



Introducción a TEI (Text Encoding Initiative)

Susanna Allés Torrent

Columbia University

susannalles@gmail.com



Madrid, 14 de julio de 2015



Programa:

1. ¿Qué es la Text Encoding Initiative?
2. Breve historia y organización de la TEI
2. TEI y edición digital
3. El texto se transforma en datos: ¿qué marcar en una edición digital?
4. Recapitulación: XML-TEI
5. *Guidelines* o las Guías directrices
6. Estructura de un documento TEI
7. El sistema de módulos
8. Personalización TEI (ROMA)
9. Herramientas y tutoriales

Materiales para la sesión:

Presentacion.pdf

Ejercicios:

- 1.Instrucciones-instalacion-oxygen.pdf
- 2.Creacion_doc_TEI.pdf
3. Crear_esquema_Roma.pdf

Ejemplos:

- Ejemplo-TEI-basico.xml
- tei_lite.rng
- poema.xml
- ejercicio2-Intro-XML.xml

Soluciones*

Objetivos:

- Entender qué es TEI y para qué sirve
- Conocer las partes de un documento TEI
- ¿Cómo afrontar proyectos de edición digital?
- Entender el sistema de módulos
- Saber crear un esquema personalizado
- Conocer algunas herramientas y recursos para aprender más



1. ¿Qué es la Text Encoding Initiative?

Definición (1)

Por Text Encoding Initiative entendemos diversas cosas:

1. 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.
2. Una iniciativa científica destinada a la codificación informática de textos de cualquier época y género de manera que puedan ser procesados digitalmente.
3. Una metodología concreta, materializada en las TEI Guidelines, basada en el lenguaje XML para la codificación de textos.

Definición (oficial)



Las [TEI Guidelines for Electronic Encoding and Interchange](#) [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**, son acompañadas de una **documentación detallada**, y publicadas bajo una licencia [open-source](#). Las *Guidelines* son mantenidas y desarrolladas por el **TEI Consortium**, a través de su **Technical Council**, con el respaldo y la participación de la **comunidad TEI**.

<http://www.tei-c.org/Guidelines/>



Puntos clave:

1. Las *Guidelines* son el núcleo central y constituye el “manual de uso” o de “buenas prácticas”.
2. Proponen un lenguaje de marcado basado en XML que consiste en aislar a través de “marcas” o “etiquetas” las características textuales: estructurales (capítulos de libro, secciones, apartados, etc.), físicas (distribución de páginas en un manuscrito), y conceptuales o semánticas (nombres de personas, de lugares, palabras clave, etc.).
3. Se usa desde hace más de 20 años en disciplinas humanísticas, Ciencias Sociales, Lingüística...
4. Funcionamiento por módulos obligatorios y optativos.
5. Se expresa a través del lenguaje XML, utiliza esquemas (generalmente también en XML) algunos ya preexistentes o bien personalizables.
6. Modelos bien documentados; independiente de cualquier software, acceso abierto, por tanto consultable libremente online, descargable, reutilizarse e, incluso, susceptible de mejoras a través del feedback de los usuarios. Todo el material está disponible en [Sourceforge](#).



2. Breve historia y organización de la Text Encoding Initiative

1. Breve historia

1987: primera reunión ([Poughkeepsie Conference](#), Vassar College) con el fin de crear unas guías directrices para la codificación de textos electrónicos

1988 – 1990: producción de TEI P1 con la participación de más de 50 colaboradores

1990 – 1992: redacción de TEI P2 y creación de grupos de trabajo (Working Groups) por disciplina (100 colaboradores)

1993 – 1994: TEI P3, primera versión oficial de las Guidelines (más de 200 colaboradores)

1995 – 1999: Gran trabajo de difusión del modelo (seminario, talleres, etc.), muchos test y, sobretodo, feedback de los colaboradores.

2000: [creación del Consortium TEI](#)

2002: TEI P4, primera versión en XML (de la P3)

2007: TEI P5

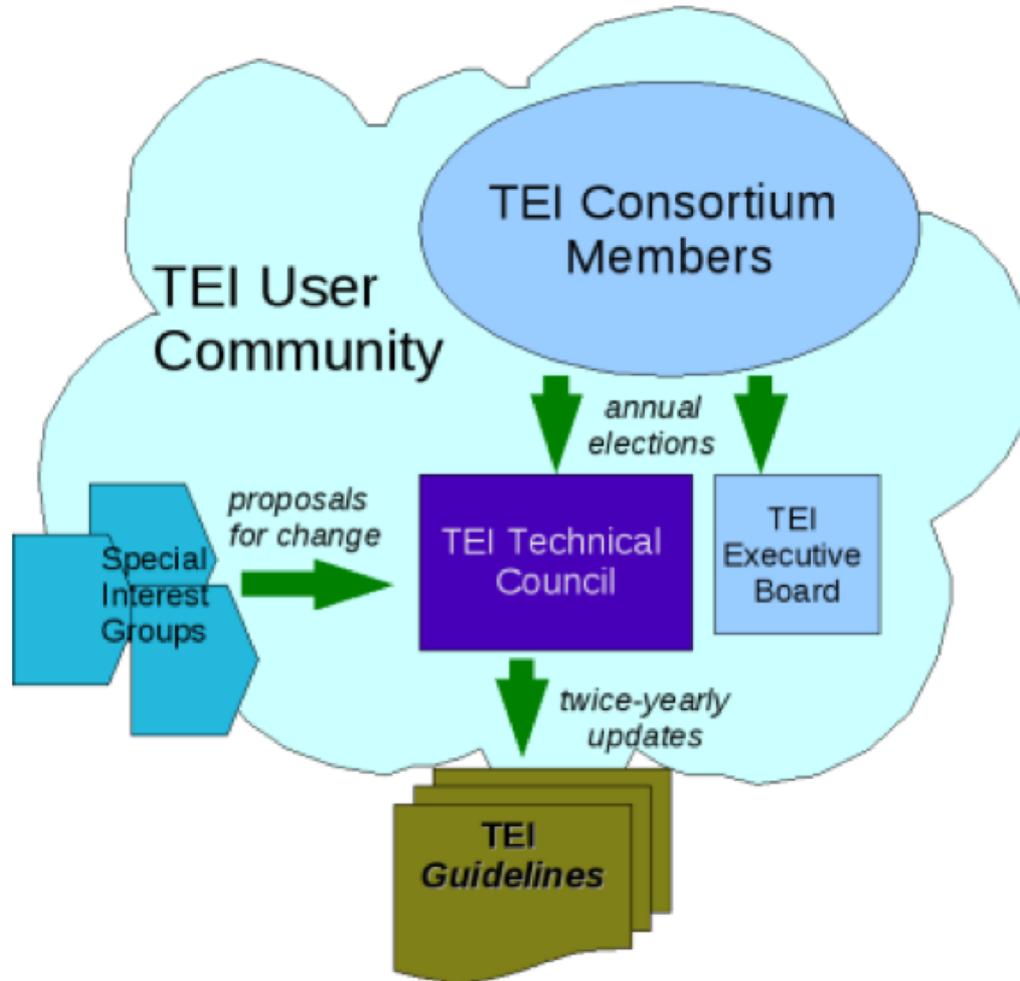
2011: TEI P5 Versión 2.0.1., todavía hoy en uso

2014: Última actualización: Versión P5 2.6.0

- Burnard, L. (2013), “[The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure](#)”, Journal of the Text Encoding Initiative 5, Junio.

- Más informaciones en [TEI: History](#)

1. Organización



1. Organización



- Estructura oficial desde el año 2000 con la creación del Consorcio TEI cuyos objetivos ([Goals and Mission](#))
 - Desarrollo de las Guías directrices TEI.
 - Difusión y promoción de las Guías directrices TEI.
 - Formación y divulgación
 - Promoción de una comunidad de investigación TEI.
- Organización, sin ánimo de lucro, autofinanciada por sus miembros y subscriptores (instituciones, tales como bibliotecas, universidades, proyectos académicos, unidades de investigación) y personas.
- [Dos niveles de gestión](#):
 - [TEI Board of Directors](#), encargado de la gestión del Consorcio
 - [TEI Technical Council](#): 12 personas, encargadas de la elaboración y la evolución de las Guías directrices
- Dos niveles de trabajo:
 - [TEI Workgroups](#):
 - [Special Interest Groups](#)
- Gran comunidad de usuarios conectada a través de [listas de correo electrónico](#) y de espacios participativos como la [Wiki](#).
- Conferencia anual: [Northwestern University](#) 2014 (Evanston, Illinois) y este año 2015 en Lyon.



3. TEI y la edición académica digital

TEI: *lingua franca*

- TEI es uno de los lenguajes más utilizados en el panorama actual de la edición digital.
- **Pros:**
 - Standard web y en formato abierto
 - Comunidad activa
 - Bien documentado a nivel de proyectos de humanidades
 - Permite el Intercambio de textos e interoperabilidad con otros proyectos de edición
- **Contras:**
 - Se necesita personal preparado para concebir el proceso de edición digital
 - Se necesitan investigadores que sepan codificar
 - Falta de fondos para preparar el personal...

No solo de edición digital vive TEI...

1. Edición de textos
2. Bibliotecas virtuales
 - [Biblioteca Pensamiento Novohispano](#)
 - [Bibliothèques Virtuelles des Humanistes](#)
3. Archivos virtuales: [The Shelley-Godwin Archive](#)
4. Descripción de manuscritos: [Greek Manuscripts in Sweden](#)
5. Datos estructurados para la creación de bases de datos: [Personography](#) (Map of Early Modern London)
6. Diccionarios: [Du Cange](#)
7. Epigrafía digital y catálogos: [US Epigraphy](#)
8. Cartas medievales / Diplomática: [CHARTA](#), [ASChart](#)
9. Corpus lingüísticos: [CORPES](#), [British National Corpus](#)
10. Editoriales: [OpenEdition](#), [Journal of the TEI](#)
11. Otras aplicaciones: [Timelines...](#)

Catálogos y listas de ediciones digitales

- Lista de proyectos en la [página TEI](#)
- [University of Oxford Text Archive](#)
- [A Catalogue of Digital Editions](#) (G. Franzini)
- [A catalog of Digital Scholarly Editions](#) (P. Sahle)

Ejemplos

En francés...

- [Montedite, édition en ligne de manuscrits de Montesquieu](#)
- [Artamène ou le Grand Cyrus](#) (1649-1653) de Madeleine de Scudéry
- [Bibliothèques virtuelles humanistes](#) (CESR)
- [Les dossiers de Bouvard et Pécuchet de Flaubert](#)
- [Les manuscrits de Madame Bovary](#) (U. de Rouen et al.)
- [Éditions de l'École nationale des chartes](#)

En inglés:

- [Folger Digital Library](#) (Folger Shakespeare Library)
- [Jane Austen Fiction Manuscripts](#) (U. Oxford – King's College)
- [The Holinshed Project](#) (U. Oxford)
- [The World of Dante](#) (U. of Virginia)
- [The Online Froissart](#) (U. of Sheffield)
- [Vincent van Gogh. The Letters](#) (Huygens ing – Van Gogh Museum)
- [William Godwin's Diary](#) (U. Oxford)
- [Petrus Plaoul. Editiones electronicas](#) (J. C. Witt)

La edición digital en España



- Proyectos en TEI:
 - [Biblioteca Virtual Miguel de Cervantes](#) (Universidad de Alicante)
 - [Corpus de la Real Academia Española](#)
 - [ReMetCa](#)
 - [Dicemto](#). Diccionario electrónico multilingüe de verbos de movimiento (Universitat Autònoma de Barcelona & al.)
 - [Archivo Iberoamericano de Cetrería](#) (U. Valladolid)
- Además:
 - Iniciativas de formación: Título de experto profesional en Humanidades Digitales (30ECTS), Laboratorio de Innovación en HD, UNED
- Algunas lecturas:
 - P. Spence – E. González-Blanco, [A historical perspective on the digital humanities in Spain](#), H-Soz-Kult (ed.), *The Status Quo of Digital Humanities in Europe*, 2014.
 - S Allés Torrent, «Edición digital y algunas tecnologías aliadas: codificación TEI y transformaciones XSLT», *Insula. Revista de Letras y Ciencias Humanas*, 822, Junio 2015.



**4. El texto se transforma en datos:
¿Qué marcar en una edición digital?**

Preguntas frecuentes

- ¿Qué información debemos codificar?

La decisión depende de las informaciones que queramos recuperar a posteriori y que queramos procesar informáticamente.

- ¿Para qué sirve un lenguaje de marcado?

Para hacer explícitos ciertos rasgos e informaciones textuales de manera que ayude el procesamiento del texto a través de programas informáticos.

- ¿Por qué marcamos los textos?

Los formatos .txt, .doc, .pdf, no son suficientes y no pueden ser procesados tan fácilmente. Ejemplos: si queremos publicar un corpus de cartas y al final decidimos que los destinatarios deben ir alineados a la izquierda, si queremos hacer listas de los emisores y los destinatarios, etc.

El objetivo siempre es:

Explicitar las partes del texto y toda la información implícita a través de marcas informáticas, para poderlas buscar a posteriori, procesarlas informáticamente por diferentes aplicaciones y presentarlas en diferentes formatos.

Modelización de datos

Antes de empezar a codificar hay que identificar las características del texto:

- **divisiones estructurales:** título de la página, capítulo, escena, poema, línea, párrafo, listas, tablas, comentarios, glosas,
- **cuestiones editoriales:** errores, cancelaciones, lecturas diferentes, adiciones, ...
- **elementos tipográficos puntuales:** cambios de letra, caracteres especiales...
- **rasgos físicos:** lagunas, saltos de línea, saltos de página, abreviaturas,...
- **informaciones semánticas:** personas, lugares, eventos, fechas, verso, rima...
- **otros elementos:** estructuras sintácticas, formas gramaticales, localización de ilustraciones, gráficos, tablas, imágenes...

Formato vs Contenido

Marcado de presentación

- HTML: presenta los contenidos en la web
- Indica el estilo y su presentación

`Título`

`<negrita>Título</negrita>`

Marcado descriptivo

- XML: describe lo que son las cosas sin preocuparse de su presentación (contenido semántico)

`<title>Título</title>`

`<titulo>Título</titulo>`

Unidad 1 Text Encoding Initiative y aplicaciones prácticas tiene

En esta primera unidad ofreceremos una introducción general a la Text Encoding Initiative y sus aplicaciones prácticas.

Apartado **PARTE I. Cuestiones generales** Título

1. Introducción Título

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 iniciativa de marcado de textos nace pues de la necesidad de procesar los textos desde un punto de vista informático.

párrafo

2. ¿Qué es la Text Encoding Initiative? Título

Subapartado 2.1. Definición y objetivos Título

La Text Encoding Initiative se ampara en primera instancia en el Consorcio TEI, cuya estructura veremos en la Parte II de esta unidad

TEI posee una plataforma web donde se publican tanto las Guías directrices como un sinfín de herramientas y de comunidades de ayuda [<http://www.tei-c.org>] (Ver vídeo Presentación página web del Consorcio TEI). Aquí encontraréis una primera definición que nos será útil para empezar:

The TEI Guidelines for Electronic Text Encoding and Interchange define and document a markup language for representing the structural, rendition, 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.

cita

TEI Consortium, Guidelines

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

1) El núcleo central de la TEI son las *Guidelines* que constituyen el manual de uso y la recomendación de buenas prácticas sobre cómo proceder.

2) La *Guidelines* proponen un tipo concreto de lenguaje de marcado que consiste en aislar a través de "marcas" o "etiquetas" las características textuales,

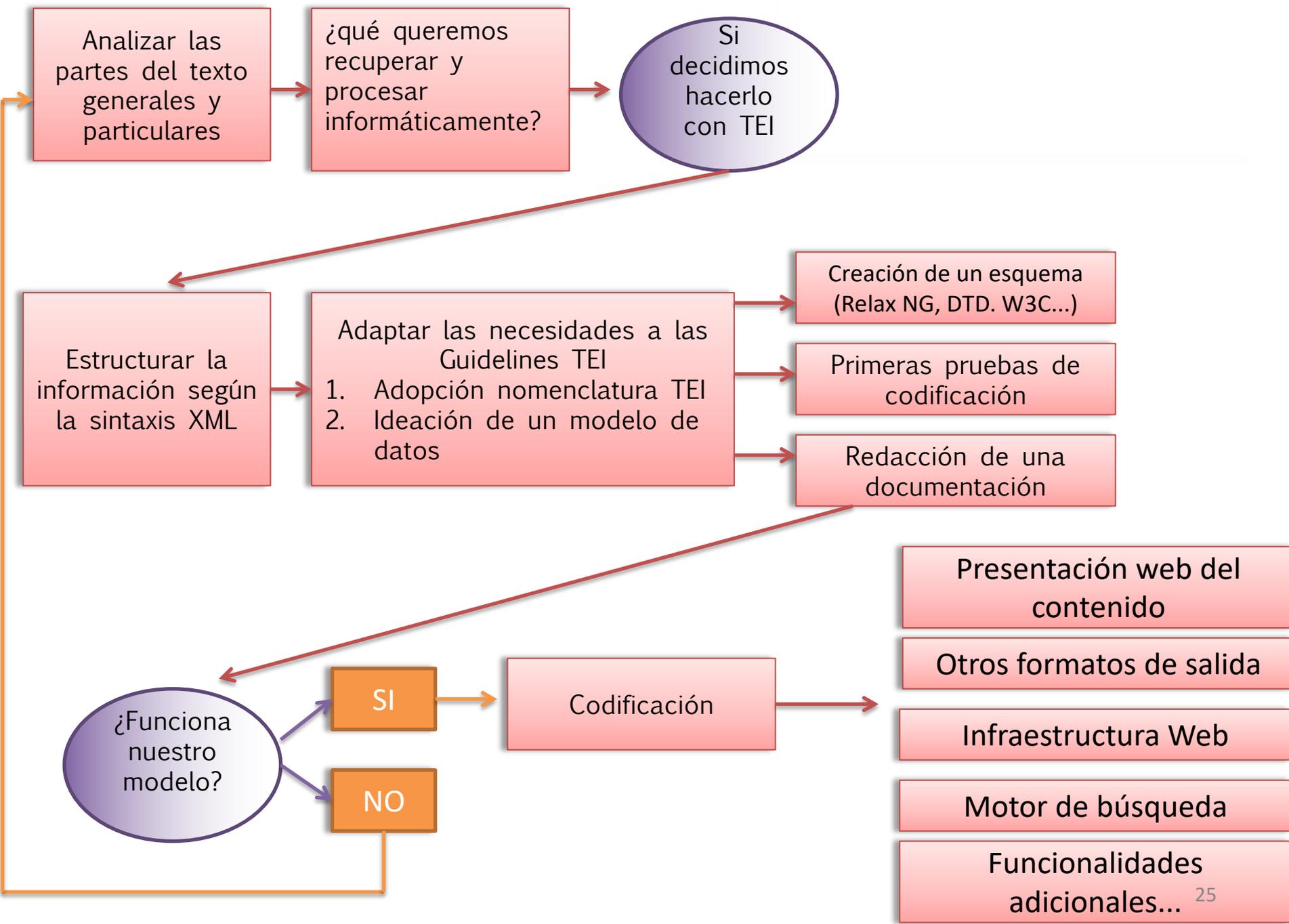
lista

3) TEI es especialmente usado en las disciplinas humanísticas, ciencias sociales o lingüística.

4) El sistema de marcado propuesto por TEI tiene una naturaleza modular que agrupa fenomenología diversa en un mismo "módulo".

- Las divisiones que corresponden a la primera parte, a los apartados (1, 2...) y a los subapartados (1.1., 1.2., etc.)
- Los títulos o encabezados
- Los párrafos
- Las citas bibliográficas
- Las listas
- Los números de página
- Las referencias cruzadas

¿Cómo expresarlo según las TEI Guidelines?





5. Recapitulación: XML -> TEI

Conceptos básicos XML-TEI (1)

TEI se expresa en XML, por tanto sigue las mismas reglas, como por ejemplo:

- el nombre del elemento puede solo contener letras y números, sin acentos ni espacios, y solo algunos signos: - _ .
- el nombre no puede empezar por un número, ni por un signo, ni por `xml`.
- ¡XML es sensible a las mayúsculas y las minúsculas!
- Los documentos XML-TEI deben:
 - Estar bien formados, es decir estar expresado correctamente según la gramática de XML.
 - Ser válido: Para ello, necesita un modelo que fije la estructura básica y las posibilidades de anidación del documento (DTD o esquemas (RelaxNG, W3CSchema) que establecen la sintaxis del documento XML).

TEI ofrece una introducción a XML: [A Gentle Introduction to XML](#)

Conceptos básicos XML-TEI (2)

- Noción de **Marcado / Codificación** = “Markup” (en.), “Balisage” (fr.), “Codifica” (it.) → aislar una porción de texto con caracteres que son interpretados por la máquina como código y que poseen un significado semántico específico.
- **Etiqueta** = ‘Tag’: unidad básica

```
<etiqueta>contenido</etiqueta>
```

- Estructura arbórea: anidación sin límites de unos elementos al interior de otros, construyendo así una especie de árbol genealógico. Por eso, necesitamos siempre un solo elemento raíz del que cuelguen todos los otros.

```
<curso>  
  <titulo>Introducción a XML-TEI</titulo>  
  <autor>Susanna Allés</autor>  
  <fecha>14 de julio de 2015</fecha>  
</curso>
```

Conceptos básicos XML-TEI (3)

Elemento: son las unidades de información semántica compuestas de una marca de apertura `<elemento>` y de cierre `</elemento>`. El nombre del elemento constituye su identificador general. Puede contener texto, otros elementos o estar vacío (`<elemento/>`).

Atributo: añaden una característica adicional al elemento; su orden no es relevante; deben llevar siempre un valor.

Valor: rellena el valor o contenido del atributo.

```
<elemento atributo="valor">contenido</elemento>
```

Ejemplos XML:

```
<fecha cuando="0972">972</fecha>  
<division tipo="prologo">El prólogo... </division>  
<frase tipo="enunciativa">Esto es una frase</frase>  
<frase tipo="interrogativa">¿Cómo te llamas?</frase>
```



6. Las *Guidelines* o Guías directrices

TEI Guidelines (1)

- Definen un lenguaje de marcado en XML para la representación de las características estructurales, físicas y conceptuales de los textos.
- Objetivo: procesamiento, explotación y análisis.
- Basado en esquemas XML extensibles
- Una selección de más de [500 etiquetas](#)
- Una estructura en [clases de modelo](#) y de [atributos](#);
- Marco de trabajo flexible basado en módulos (un total de 21; 4 obligatorios)
- Documentación detallada
- Publicación bajo open-source
- Son mantenidas y actualizadas a través de TEI Consortium, Technical Council y la participación de la comunidad TEI.

Véase “[1 The TEI Infrastructure](#)”, *TEI Guidelines P5*

<arc>

Principal
C Elementos

<arc> codifica un arco, es decir, la conexión entre un nodo y otro al interno de un gráfico [\[19.1 Graphs and Digraphs\]](#)

Módulo nets — Graphs, Networks, and Trees

Atributos [att.global](#) (@xml:id, @n, @xml:lang, @xml:base, @xml:space) ([att.global.rendition](#) (@rend, @style, @rendition)) ([att.global.linking](#) (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) ([att.global.analytic](#) (@ana)) ([att.global.facs](#) (@facs)) ([att.global.change](#) (@change)) ([att.global.responsibility](#) (@cert, @resp))

@from proporciona el identificador al nodo adyacente al nodo corriente.

Estado Necesario

Tipo de datos

[data.pointer](#)

De formato compacto a XML

@to proporciona el identificador al nodo adyacente al nodo corriente.

Estado Necesario

Tipo de datos

[data.pointer](#)

De formato compacto a XML

Contenido en nets: [graph](#)

Puede contener core: [label](#)

Declaración

```

element arc
{
  att.global.attributes,
  att.global.rendition.attributes,
  att.global.linking.attributes,
  att.global.analytic.attributes,
  att.global.facs.attributes,
  att.global.change.attributes,
  att.global.responsibility.attributes,
  attribute from { data.pointer },
  attribute to { data.pointer },
  ( label, label? )?
}
    
```

Ejemplo

```

<arc from="#T3" to="#T3">
  <label>OLD</label>
  <label>VIEUX</label>
</arc>
    
```

[Mostrar todo](#)

- a) una definición,
- b) el módulo de pertenencia
- c) clase de atributos que puede conllevar,
- d) clase de modelo del que forma parte
- e) indicación de los elementos donde puede ser utilizado,
- f) serie de elementos que puede contener en su interior;
- g) código del esquema (RelaxNG) que define su comportamiento;
- h) Ejemplos

Estructura y formatos de publicación

- Las GD se fundamentan sobre el principio del acceso abierto, por ello todo lo que las rodea se publica bajo esa licencia y puede descargarse por completo:
 - Esquemas
 - Código fuente
 - Documentación
- Las GD pueden descargarse en múltiples formatos (lo más habitual –debido al volumen de las mismas– es hacerlo en su versión HTML en inglés o parcialmente en español)

[Versión HTML en línea](#)

[En formato PDF](#)

[En formato EPUB](#)

[En formato MOB](#)



7. Estructura de un documento TEI

Hands-on...

- Instalación de <oxygen/> u otro editor XML: 1.Instrucciones-instalacion-oxygen.pdf
- Abrid el programa y exploremos...

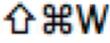
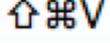
Hands-on...

- Abrid el documento `documento_TEI_basico.xml` con el programa `<oXygen/>` u otro editor de texto XML
- Podéis ir comentando vuestro fichero XML-TEI.

Pequeños trucos en <oXygen/>

- Control + E = pop up con la lista de los elementos
- Control + Mayus + , = convierte selección en comentario

Recordad:

- Abrid la ventana de los elementos:
Window > Show View > Elements
- Indentar el documento 
- Comprobar que el documento:
 - esté bien formado  Check Well-Formedness  ⌘W
 - sea válido  Validate  ⌘V

Estructura mínima

1. Prólogo XML:

- La declaración XML nos indica que se trata de un documento XML y que el tipo de codificación utilizado es el UTF-8.
- La asociación del modelo de esquema: en este caso se trata de un esquema RelaxNG (.rng).

2. El elemento raíz <TEI> que engloba la totalidad del documento. Aquí irán todos los otros elementos XML-TEI: el encabezado y el texto, con la siguiente estructura:

3. Encabezado o <teiHeader>: se sitúan los metadatos, es decir, las diferentes informaciones relativos al texto que estamos marcando.

4. Cuerpo del texto o <text>

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model
href="http://www.tei-
c.org/release/xml/tei/custom/schema/relaxng/teilight.rng"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Título</title>
      </titleStmt>
      <publicationStmt>
        <p>Información sobre la publicación electrónica</p>
      </publicationStmt>
      <sourceDesc>
        <p>Información sobre la fuente</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Algo de texto.</p>
      <figure><graphic url="http://www.tei-c.org/logos/TEI-
glow.png"/></figure>
    </body>
  </text>
</TEI>
```

Véase: Ejemplo-TEI-basico.xml

```
<?xml version="1.0" encoding="UTF-8"?>
```

Declaración XML

```
<?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/teilight.rng" schematypens="http://relaxng.org/ns/structure/1.0"?>
```

Asociación al esquema .rng

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
```

elemento raíz, espacio de nombre

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Título del fichero XML</title>
    </titleStmt>
    <publicationStmt>
      <p>Información sobre la publicación digital</p>
    </publicationStmt>
    <sourceDesc>
      <p>Información sobre el texto original</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Encabezado

```
<text>
  <body>
    <p>El texto </p>
    <figure><graphic url="http://www.tei-c.org/logos/TEI-glow.png"/></figure>
  </body>
</text>
```

El texto

```
</TEI>
```

Encabezado o <teiHeader>

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Título del
fichero XML</title>
    </titleStmt>
    <publicationStmt>
      <p>Información sobre la
publicación digital</p>
    </publicationStmt>
    <sourceDesc>
      <p>Información sobre el
texto original</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

- El <teiHeader> tiene un solo elemento obligatorio, llamado <fileDesc>. Este elemento es el responsable de contener tanto la información sobre el fichero XML-TEI con el que estamos trabajando, como los detalles de la fuente primaria:
- <titleStmt> contiene la información sobre el título del fichero propiamente dicho; éste a su vez debe ir dentro de un elemento más específico llamado <title>.
- <publicationStmt> recoge la información sobre la publicación digital; la información debe ir organizada en párrafos: <p>
- <sourceDesc> contiene la información de la fuente original.

Cuerpo del documento o `<text>`

```
<text>
  <front>
    <div type="prologo">
  </div>
  </front>
  <body>
    <div type="tomo" n="1">
      <div type="capitulo" n="1">...</div>
      <div type="capitulo" n="2">...</div>
      <div type="capitulo" n="3">...</div>
    </div>
    <div type="tomo" n="2">
      <div type="capitulo" n="1">...</div>
      <div type="capitulo" n="2">...</div>
      <div type="capitulo" n="3">...</div>
    </div>
  </body>
</text>
```

- `<front>`: se utiliza para marcar elementos paratextuales que preceden el texto, tales como prefacios, prólogos, cartas dedicatorias, una lista de personajes, etc. No es obligatorio.
- `<body>`: se consagra al texto propiamente dicho del documento y es obligatorio.
- `<back>`: puede utilizarse para añadir los apéndices, índices, cronologías, bibliografías, etc. No es obligatorio.

Divisiones o <div>

Uno de los elementos más utilizados en el cuerpo del documento es el de “división”, `<div>`, que es utilizado para marcar cualquier tipo de sección estructural del texto, como tomos o volúmenes, libros, capítulos, partes, apartados, etc.

Del ejemplo anterior, podéis observar cómo hemos utilizado el elemento `<div>` para:

- caracterizar el “prólogo”.
- diferenciar los diferentes “tomos”.
- estructurar los diferentes “capítulos”.

Dos atributos muy utilizados:

- `@type`
- `@xml:id`

```
<div type="tomo" xml:id="tomo_1"> ...</div>
```



8. El sistema de módulos

Sistema por módulos

En primer lugar, conviene recordar que el sistema de codificación TEI se divide en módulos, cada uno destinado a una tipología textual: verso, drama, transcripciones de discurso, diccionarios, etc. Cada uno de estos módulos es descrito en las [Guidelines](#).

Cada módulo define una serie de etiquetas propias (de un total de aproximadamente [500 elementos](#)), así como las [clases de modelo](#) y de [los atributos](#). Cada módulo tiene una presentación general, donde se explica el funcionamiento de cada uno de sus elementos.

Los módulos pueden ser obligatorios o optativos. Así, los módulos obligatorios son cuatro: [tei](#), [core](#), [header](#), [textstructure](#); sin estos no tendríamos un documento válido.

Lista de módulos (1)

1. Infraestructura TEI
2. TEI Header (encabezado) = header
3. Elementos disponibles en todos los docs TEI = core
4. Estructura del texto por defecto = textstructure
5. Caracteres no estándares y Glyph = gaiji
6. Verso = verse
7. Drama = drama
8. Transcripción de discurso = spoken
9. Diccionarios = dictionaries
10. Descripción de manuscritos = msdescription
11. Representación de fuentes primarias = transcr
12. Aparato crítico = textcrit
13. Nombres, fechas, personas y lugares = namesdates
14. Tablas, fórmulas, gráficos i música anotada = figures

Lista de módulos (1)

15. Corpora textuales = `corpus`
 16. Linking, segmentación y alineación = `linking`
 17. Mecanismos analíticos simples = `analysis`
 18. Estructuras de rasgos (*Feature Structures*) = `iso-fs`
 19. Gráficos, Networks y Árboles = `nets`
 20. Estructuras no jerárquicas = `header`
 21. Certitud, Precisión y Responsabilidad = `certainty`
 22. Elementos para la documentación = `tagdocs`
- (Describe el esquema XML usado para expresar las Guidelines)

The following table lists the modules defined by the current release of the Guidelines:

Module name	Formal public identifier	Where defined
analysis	Analysis and Interpretation	17 Simple Analytic Mechanisms
certainty	Certainty and Uncertainty	21 Certainty, Precision, and Responsibility
core	Common Core	3 Elements Available in All TEI Documents
corpus	Metadata for Language Corpora	15 Language Corpora
dictionaries	Print Dictionaries	9 Dictionaries
drama	Performance Texts	7 Performance Texts
figures	Tables, Formulae, Figures	14 Tables, Formulæ, Graphics and Notated Music
gaiji	Character and Glyph Documentation	5 Characters, Glyphs, and Writing Modes
header	Common Metadata	2 The TEI Header
iso-fs	Feature Structures	18 Feature Structures
linking	Linking, Segmentation, and Alignment	16 Linking, Segmentation, and Alignment
msdescription	Manuscript Description	10 Manuscript Description
namesdates	Names, Dates, People, and Places	13 Names, Dates, People, and Places
nets	Graphs, Networks, and Trees	19 Graphs, Networks, and Trees
spoken	Transcribed Speech	8 Transcriptions of Speech
tagdocs	Documentation Elements	22 Documentation Elements
tei	TEI Infrastructure	1 The TEI Infrastructure
textcrit	Text Criticism	12 Critical Apparatus
textstructure	Default Text Structure	4 Default Text Structure
transcr	Transcription of Primary Sources	11 Representation of Primary Sources
verse	Verse	6 Verse

Módulos principales



tei: definición de las clases, macros y tipos de datos

textstructure: elementos de base para estructurar un texto de tipo libro

core: elementos disponibles en todos los documentos

TEI

header: encabezado TEI (metadatos del documento)

Elementos generales: core

1. Párrafos y cuestiones de puntuación: [<p>](#), [<seg>](#), [<s>](#), [<pc>](#)
2. Elementos tipográficos: [<hi>](#), [<foreign>](#), [<emph>](#), [<distinct>](#), [<mentioned>](#), [<soCalled>](#)
3. Citas: [<q>](#), [<said>](#), [<quote>](#), [<cit>](#), [<bibl>](#), [<biblScope>](#), [<title>](#), [<author>](#)
4. Nombres, números, fechas: [<name>](#), [<rs>](#), [<date>](#), [<num>](#), [<measure>](#)
5. Listas, notas, anotaciones, indexación, referencias: [<list>](#), [<item>](#), [<label>](#), [<note>](#), [<index>](#), [<term>](#), [<ptr>](#), [<ref>](#)
6. Otros elementos frecuentemente utilizados:
 1. Cambios editoriales: [<choice>](#), [<sic>](#), [<corr>](#), [<orig>](#), [<reg>](#), [<add>](#), [](#), [<gap>](#), [<unclear>](#)
 2. Milestones: [<milestone>](#), [<gb>](#), [<pb>](#), [<lb>](#), [<cb>](#)

Atributos

En la mayoría de los casos los elementos poseen atributos, que pueden ser generales o específicos de un elemento:

Atributos generales muy utilizados:

@xml:id: consigna un identificador a un elemento,

```
<note xml:id="nota1">...</note>
```

@target: indica la localización hacia un identificador,

```
<ref target= "#nota1">...</ref>
```

@type: caracteriza un elemento,

```
<div type="prologo"> ...</div>
```

@n: añade un número,

```
<div n="1">...</div>
```

@xml:lang: especifica la lengua del documento o de la sección,

```
<gloss xml:lang="es">La glosa</gloss>
```

@rend: especifica una presentación diferente ("rendition"),

```
<title rend="secundario">Título</title>
```

Módulos: verso

<lg>, <l>, <seg>, <caesura>, <rhyme>

```
<lg type="soneto">  
  <lg type="cuarteto" n="01" rhyme="abba">  
    <l>Escrito está en mi alma vuestro gesto, </l>  
    <l>y cuanto yo escribir de vos deseo; </l>  
    <l>vos sola lo escribisteis, yo lo leo </l>  
    <l>tan solo, que aun de vos me guardo en esto.</l>  
</lg>
```

Módulos: drama

<actor>, <camera>, <castGroup>, <epilogue>, <role>,
<set>,<sp>, <speech>, <speaker>, <sound>...

```
<div type="acto" n="01">  
  <div type="escena" n="01">  
    <stage>Sale Felisardo con espada y broquel, y  
Mendoça, su lacayo, detrás.</stage>  
    <sp xml:id="sp-1" who="#Men">  
      <speaker>Men.</speaker>  
      <ab>Casta muger.</ab>  
    </sp>  
    <sp xml:id="sp-2" who="#Fel">  
      <speaker>Fel.</speaker>  
      <ab>Para casta lo deve Octavia de ser.</ab>  
    </sp>  
    <!-- ... -->  
  </div>  
</div>
```

Módulos: diccionarios

<case>, <entry>, <entryFree>, <lemma>, <def>, <gramGrp>, <form>, <orth>, <etym>, <lang>, <pos>, <xr>...

```
<entry>
  <form><orth>ventana</orth></form>
  <sense n="1">
    <gramGrp><gen>f.</gen></gramGrp>
    <def>Abertura más o menos elevada sobre el suelo, que se deja en una pared para dar luz y ventilación.</def>
  </sense>
  <sense n="2">
    <gramGrp><gen>f.</gen></gramGrp>
    <def>Hola o hojas de madera y de cristales con que se cierra esa abertura.</def>
  </sense>
</entry>
```

Módulos: Representación de fuentes primarias

<damage>, <facsimile>, <fw>, <handNotes>, <handShift>, <line>, <space>, <supplied>, <zone>...

```
<text>  
  <pb facs="page1.png" />  
<!--el texto contenido en la página 1 se marcará aquí -->  
  <pb facs="page2.png" />  
<!-- Aquí el contenido de la página 2-->  
</text>
```

Módulos: Aparato crítico

<app>, <lem>, <listWit>, <wit>...

```
<1 n="06">
  <app>
    <rdg wit="#M1">El contenido de la línea 06 que
      aparece en el manuscrito M1
      <note>Nota relativa al manuscrito.</note>
    </rdg>
    <rdg wit="#M2">El contenido de la línea 06 que
      aparece en el manuscrito 2
    <rdg wit="#M3"/>
  </app>
</1>
```

Módulos: Nombres, fechas, personas, lugares

<affiliation>, <age>, <birth>, <climate>, <country>, <death>, <event>, <forename>, <person>, <place>, <placeName>, <region>, <sex>, <surname>...

```
<persName>  
  <forename>Margarita</forename>  
  <surname type="paterno">García</surname>  
  <surname type="materno">González</surname>  
</persName>
```



9. Personalización TEI: ROMA

Esquema y validación

Un documento XML depende de un **modelo abstracto** que especifica la estructura del texto codificado. Este modelo de datos se materializa en un “**esquema**” que puede tener diferentes formatos (DTD, RelaxNG, Schematron, etc.). Si el documento XML cumple los requisitos del esquema, será un **documento válido**.

- Se localiza, en general, en un fichero diferente, de esta manera podemos referirnos a él desde múltiples documentos XML.
- Su función principal es la de **establecer y definir los requisitos** que debe respetar el documento XML.
- Establece la representación formal de:
 - los elementos y atributos permitidos
 - el elemento raíz del documento
 - los nombres de los elementos utilizados
 - los nombres y los tipos de datos (cadena de caracteres, números, etc.)
 - las reglas de anidación
 - a veces, los valores por defecto de los atributos etc.

Esquemas prediseñados

- Todo documento TEI necesita un esquema que establezca una sintaxis precisa. TEI ofrece algunos modelos, como una versión light ([tei_lite.rng](http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_lite.rng)) o una versión con todos los elementos ([tei_all.rng](http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng)) (que es poco recomendable).
- La buena práctica sugiere que cada proyecto construya a su medida un esquema, adoptando los módulos que crea convenientes, eliminando y añadiendo elementos, eliminando y concretando atributos y valores. Así, la labor de marcado será mucho más fácil y consistente.
- Lista de esquemas RelaxNG prediseñados:
<http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/>

El más utilizado en TEI: Relax NG



Al trabajar con XML-TEI, lo más habitual es el uso de esquemas Relax NG, pues son los que permiten más granularidad y flexibilidad. Relax NG responde a las siglas: *REgular LAnguage for XML Next Generation* y su especificación fue publicada por el Consortium OASIS el 3 de diciembre de 2001. Poco después, en 2003, se convirtió en norma ISO (ISO/IEC 19757-2:2003) y en 2008 hubo una nueva publicación que remplazaba la segunda: [ISO/IEC 19757-2:2008](#) *Document Schema Definition Language (DSDL) – Part 2: Regular grammar-based validation – RELAX NG*.

ROMA: la herramientas para construir i modificar esquemas

- Para crear nuestro esquema, disponemos de una herramienta muy útil y fácil de manejar que se llama ROMA.

<http://www.tei-c.org/Roma/>

 **TEI Roma: generating customizations for the TEI**

Más concretamente, Roma permite:

- Generar y modificar un esquema
- Elegir la lengua de la documentación
- Elegir los módulos que nos conviene
- Añadir y eliminar elementos en una determinada clase
- Modificar clases de atributos
- Elegir el tipo de esquema: RelaxNG, Schematron, W3C Schema, DTD
- Generar una documentación (en HTML, por ejemplo)
- Guardar nuestra personalización a través de un

 **Roma: generating customizations for the TEI**

Hands-on...

- Para la creación de un esquema, seguid las instrucciones del fichero: 3.Crear_esquema_Roma.pdf
- Y dirigiros a: <http://www.tei-c.org/Roma/>
- Si no os da tiempo: ¡tranquilos! ¡Nos vemos en el foro del aula virtual!



10. Herramientas y tutoriales

Documentaciones

Ejemplos de documentaciones

- Diccionario de latín medieval [Du Cange](#)
- [Bibliothèque Virtuelle des Humanistes](#)
- [ENRICH Project: European Networking Resources and Information concerning Cultural Heritage](#)
- [Charta](#): Corpus Hispánico y Americano en la red: textos antiguos

Herramientas

Lista de herramientas:

- [Interedtion Wiki. Existing Tools](#)
- [TEI Tools](#)

Herramientas de transformación:

- [Oxgarage](#)
- [Développement de l'ENC](#)

Plataformas y marcos de publicación:

- [TextGrid](#)
- [TEI Viewer](#)
- [TEI Boilerplate](#)
- [V-Machine](#)
- [Tapas project](#)
- [ModVers](#)
- [Ediarium](#)
- [TEI Publisher](#)

Comparación de documentos:

- [Juxta Project](#)
- [TEI Comparator](#)

Tutoriales

- [Learn TEI](#)
- [TEI WIKI](#)
- [TEI by Example](#)
- [WWW Seminars on Scholarly Text Encoding](#)
- [TEI @ Oxford](#)
- [Lou Burnard's GitHub](#)
- [Getting started using TEI](#)
- [Critical Editions Cheatsheet](#)

Ante una duda...

- Recurrir a las Guidelines: <http://www.tei-c.org>
- Si no encontramos una solución clara, escribid a la [lista TEI](#)
- Si se trata de un defecto, se sugiere la mejora en “Feature Request” o “Bug Reports” en Sourceforge: <http://tei.sf.net>

Podéis plantear cualquier duda en el foro o contactarme directamente:
susannalles@gmail.com

¡Gracias!