupos de personas con intereses específicos normalmente en contacto a través de las listas de discusión, como por ejemplo, el caso de la codificación de la correspondencia epistolar, el uso de TEI en bibliotecas o la codificación de música, el uso de ontologías o la creación de herramientas, entre otras cuestiones.

Todo ello, se engloba en la gran comunidad de usuarios conectada a través de la [lista de correo electrónico](#) y de espacios participativos como la [Wiki](#).

Además de los talleres y seminarios que pueden tener lugar a lo largo del año, el Consorcio organiza des del año 2001 una conferencia anual, las últimas de las cuales han tenido lugar en [Lyon](#) (2015), [Vienna](#) (2016), y [Victoria](#) (2017), [Tokyo](#) (2018) y [Graz](#) (2019).

En este esquema se refleja el engranaje de la comunidad TEI:

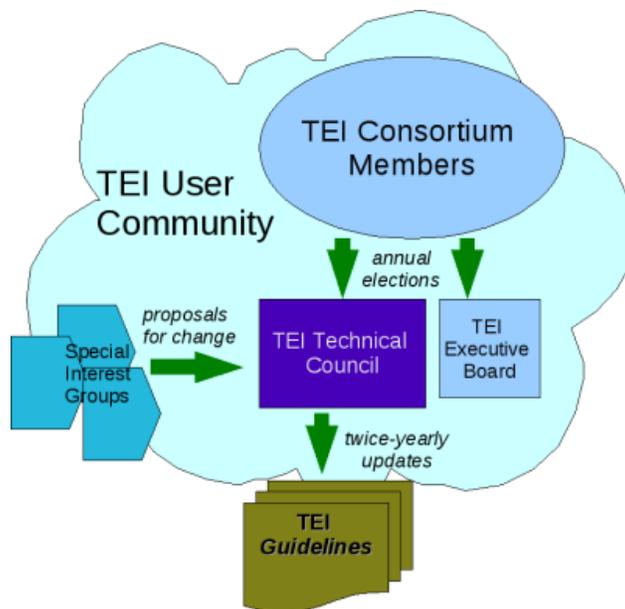


Figura 8: TEI organizational structure, 2012 (Burnard 2013)

III. Aplicaciones prácticas y tipos de proyectos

El sistema de marcado XML-TEI se utiliza en una gran variedad de contextos diferentes. En línea de máxima, se emplea para la edición de textos, pero su uso se ha extendido también a otros tipos de proyectos.

El mismo Consorcio ofrece una lista oficial de [proyectos que utilizan TEI](#) accesible en su web;

también la [universidad de Oxford](#) ha recogido otra serie de proyectos con el mismo propósito; ambas listas no son, ni mucho menos, completas.

A continuación, damos algunos ejemplos de los escenarios más típicos de su aplicación:

Edición de textos

El uso más extendido de XML-TEI es, sin duda, la edición de textos. Esta es su aplicación principal y, de hecho, muchos proyectos de mayor envergadura contienen un componente editorial. En este sentido, podemos encontrar una gran variedad de enfoques que pueden abrazar todos los diferentes métodos y escuelas de la crítica textual, desde ediciones de textos modernos, a otras diplomáticas o de crítica genética e incluso auténticas ediciones críticas en línea. Véanse como ejemplos las siguientes iniciativas:

- [La Dama Boba](#)
- [Le Petit Thalamus de Montpellier](#)
- [Les manuscrits de Madame Bovary](#)
- [Mark Twain Project](#)
- [Peter Plaoul: The Electronic Critical Edition of Plaoul's Commentary on the Sentences](#)

Algunas de las ediciones aprovechan las posibilidades del medio digital para ofrecer los textos de múltiples ediciones en paralelo e incluso los textos originales junto con sus traducciones, como por ejemplo los dos proyectos siguientes:

- [Holinsheds Chronicles](#)
- [Vincent Van Gogh, *The Letters*](#)

Otros proyectos van más allá y plantean sus ediciones como una plataforma de recursos en torno a una obra o a un autor; muchos de estos son verdaderos instrumentos pedagógicos y didácticos; véanse estos tres casos:

- [The World of Dante](#)
- [Decameron Web](#)
- [The Chymistry of Isaac Newton](#)

Dos recursos muy útiles a la hora de buscar ediciones en línea son los dos catálogos más completos hasta el momento:

- [G. Franzini, *A Catalogue of Digital Editions*](#)
- [P. Sahle, *Digital Scholarly Editions*](#)

Bibliotecas virtuales

En muchas ocasiones, la edición de los textos se inscribe en proyectos de más amplia envergadura, como el caso de las bibliotecas virtuales. Los documentos XML-TEI ofrecen la posibilidad de contener una serie muy rica de metadatos lo que facilita su inserción y conversión en bases de datos de fácil búsqueda. Algunos ejemplos, son:

- [Folger Digital Library](#)

- [Biblioteca Novohispánico](#)
- [Biblioteca Virtual Miguel de Cervantes](#)
- [Les Bibliothèques Virtuelles Humanistes](#)
- [Women Writers projects](#)

Archivos virtuales

En la misma línea que las bibliotecas virtuales, están los archivos en línea, normalmente centrados en los escritos de un autor; especialmente interesantes son aquellos casos en que su obra se encuentra en localizaciones diferentes y puede ser ofrecida en una única plataforma:

- [The Charles Brockden Brown. Electronic Archive and Scholarly Edition](#)
- [The Shelley-Godwin Archive](#)

Descripción de manuscritos

XML-TEI también es utilizado en la descripción de los fondos de las bibliotecas. Aunque estas por regla general tienen sus propios sistemas de catalogación, normalmente basado en [MARC](#), algunas emplean TEI para la descripción detallada de manuscritos y obras antiguas; esta tendencia facilita eventualmente la creación de ediciones o exposiciones virtuales de los manuscritos.

Se ha publicado, de hecho, una serie de buenas prácticas para el uso de TEI en las bibliotecas: *Best practices for TEI in Libraries*.

Algunas iniciativas son:

- [Greek Manuscripts in Sweden](#)
- [Oxford & Islamic Manuscripts Catalogue Online](#)

De la misma manera puede también utilizarse para la descripción de documentos de archivo, pero en este caso se prefiere el uso de la [Encoding Archival Description](#), también basada en el lenguaje XML y en muchos casos muy parecida a la Text Encoding Initiative.

Datos estructurados para la creación de bases de datos

XML-TEI también puede utilizarse para crear bases de datos, como listas de personajes, cronologías, o colecciones de objetos, etc. Un ejemplo representativo se puede encontrar en:

- [Personography, Map of Early Modern London](#)

Diccionarios

Cada vez más habitual es el uso de XML-TEI para diccionarios:

- [Diccionario griego español](#)
- [DuCange et al. Glossarium mediae et infimae latinitatis. Niort: L. Favre, 1883-1887](#)
- [Lexicon of Greek Personal Names](#)

Codificación de inscripciones y catálogos

XML-TEI en algunos casos se ha especializado en la codificación de ciertas tipologías documentales. Es el caso de la iniciativa [EpiDoc](#), una adaptación de las guías TEI, para la codificación de inscripciones epigráficas. Véanse estos casos:

- [IOSPE. Ancient Inscriptions of the Northern Black Sea](#)
- [US Epigraphy Project](#)
- [Inscriptions of Israel / Palestine](#)

Cartularios medievales

XML-TEI se utiliza también para la codificación de cartas y documentos medievales, tanto en latín como en otras lenguas. Existe, como en el caso anterior, alguna iniciativa para adaptar las amplias *Guías Directrices* a este tipo textual, como la [Charters Encoding Initiative](#), pero, en este caso, TEI sigue siendo la más usada.

Algunos de estos proyectos son:

- [ASChart Anglo-Saxon Charters](#)
- [CHARTA. Corpus Hispánico y Americano en la Red: Textos Antiguos](#), que en su día publicó una *Guía para editar textos CHARTA según el estándar TEI: una propuesta*.
- [Les cartulaires d'île-de-France \(Elec, ENC\)](#)

Corpus lingüísticos

TEI permite tanto una codificación generalista (por corpus, subcorpus, obras, autores, períodos, etc.), como un nivel de granularidad muy fino que comprende incluso las unidades léxicas; por ello, en algunos de los corpus la utilización de este sistema es especialmente interesante para la lematización de sus textos, es decir, para añadir la información gramatical a cada una de las palabras.

Algunos corpus que utilizan TEI son:

- [Base de Français Médiéval](#)
- [British National Corpus](#)
- [CORPES \(Corpus del Español del siglo XXI\)](#)

Editoriales

Algunas editoriales han decidido adoptar el marcado TEI-XML como método de publicación para todas o algunas de sus colecciones, e incluso también revistas científicas:

- [Journal of the TEI](#)
- [OpenEdition](#)

Otras aplicaciones

TEI-XML puede aplicarse también en función de la utilización de otras herramientas. Así, y sólo por citar dos, existen ejemplos para elaborar líneas cronológicas (*Timeline*), o para realizar transcripciones de facsímiles por zonas de píxeles, como en el proyecto del *Album interactif de paléographie médiévale*.

IV. Algunos recursos donde aprender XML-TEI

Además de la sección “Aprende” en nuestro proyecto *TTHub. Tecnologías del texto y edición digital*, donde encontrarás lecciones, recursos, ejemplos, ejercicios y un espacio colaborativo, recomiendo:

- Burghart, Marjorie & Elena Pierazzo. “Digital Scholarly Editions: Manuscripts, Texts and TEI Encoding”. #dariahTeach. DH Teaching material, 2017.
- Schreibman, Susan & Roman Bleier. “Text encoding and the TEI”. #dariahTeach. DH Teaching material, 2017.
- *Teach Yourself TEI*: Esta es la página de recursos que ofrece el mismo Consorcio TEI.
- *TEI By Example*: Aunque tiene ahora ya algunos años, sigue siendo uno de los tutoriales más completos para estudiar TEI-XML.
- *TEI@Oxford Teaching Pages*: La universidad de Oxford es conocida especialmente por sus cursos de verano sobre HD, aquí podréis encontrar una gran cantidad de presentaciones, teoría y ejercicios que os serán de mucha utilidad.
- *WWP Resources for teaching and learning text encoding*: El proyecto Women Writers Projects, ahora en la universidad de Northeastern, ofrece una serie de materiales muy interesantes para aprender TEI y otros lenguajes relacionados.

V. Bibliografía

Burnard, L., “Introduction”, en Burnard, L. (2014), *What is the Text Encoding Initiative. How to add intelligent markup to digital resources*, Marserville: OpenEdition Press. <https://books.openedition.org/oep/679>

Burnard, L. (2013), “The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure”, *Journal of the Text Encoding Initiative* 5, Junio. <https://journals.openedition.org/jtei/811>

McGann, J. (2004), “Marking Texts of Many Dimensions”, en S. Schreibman, R. Siemens, J. Unsworth, *A companion to Digital Humanities*, Oxford: Blackwell, pp. 218-239.

Pierazzo, Elena (2016). “Textual Scholarship and Text Encoding”. en S. Schreibman, R. Siemens, J. Unsworth, *A New Companion to Digital Humanities*, Oxford: Wiley Blackwell.

Renear, A. H. (2004), “Text Encoding”, en S. Schreibman, R. Siemens, J. Unsworth, *A companion to Digital Humanities*, Oxford: Blackwell, pp. 218-239.

TEI Consortium (2014), “TEI: History”. <http://www.tei-c.org/about/history/>.

Cita propuesta:

Allés Torrent, Susanna (2019). “Introducción a la Text Encoding Initiative. Definición, aplicaciones prácticas y recursos”. *TTHUB. Text Technologies Hub: Recursos sobre tecnologías del texto y edición digital*. <https://tthub.io/aprende/11-intro-a-tei/> DOI: 10.5281/zenodo.3530772