

# Text Encoding Initiative, lenguaje XML y marcado semántico de textos

## Jornada 1

**Susanna Allés Torrent**  
University of Miami  
<https://susannalles.com/>

**Gimena del Rio Riande**  
CONICET  
<http://hdlab.space/>

# Programa

## Jornada 1

Miércoles, 29 junio 2022

10:00 – 13:30 pm

16:00 – 20:00

**10:00-12:00**

1. Humanidades digitales: Introducción y tipos de proyectos
2. Los datos en humanidades ¿cuáles y para qué?
3. La codificación semántica de los textos: La Text Encoding Initiative (TEI)

**12:00 -12:30 - Pausa**

- a. Instalación Oxygen

**13:30 - 16:00 - Pausa**

**16:00-20:00**

4. El lenguaje XML y las Guías de la TEI
  5. Modelado de los textos y posibilidades de codificación
  6. Manos a la obra: Codificación de un texto modelo (I)
- 18:00 - 18:30 - Pausa**
7. Manos a la obra: Codificación de un texto modelo (II)

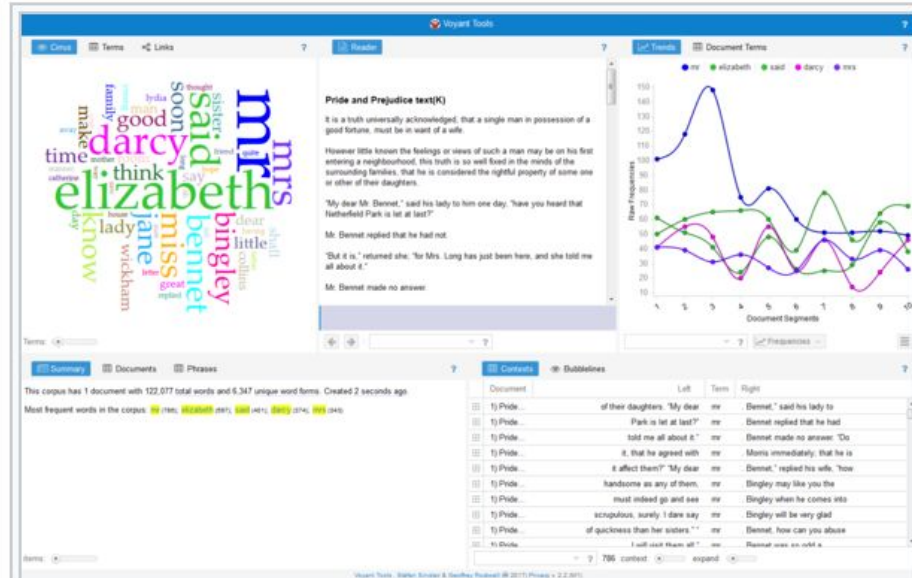
# Sesión 1.

## Humanidades Digitales:

### Introducción y tipos de proyectos

**Digital humanities (DH)** is an area of scholarly activity at the intersection of computing or digital technologies and the disciplines of the humanities. It includes the systematic use of digital resources in the humanities, as well as the analysis of their application.<sup>[1][2]</sup> DH can be defined as new ways of doing scholarship that involve collaborative, transdisciplinary, and computationally engaged research, teaching, and publishing.<sup>[3]</sup> It brings digital tools and methods to the study of the humanities with the recognition that the printed word is no longer the main medium

Las HD constituyen un área de investigación en la intersección de las tecnologías informáticas o digitales y las disciplinas de las humanidades. Las HD persiguen la investigación, docencia y publicación de forma colaborativa, transdisciplinar y computacionalmente comprometida. Las HD aportan herramientas y métodos digitales al estudio de las humanidades con el reconocimiento de que la palabra impresa ya no es el medio principal para la producción y distribución del conocimiento.

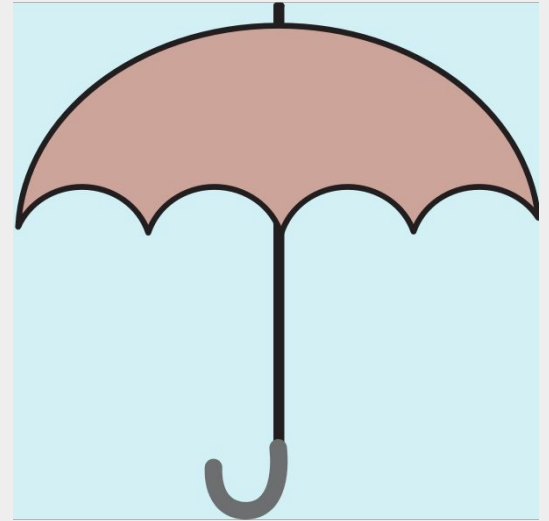


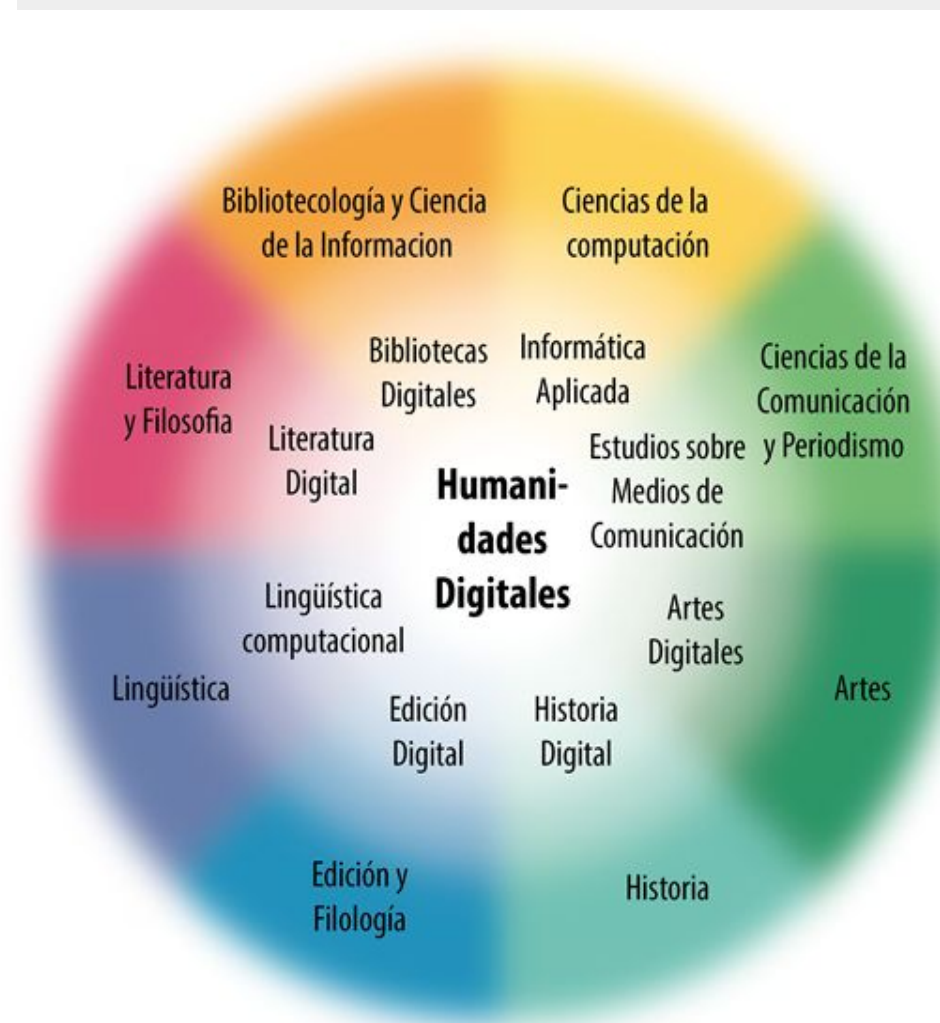
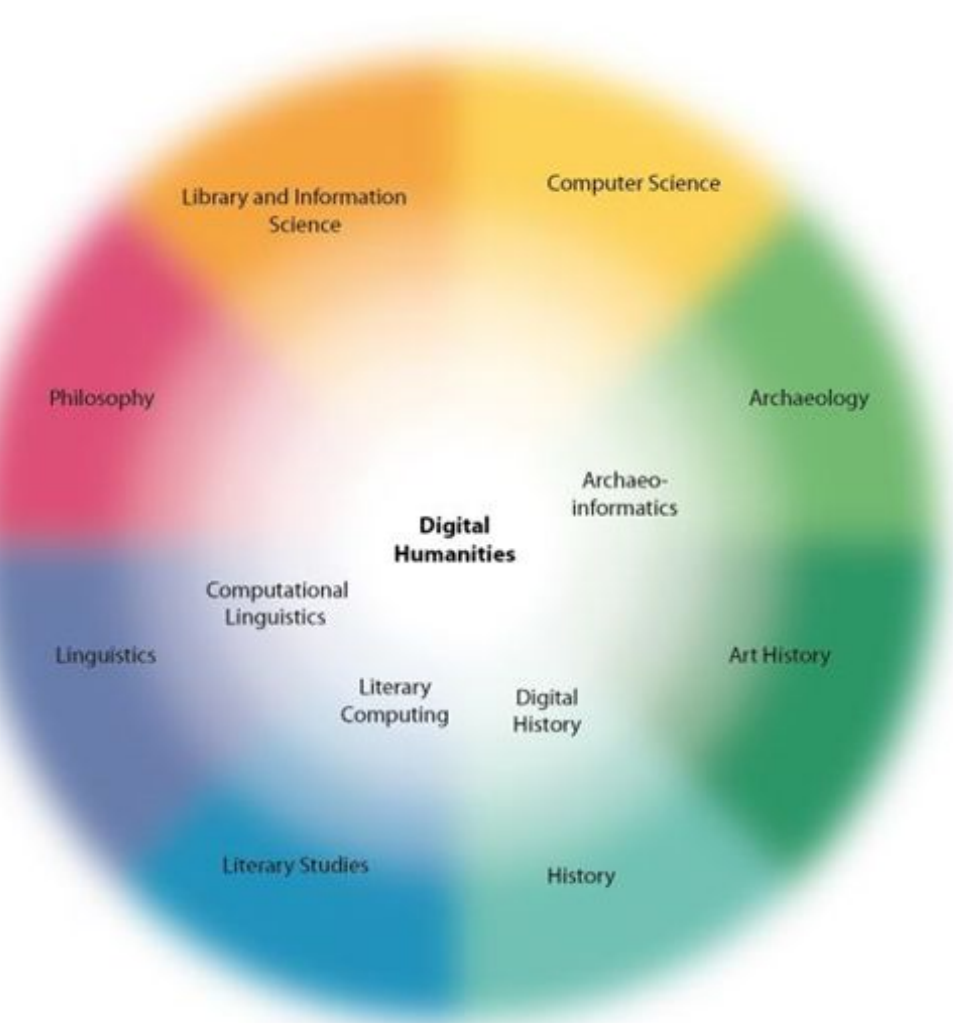
Example of a textual analysis program being used to study a novel, with Jane Austen's *Pride and Prejudice* in Voyant Tools

## What is digital humanities?

“DH is an umbrella term that, depending on who you are talking to, covers a huge territory: everything from applied text analysis and corpus stylistics to more esoteric and theoretical realms of video game criticism”

—Matthew Jockers, “Day of DH 2010”  
(*Defining Digital Humanities*)





# TEI como práctica consolidada

“If one humanities computing activity is to be highlighted above all others, in my view it must be the TEI [Text Encoding Initiative]. It represents the most significant intellectual advances that have been made in our area, and has influenced the markup community as a whole.”



Susan Hockey. “The History of Humanities Computing,”  
*Companion to Digital Humanities* (2003)



***Let's Explore***



# Algunos ejemplos de TEI en uso. Ediciones digitales

Las cartas de Van Gogh: <http://vangoghletters.org/vg/>

La dama boba: <http://damaboba.unibo.it/index.html>

# Algunos ejemplos de TEI en uso

Otra forma de ver los textos:

Relación de un viaje al Río de la Plata:

<http://hdlab.space/Relacion-de-un-viaje/>

Drácula:

<https://beeplosion.github.io/TheDracIsBack/index.html>

## Algunos ejemplos de TEI en uso. Crítica textual y crítica genética

*De quinque septenis* (Hugo de San Víctor):

<https://scholarlyediting.org/2016/editions/intro.dunning.html>

Manuscrito digital de Juan Goytisolo:

<http://goytisolo.unibe.ch/index.html>

# Algunos ejemplos de TEI en uso para seguir explorando

Bestiario de Aberdeen:

<https://www.abdn.ac.uk/bestiary/>

Proyecto Hesperia:

<http://hesperia.ucm.es/index.php>

Mini Lazarillo:

[Minilazarillo.github.io](http://Minilazarillo.github.io)

Poesía Medieval:

<http://hdlab.space/Poesia-Medieval/>

# Sesión 2: Los datos en humanidades ¿Cuáles y para qué?

11:00 - 12:00am

# Tradicionalmente, ls humanists se resisten a hablar de datos

- “Fuentes primarias” = Textos, artefactos, objetos de estudio

Más a menudo, son objetos mediados y contextualizados de alguna manera (es decir, una edición, transcripción o similar)

- “Fuentes secundarias” = Trabajos de otros académicos (a menudo basados en “Fuentes primarias”)

“Lecturas” (1) = Pasajes, extractos, citas para interpretación o apoyo

“Lecturas” (2) = Interpretación, producto final de la investigación (estudio literario)

# Los datos en humanidades

- Generalmente recogidos y no generados
- Complejos: anotaciones, comentarios, textos híbridos
- Multiformato, multilingües
- Objetos de estudio y de referencia (el mismo objeto puede usarse para diferentes fines)
- Crean datasets heterogéneos
- La investigación en Humanidades muchas veces apunta más a explorar preguntas que a resolverlas (interés en los procesos y no en los productos) (Julianne Nyhan)

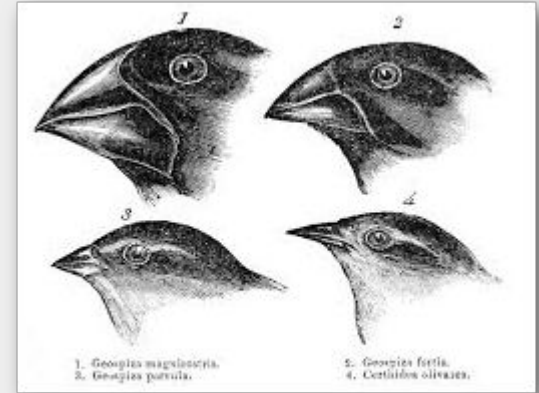
# ¿Cómo funcionan los datos en otros campos?

La resistencia tiene sentido, porque los datos de Humanidades son diferentes de otras formas de datos

En otros dominios, "datos" ("cosas dadas") son más propiamente "capta" ("capturado, tomado"): generado a través de experimentos, observación y medición, luego observado

Piensa en Darwin y su trabajo en las Islas Galápagos

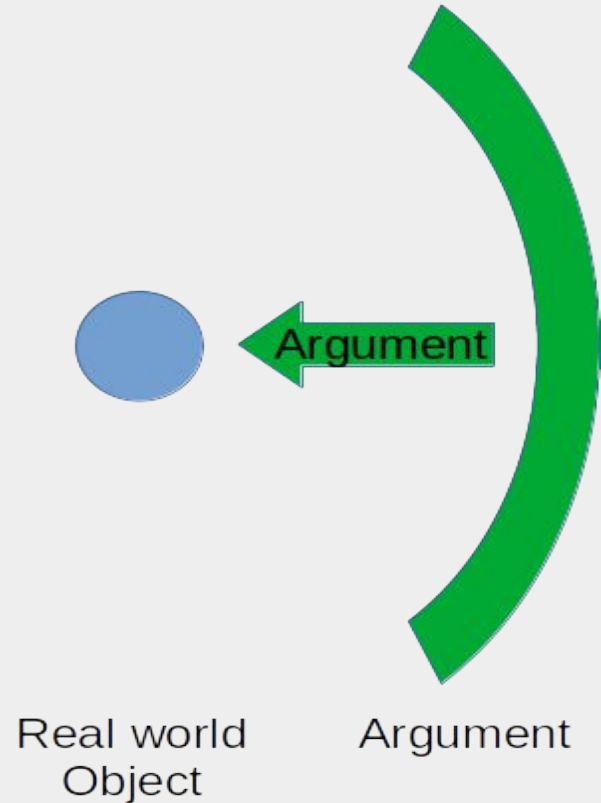
¿Cuáles son sus datos?





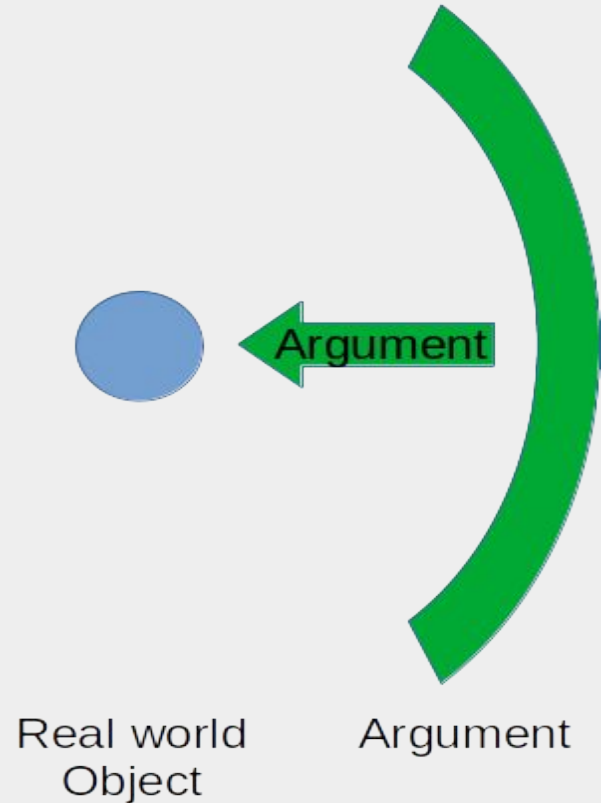
# Pero en las Humanidades

- Pueden ser tanto "data" como "capta"
- Se trata de datasets muy específicos y a menudo provisionales y de pequeño tamaño
- Dependen de la interpretación y el argumento (discutir si algo son datos)
- Los humanistas visitamos con frecuencia los mismos conjuntos de datos para verlos, analizarlos de manera diferente, proporcionando nuevos contextos, reutilizándolos



# Y en las Humanidades Digitales

- Podemos tener "capta" ("observaciones" extraídas algorítmicamente para formar grandes conjuntos de datos que luego requieren interpretación)
- Se introduce la posibilidad de trabajo deductivo.
- Hace que las preguntas de método sean más importantes que cuando trabajábamos de forma inductiva a partir de las colecciones a las que podía acceder



# Sesión 3:

## La codificación semántica de los textos: La Text Encoding Initiative

1. Algunas reflexiones
2. Qué es la TEI
3. Breve historia
4. Organización del consorcio TEI
5. Página web de TEI
6. Pros y contras

## Antes de empezar, reflexionemos...

¿En qué consiste la digitalización de un texto?

¿Qué necesitamos para transformar un texto en formato digital?

¿Qué entendemos por ‘texto’? \*

¿Cómo transformamos un texto en datos informáticos?

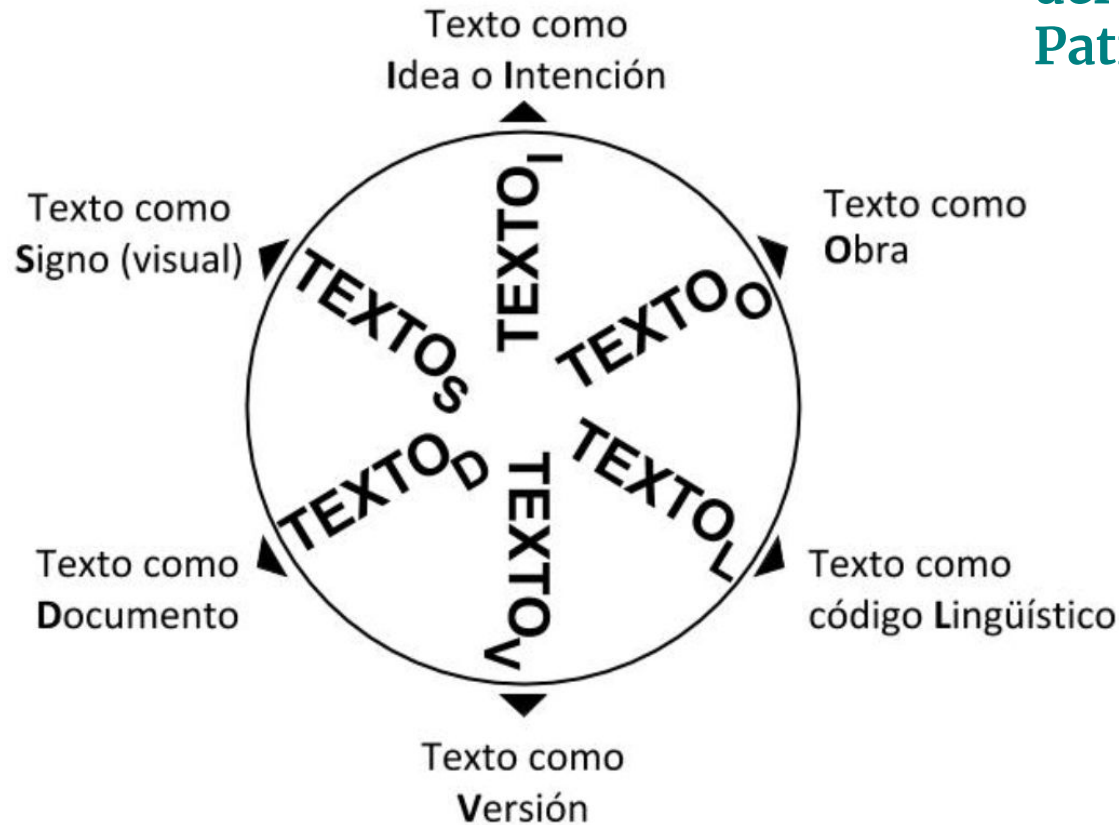
¿Qué queremos ‘hacer’ con el texto digital?

¿Qué significa ‘marcado’ o ‘etiquetado’ de textos? \*

¿En qué consiste un ‘marcado semántico’? \*

# Una noción pluralística del texto

Patrick Sahle (2015)



SÍGUESEN LAS GENERACIONES, SEMBLANZAS E OBRAS DE LOS REYNTES REYES DE ESPAÑA DON ENRIQUE EL TERCERO E DON IOHAN EL SEGUNDO E DE LOS VENERABLES PERLADOS E NOTABLES CAI ALLEROS QUE EN LOS TIEMPOS DESTOS REYES FUERON. 5

Muchas vezes acaçe que las coronicas e estorias que fablan de los poderosos reyes e notables príncipes e grandes çibdades, son aidadas por sospechosas e inçiertas e les es dada poca fe e abtoridat, lo çual, entre otras çabsas, acaçe e biene por dos: la primera, porque algunos que se entremeten de escribir e notar las antigüedades son ombres de poca vergueña, e mas les plazet relatar cosas estrañas e maravillosas que verdaderas e çiertas, creyendo que non sera auida por notable la estoria que non contare cosas muy grandes e graves de creer así que sean mas dignas de maravilla que de fe, como en otros nuestros tiempos fizó un liuiano e pre-

5. A excepción de esta larga rúbrica al frente del prohemio, faltan los encabezamientos en todos los capítulos del ms. Esc. En el Mat. se denominan «Raçamientos».

18. «otros nuestros». Así también en Mat. Falta *otres* en la ed. princeps. Liaguano sustituyó dicha palabra por *estes*.

SÍGUESEN LAS GENERACIONES, SEMBLANZAS E OBRAS DE LOS REYNTES REYES DE ESPAÑA DON ENRIQUE EL TERCERO E DON IOHAN EL SEGUNDO E DE LOS VENERABLES PERLADOS E NOTABLES CAI ALLEROS QUE EN LOS TIEMPOS DESTOS REYES FUERON. 5

Muchas vezes acaçe que las coronicas e estorias que fablan de los poderosos reyes e notables príncipes e grandes çibdades, son aidadas por sospechosas e inçiertas e les es dada poca fe e abtoridat, lo çual, entre otras çabsas, acaçe e biene por dos: la primera, porque algunos que se entremeten de escribir e notar las antigüedades son ombres de poca vergueña, e mas les plazet relatar cosas estrañas e maravillosas que verdaderas e çiertas, creyendo que non sera auida por notable la estoria que non contare cosas muy grandes e graves de creer así que sean mas dignas de maravilla que de fe, como en otros nuestros tiempos fizó un liuiano e pre-

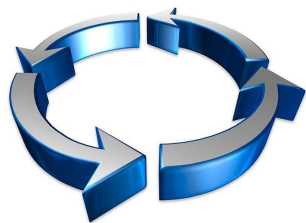
5. A excepción de esta larga rúbrica al frente del prohemio, faltan los encabezamientos en todos los capítulos del ms. Esc. En el Mat. se denominan «Raçamientos».

18. «otros nuestros». Así también en Mat. Falta *otres* en la ed. princeps. Liaguano sustituyó dicha palabra por *estes*.

```
56 <text xml:id="FPG_00" xml:lang="es">
57 <body>
58 <head>Prólogo</head>
59 <p><pb edRef="#DBordonaFPG" n="3"/>Síguense las generaciones, semblanzas
60 eçlentes <roleName>Reyes de España</roleName> don <persName>Enrique e
61 don <persName>Johan el Segundo</persName> e de los venerables perlados
62 que en los tiempos destos reyes fueron.</p>
63 <p>Muchas vezes acaçe que las coronicas e estorias que fablan de los po
64 notables príncipes e grandes çibdades, son avidas por sospechosas e in
65 poca fe e abtoridat, lo çual, entre otras çabsas, acaçe e biene por da
66 algunos que se entremeten de escribir e notar las antigüedades son onbi
67 e más les plazet relatar cosas estrañas e maravillosas que verdaderas e
68 non será avida por notable la estoria que non contare cosas muy grande:
69 así que sean más dignas de maravilla que de fe, como en otros nuestro:
70 liviano e presuntooso <pb edRef="#DBordonaFPG" n="4"/> onbre, llamado .
71 ref="#Pedro_de_Corral">Pedro de Coral</persName>, en una que se llame
72 Serrazina</title> (otros la llamavan <title>del <persName>Rey Rodrigo
73 que más propiamente se puede llamar trufa o mentira paladina: por lo ç
74 tiempo se platicase en <placeName>Castilla</placeName> aquel muy notabi
75 en el tiempo antiguo que <placeName>Roma</placeName> usava de grant po
```

## Obra física

## Digitalización Haithi Trust (OCR, pdf)



## ArchBio

### Generaciones y semblanzas

Fernán Pérez de Guzmán

Home > Works > Generaciones y semblanzas > Index

#### Index

- Prólogo
- Enrique III
- Catalina de Lancaster
- Fernando I
- Ruy López Dávalos
- Alonso Enriquez
- Pero López de Ayala
- Diego López de Sotoliga
- Diego Hurtado de Mendoza
- Gonzalo Núñez de Guzmán
- Juan García Manrique
- Juan de Velasco
- Sancho de Rojas

Text XML-TEI

#### Legend

- People
- Places
- Role
- Family name

Top  
Meta

### Prólogo

Síguense las generaciones, semblanzas e obras de los eçlentes Reyes de España don Enrique el Tercero e don Johan el Segundo e de los venerables perlados e notables cavalleros que en los tiempos destos reyes fueron.

Muchas vezes acaçe que las coronicas e estorias que fablan de los poderosos reyes e notables príncipes e grandes çibdades, son avidas por sospechosas e inçiertas e les es dada poca fe e abtoridat, lo çual, entre otras çabsas, acaçe e biene por dos: la primera, porque algunos que se entremeten de escribir e notar las antigüedades son ombres de poca vergueña, e más les plazet relatar cosas estrañas e maravillosas que verdaderas e çiertas, creyendo que non será avida por notable la estoria que non contare cosas muy grandes e graves de creer así que sean más dignas de maravilla que de fe, como en otros nuestros tiempos fizó un liviano e presuntooso onbre,

```
<text xml:id="FPG_02" xml:lang="es">
  <body>
    <head>Doña Catalina de Lancaster</head>
    <p><pb edRef="#DBordonaFPG" n="19"/>La <persName>reina doña Catalina</persName>, muger d'este
      <persName>rey don Enrique</persName>, fue fija de don <persName>Johan de
      <persName>Lencastre</persName>, fijo ligítimo del <persName>rey Aduarte de Inguelaterra</persName>,
      el cual duque caso con doña <persName>Costança</persName>, fija del rey don <persName>Pedro
      de Castilla</persName> e de doña <persName>María de Padilla</persName>.</p>
    <p>Fue esta reina alta de cuerpo e muy gruesa, blanca e colorada e rubia. En el talle e meneo
      del cuerpo tanto parecía onbre como muger. Fue muy onesta e guardada en su presona e fama,
      liberal e manífica, pero muy sometida a privados e muy regida d'ellos, lo cual, por la mayor
      parte, es biçio común de los reyes. No era bien regida en su presona; ovo <pb
      edRef="#DBordonaFPG" n="20"/> una grande dolencia de perlesia, de la cual non quedó bien
      suelta de la lengua nin libre del cuerpo.</p>
    <p>Murió en <placeName>Valladolid</placeName> en hedad de çinquenta años, <date when="1418"
      >año de mill e quatroçientos e dies e ocho años</date>.</p>
  </body>
</text>
```

# ¿Qué es la Text Encoding Initiative (TEI)?

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) <https://tei-c.org/>

Las [TEI Guidelines for Electronic Encoding and Interchange](https://tei-c.org/) [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.

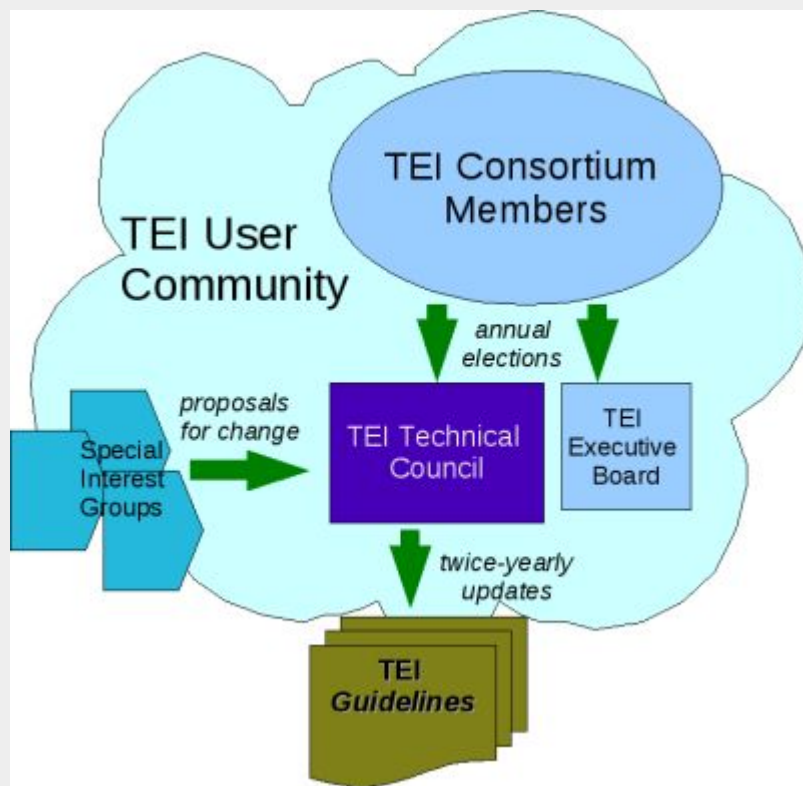
# Puntos clave

1. Las Guidelines son el núcleo central y constituyen 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 [GitHub](#).

# 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 4.4.0
- 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](#)

# Organización de TEI

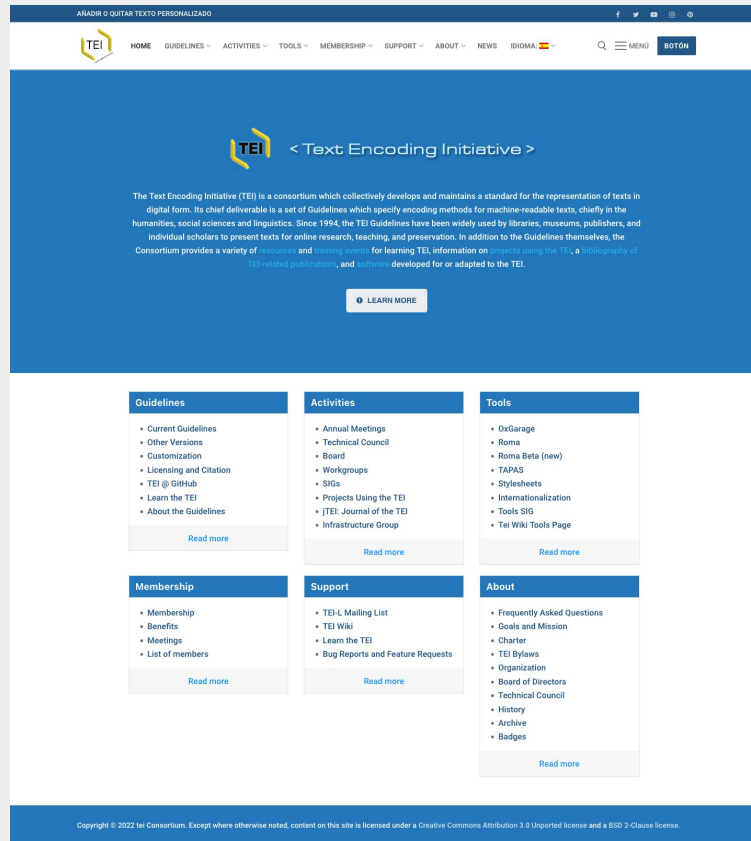


1. Estructura oficial desde el año 2000 con la creación del Consorcio TEI cuyos objetivos ([Goals and Mission](#))
  - Desarrollo de las Guías directrices TEI.
  - Difusión y promoción de las Guías directrices TEI.
  - Formación y divulgación
  - Promoción de una comunidad de investigación TEI.
- ÿ2. Organización, sin ánimo de lucro, autofinanciada por sus miembros y suscriptores (instituciones, tales como bibliotecas, universidades, proyectos académicos, unidades de investigación) y personas.
- ÿ3. [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
- ÿ4. Dos niveles de trabajo:
  - [TEI Workgroups](#)
  - [Special Interest Groups](#)
- ÿ5. Gran comunidad de usuarios conectada a través de listas de correo electrónico y de espacios participativos como la Wiki.
- ÿ6. [Conferencia anual](#)

# Página Web en transición:



The screenshot shows the old website layout. At the top is a blue header with the TEI logo and the text '< Text Encoding Initiative >'. Below this is a navigation bar with links for Home, Guidelines, Activities, Tools, Membership, Support, About, and News. The main content area has a title 'Text Encoding Initiative' followed by two paragraphs of text. The first paragraph describes the TEI as a consortium that develops and maintains a standard for the representation of texts in digital form. The second paragraph describes the TEI Consortium as a nonprofit membership organization. Below the text are social media links for Twitter (@TEIConsortium), a Code of Conduct, and a link to Contact the TEI. At the bottom right is the Huma-Num logo.



The screenshot shows the new website layout. At the top is a dark blue header with the TEI logo and the text '< Text Encoding Initiative >'. Below this is a navigation bar with links for HOME, GUIDELINES, ACTIVITIES, TOOLS, MEMBERSHIP, SUPPORT, ABOUT, NEWS, IDIOMA, and a search icon. The main content area has a title '< Text Encoding Initiative >' followed by a paragraph of text and a 'LEARN MORE' button. Below this is a grid of six boxes, each with a title and a list of items. The boxes are: Guidelines, Activities, Tools, Membership, Support, and About. Each box has a 'Read more' link at the bottom. At the bottom of the page is a dark blue footer with copyright information.



<https://tei-c.org/>

# TEI: El lenguaje utilizado en la mayoría de proyectos de edición digital

## PROS:

- Standard web y de 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
- Trabajo laborioso que hoy tiende a automatizarse perdiendo su objetivo de “comprensión semántica” del texto

## 2. Campos de aplicación y ejemplos

1. Campos de aplicación
2. Catálogos de ediciones digitales
3. Ejemplos de proyectos que utilizan TEI

# Algunos campos de aplicación

1. Edición de textos
2. Bibliotecas virtuales
  - [Biblioteca Pensamiento Novohispano](#)
  - [Bibliothèques Virtuelles des Humanistes](#)
3. Archivos virtuales: [The Shelley-Godwin Archive](#), [Archive of Biographical Writings](#)
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...](#)



## Dos catálogos de ediciones digitales:

- Catalogue of Digital Editions (Greta Franzini), <https://dig-ed-cat.acdh.oeaw.ac.at/>
- A Catalog of Digital Scholarly Edition (Patrick Sahle), <https://www.digitale-edition.de/index.html>

Otras listas con proyectos que usan TEI:

- Projects Using TEI, <https://tei-c.org/Activities/Projects/>

## Más ejemplos:

- [Les manuscrits de Madame Bovary](#) (U. de Rouen et al.)
- [Folger Digital Library](#) (Folger Shakespeare Library)
- [Jane Austen Fiction Manuscripts](#) (U. Oxford – King's College)
- [The World of Dante](#) (U. of Virginia)
- [Vincent van Gogh. The Letters](#) (Huygens ing – Van Gogh Museum)
- [Biblioteca Virtual Miguel de Cervantes](#) (U. de Alicante)

# TTHUB

Text Technologies Hub. Recursos sobre tecnologías del texto y edición digital

ACERCA DE TTHUB APRENDE ÚNETE RECURSOS CONTACTO

APRENDE



LECCIONES



AYUDA A TRADUCIR



BÚSQUEDA

Search ...



CATEGORIAS

Elegir la categoría ▾

ARCHIVO

Elegir el mes ▾



RECURSOS



EJEMPLOS DE  
CODIFICACIÓN



BIBLIOGRAFÍA EN  
ESPAÑOL



<https://tthub.io/>

THE  
ANDREW W.  
**MELLON**  
FOUNDATION

PROGRAMS

**Public Knowledge**

UNIVERSITY OF MIAMI  
COLLEGE of  
ARTS & SCIENCES



CONICET



TTHub

Powered by WordPress & Bezel Theme - Developed by Susanna Allés Torrent

Sobre tthub Política de Privacidad Términos y Condiciones Política de Cookies Contacto

12:30 - 13:30am



## Actividad 1: Instalación del programa Oxygen

Instalación de Oxygen

[https://www.oxygenxml.com/xml\\_editor/download\\_oxygenxml\\_editor.html](https://www.oxygenxml.com/xml_editor/download_oxygenxml_editor.html)

Instrucciones:

<https://github.com/tthub-repo/ejercicios/blob/master/1.Instrucciones-instalacion-oxygen.md>

<https://github.com/tthub-repo/ejercicios/blob/master/2.Creacion-doc-xml.md>

[http://tthub.io/UCV/docs\\_Valencia.zip](http://tthub.io/UCV/docs_Valencia.zip)

## Actividad 2: ¿Cómo se codificaría un poema en lenguaje XML? Ideas...?

¿Qué tipo de texto es?

¿Quién es el autor?

¿Cuál es el título?

¿Qué tipo de composición poética es?

¿Cómo se estructura?

¡Proponed una codificación en XML inventando vosotros mismos las etiquetas!

Tomad en consideración este poema de Quevedo:

*Un soneto me manda hacer Violante*

Lope de Vega

Un soneto me manda hacer Violante,  
y en mi vida me he visto en tal aprieto:  
Catorce versos dicen que es Soneto:  
Burla burlando van los tres delante.

Yo pensé que no hallara consonante  
y estoy a la mitad de otro cuarteto:  
Mas si me veo en el primer terceto  
no hay cosa en los cuartetos que me espante.

Por el primer terceto voy entrando  
y aún parece que entré con pie derecho,  
pues fin con este verso le voy dando.

Ya estoy en el segundo, y aun sospecho  
que estoy los trece versos acabando:  
contad si son catorce, y está hecho.

# Sesión 4:

## El lenguaje XML y las Guías de la TEI

1. Lenguaje XML
2. Concepto de marcado
3. “Leyes” del XML
4. Estructura de un documento TEI
5. Las Guías directrices
6. Sesión práctica

# 1. El lenguaje XML

Al trabajar con TEI,  
se utiliza el lenguaje  
XML



```
<?xml version="1.0" encoding="UTF-8"?>
<lista>
  <libro>
    <autor id="esquilo">Esquilo</autor>
    <titulo>Prometeo encadenado</titulo>
  </libro>
  <libro>
    <autor id="sofocles">Sófocles</autor>
    <titulo>Edipo Rey</titulo>
  </libro>
  <libro>
    <autor id="euripides">Eurípides</autor>
    <titulo>Medea</titulo>
  </libro>
</lista>
```

[https://github.com/tthub-repo/ejemplos/blob/master/L2\\_ejemplo-1.xml](https://github.com/tthub-repo/ejemplos/blob/master/L2_ejemplo-1.xml)

## 2. El marcado / La codificación

- Noción de Marcado / Codificación = “Markup” (en.), “Balisage” (fr.), “Codifica” (it.) consiste en aislar una porción de texto con caracteres que son interpretados por la máquina como código y que poseen un significado semántico específico.

ÿ

- 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>
```



# Elementos de una etiqueta

- **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 ficticios:

```
<fecha cuando="2022">2022</fecha>  
<division tipo="prologo">El prólogo...</division>  
<frase tipo="enunciativa">Esto es una frase</frase>  
<pregunta tipo="interrogativa">¿Cómo te llamas?</pregunta>
```

### 3. Las “leyes” del XML

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

- Y el nombre del elemento puede solo contener letras y números, sin acentos ni espacios, y solo algunos signos: - \_ .
- Y el nombre no puede empezar por un número, ni por un signo, ni por `<xml`.
- Y ¡XML es sensible a las mayúsculas y las minúsculas! Y
- Los documentos XML-TEI deben:
  - Y Estar bien formados, es decir estar expresado correctamente según la gramática de XML.
  - Y 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).

“El lenguaje XML y algunos conceptos generales”, TTHUB (2019),  
<https://tthub.io/aprende/l2-xml/>

# N.B. Formato VS Contenido

## Marcado de presentación

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

```
<b>Título</b>
```

```
<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>
```

## 4. La estructura de un documento TEI

1. Estructura mínima
2. Encabezado `<teiHeader>`
3. Cuerpo del documento `<text>`
4. Divisiones `<div>`

Abrid el documento: `2.Ejemplo_minimo_XML-TEI.xml` para seguir la explicación

# Estructura mínima

1. Prólogo XML:
  - a. La declaración XML nos indica que se trata de un documento XML y que el tipo de codificación utilizado es el UTF-8.
  - b. 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"?>
```

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>
```

- Las divisiones que corresponden a la primera parte, a los apartados (1, 2...) y a los subapartados (1.1., 1.2., etc.)
- ÿ Los títulos o encabezados
- ÿ Los párrafos
- ÿ Las citas bibliográficas
- ÿ Las listas
- ÿ Los números de página
- ÿ Las referencias cruzadas
- ¿Cómo expresarlo según las TEI Guidelines?

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="mi_esquema.rng"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Informaciones del archivo digital</title>
      </titleStmt>
      <publicationStmt>
        <p>Detalles de la publicación digital</p>
      </publicationStmt>
      <sourceDesc>
        <p>Información de la fuente a transcribir</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Aquí va vuestro texto codificado en XML-TEI</p>
    </body>
  </text>
</TEI>
```

# Encabezado <teiHeader>

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Título del archivo
digital</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>
```

## 5. Las guías directrices

1. ¿Qué son las Guías directrices?
2. Funcionamiento y formatos de publicación
3. Iniciativa para su traducción

# ¿Qué son las Guías directrices?

- 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 en 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



# Contenido 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





## TEI Specification Translator

### Getting started

Clicking on "Browse" or "Settings" above will first ask you to log in via [GitHub](#), if you haven't already done so. You can create a GitHub account if you don't already have one.

After you log in you'll be taken to the Settings page. Before you can save your work, you'll need to choose a repository to save to. This must be a fork of the TEI repo. It can be in your own account or in a GitHub organization you're a member of.

The next step is to choose or create a branch to save into. Branches can have any name that doesn't conflict with a branch name in the TEI repo. If you already have a branch you want to work in, you can type or choose its name. If you type a new name, you'll be prompted to create the branch. You should then be able to choose what you want to work on.

The browse view gives you lists of elements, attribute classes, model classes, macros, and datatypes. For each language, there will be icons telling you what the state of the translations are. A  means everything has been translated. A  means all translations are up to date. A  means there are missing translations. A  means there are out of date translations.

Clicking on a name link will take you to the specification page for that item. You can then choose a language to work in, enter your translations, and save them when you're done.

Saving your work creates a commit on GitHub, in your working branch. You can visit your repo on GitHub to see all the changes you've made via the link on the browse page. When you're ready to submit your work, click the "Submit Your Work" button, and a Pull Request will be created for review on the TEI repo. You can continue to save changes until the Pull Request is closed.

### Setting up a translation campaign

If you want to collaborate with a group in translating pages, you will first need to set up an [organization](#) on GitHub, create a fork of [TEI/TEI](#) within it, and then invite your collaborators to work on the fork (see [Managing access to your organization's repositories](#)). When your collaborators sign in to [translate.tei-c.org](#), they will be able to enter the organization name instead of their username to access the organization's fork. You will want to set up a branch for everyone to work in and let your collaborators know which branch to use.

# Iniciativa en marcha para la traducción a diferentes lenguas - TEI Specification Translator:

[https://translate.tei-c.org/doc\\_en.html](https://translate.tei-c.org/doc_en.html)

# ¿Cómo empezar a marcar y saber qué marcar?

## El modelado de datos

Modelización es es la actividad de creación de una representación abstracta de un fenómeno complejo, por lo general en una forma legible mecánicamente y posiblemente de una manera interactiva (es decir, incluyendo una simulación). Los modelos se convierten en legibles por máquina cuando el modelado produce un esquema que describe los elementos y la estructura de un objeto de investigación de manera explícita. La modelización también puede referirse a la actividad de transformar o manipular un objeto digital de una manera tal como para que sea compatible con un modelo o esquema previamente construido. El mapeo, por ejemplo, es un ejemplo de modelo espacial. El Diseño de flujo de trabajo se incluye como parte de la modelización, utilizando un objeto tal como Proceso.

TaDiRAH - Taxonomía sobre Actividades de investigación digital en humanidades  
<https://www.vocabularyserver.com/>



# Modelado 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...

Los documentos XML-TEI dependen de un ESQUEMA, es decir, de otro documento (normalmente expresado también en XML) que establece la sintaxis de una codificación concreta.

Símil:

- XML - Lengua
- TEI - Gramática
- Esquema - Sintaxis

```
<?xml-model href="https://archbio.miami.edu/corpus/archbio.rng"
schematypens="http://relaxng.org/ns/structure/1.0"?>
```

## ¿Para qué sirven los esquemas?

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.

# Un documento XML-TEI siempre va relacionado a un esquema

Documento Codificación XML

 XML Document



Documento Esquema RNG

 RELAX NG Schema - XML

<https://www.oasis-open.org/committees/relax-ng/charter.php>

```
<?xml-model href="/esquema/nuestro_esquema.rng"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
...
</TEI>
```

Ejemplo simple de un esquema RNG comentado:

[https://github.com/tthub-repo/ejemplos/blob/master/L6\\_Ejemplo\\_simple\\_comentado.rng](https://github.com/tthub-repo/ejemplos/blob/master/L6_Ejemplo_simple_comentado.rng)

# Sesión 6

## Manos a la obra: Codificación de un texto modelo en XML-TEI

## Actividad 3: Creación de un documento XML-TEI

Para este primer ejercicio práctico seguiremos los pasos propuestos en este ejercicio “Creación de un documento TEI” <https://tthub-repo.github.io/ejercicios/3.Creacion-doc-tei>

Objetivos:

- Crear un documento bien formado
- Crear un documento válido según el esquema TEI\_All
- Familiarizarse con el programa Oxygen (vista elementos, indentación, validación)
- Creación de un comentario `<!-- un comentario -->`

## Parte 2: Codificación de un poema

A partir de la actividad anterior y del documento 3.Ejercicio\_codificacion\_soneto.xml codifica el siguiente poema de Francisco de Quevedo:

*Es hielo abrasador, es fuego helado,  
es herida que duele y no se siente,  
es un soñado bien, un mal presente,  
es un breve descanso muy cansado.  
Es un descuido que nos da cuidado,  
un cobarde, con nombre de valiente,  
un andar solitario entre la gente,  
un amar solamente ser amado.  
Es una libertad encarcelada,  
que dura hasta el postrero parasismo,  
enfermedad que crece si es curada.  
Este es el niño Amor, este es su abismo.  
¡Mirad cuál amistad tendrá con nada  
el que en todo es contrario de sí mismo!*

- ¿Qué elementos estructurales tiene el texto?
- ¿Qué etiquetas o elementos TEI podríamos utilizar para marcarlos?
- ¿Qué elementos semánticos tiene el texto?
- ¿Cuáles crees que serían interesantes para este texto?
- ¿Qué elementos deberíamos utilizar para marcarlos?

Usaremos el esquema `tei_all`

Descargarlo aquí: [bit.ly/teiall](https://bit.ly/teiall)



# ¡Gracias y nos vemos mañana!

Podéis contactarnos a: [susanna\\_alles@miami.edu](mailto:susanna_alles@miami.edu) o  
[gdelrio@conicet.gov.ar](mailto:gdelrio@conicet.gov.ar)

# Text Encoding Initiative, lenguaje XML y marcado semántico de textos

## Jornada 2

**Susanna Allés Torrent**  
University of Miami  
<https://susannalles.com/>

**Gimena del Rio Riande**  
CONICET  
<http://hdlab.space/>

# Programa

## Jornada 2

Jueves, 30 junio 2022

10:00 – 13:30 pm

16:00 – 20:00

**10:00-12:00**

1. ¿Qué son y cuál es la finalidad de los “esquemas”?
2. Manos a la obra: Personalización de un esquema
3. La función de la ODD
4. Manos a la obra: Personalización de un esquema y una ODD

**12:00 -12:30 - Pausa**

5. Manos a la obra: Codificación de un texto y transformación (1)

**13:00 - 16:00 - Pausa**

**16:00-17:30**

6. ¿Y después de la codificación?  
Escenarios de transformación

**17:30 - 18:00 - Pausa**

7. Manos a la obra: Transformaciones de XML a HTML
8. Balance del curso y turno abierto de preguntas

# 1. El sistema de módulos

# El sistema de módulos establecido por las Guías directrices

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. <https://tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>

## Lista módulos

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 módulos

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



# 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:
  - Cambios editoriales: `<choice>`, `<sic>`, `<corr>`, `<orig>`, `<reg>`, `<add>`, `<del>`, `<gap>`, `<unclear>`
  - 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>
```

# Módulo: verso

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

```
<lg type="soneto">2022
  <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, y lo leo</l>
    <l>tan solo, que aun de vos me guardo en esto.</l>
  </lg>
</lg>
```

# Módulo: 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>
```

# Módulo: diccionario

<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>
```

# Módulo: Representación de fuentes primarias

<damage>, <facsimile>, <fw>, <handNotes>, <handShisft>, <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>
```

# Módulo: 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>Susanna</forename>  
  <surname type="paterno">Allés</surname>  
  <surname type="materno">Torrent</surname>  
</persName>
```

# Sesión 1:

## ¿Qué son y cuál es la función de los esquemas?

# ¿Para qué sirven los esquemas?

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.

# Un documento XML-TEI siempre va relacionado a un esquema

Documento Codificación XML

 XML Document



Documento Esquema RNG

 RELAX NG Schema - XML

<https://www.oasis-open.org/committees/relax-ng/charter.php>

```
<?xml-model href="/esquema/nuestro_esquema.rng"
schematypens="http://relaxng.org/ns/structure/1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
...
</TEI>
```

Ejemplo simple de un esquema RNG comentado:

[https://github.com/tthub-repo/ejemplos/blob/master/L6\\_Ejemplo\\_simple\\_comentado.rng](https://github.com/tthub-repo/ejemplos/blob/master/L6_Ejemplo_simple_comentado.rng)

# Tipos de modelos

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

## Los esquemas RNG

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 reemplaza la segunda: [ISO/IEC 19757-2:2008](#) Document Schema Definition Language (DSDL) – Part 2: Regular grammar-based validation – RELAX NG.

Más detalles sobre los esquemas en:

“Esquemas RNG y personalización de TEI”. TTHub,  
<https://tthub.io/aprende/16-esquemas/>

# La plataforma ROMA: creación de esquemas y ODD

## Roma: generating customizations for the TEI

A new version of Roma is available! It is currently in beta, [try it at romabeta.tei-c.org](https://romabeta.tei-c.org/)

TEI Roma is a tool for working with TEI customizations. A TEI customization is a document from which you can generate a schema, along with customized HTML or PDF documentation of it. The schema generated can be expressed in an ODD.

You can make or modify your TEI customization in several different ways:

- Build up: create a new customization by adding elements and modules to the smallest recommended schema
- Reduce: create a new customization by removing elements and modules from the largest possible schema
- Create a new customization starting from a template
- Use or modify an existing TEI-defined customization
- Upload a customization  No file selected.

Community-maintained customizations can be downloaded from [the TEI website](#)

**Start**

A TEI customization is informally referred to as an ODD (for "One Document Does it all")

Roma was written by Arno Mittelbach and Sebastian Rahtz. Sanity check written by Ioan Bernevig. It is maintained by the TEI Consortium's Technical Council. This is Roma version 5.0.1, last updated 2019-08-25.

Aplicación creada por el Consorcio que permite la creación de esquema en diferentes formatos y la descarga de la ODD correspondiente.

<https://roma2.tei-c.org/>

Hay una nueva versión en curso:

<https://romabeta.tei-c.org/>



## Actividad 4: Creación de un esquema

Pasos para la creación de un esquema personalizado:

<https://tthub.io/aprende/l6-esquemas/#iv-aplicación-roma>

# Sesión 2

## La función de la ODD: *One Document Does it All*

## ¿Qué es una ODD?

El documento ODD consiste en un archivo XML-TEI, a partir del cual podremos generar el esquema en diferentes lenguajes y recuperar la documentación de nuestro proyecto.

- El documento ODD es el único que nos permitirá **recuperar nuestro esquema personalizado**. En cualquier proyecto de edición, especialmente en las primeras etapas, es normal modificar el esquema de codificación, de manera que cada vez que lo modifiquemos deberemos crear también un documento ODD que será el responsable de generarnos el esquema deseado.
- El documento ODD también nos permitirá generar, además de los esquemas, una nueva **documentación** tal y como la habremos redactado para nuestro proyecto.
- En fin, el documento ODD es la pieza clave para poder **difundir y compartir nuestro esquema de codificación**. No solo con la comunidad TEI, sino también entre los mismos miembros de un equipo.

<https://tthub.io/aprende/l7-odd/#i-qué-es-un-odd>

# Un documento XML-TEI siempre va relacionado a un esquema

Documento Codificación XML

 XML Document



Documento Esquema RNG

 RELAX NG Schema - XML

<https://www.oasis-open.org/committees/relax-ng/charter.php>

Documento ODD

 XML Document

<https://tei-c.org/guidelines/customization/getting-started-with-p5-odds/>

Ejemplos simples de un documento ODD:

[https://github.com/tthub-repo/ejemplos/blob/master/L7\\_ODD\\_basicaRoma.xml](https://github.com/tthub-repo/ejemplos/blob/master/L7_ODD_basicaRoma.xml)

[https://github.com/tthub-repo/ejemplos/blob/master/L7\\_ODD\\_ejemploVerse.xml](https://github.com/tthub-repo/ejemplos/blob/master/L7_ODD_ejemploVerse.xml)

## Actividad 5: Creación de una ODD

Pasos para la creación de una ODD:

<https://tthub.io/aprende/l7-odd/#ii-creación-del-odd>

## ¿Qué tipo de texto vamos a codificar?

- Modelado de los datos (textuales)
- Creación de un esquema y personalización (incluyendo los módulos necesarios y haciendo las modificaciones convenientes)
- Creación de la ODD
- Codificación

# Manos a la obra: Personalización de un esquema y una ODD Codificación de un texto en prosa

Indicaciones: “Codificación de la biografía de Catalina de Lancaster (Fernán Pérez de Guzmán, s. XV)”

<https://github.com/tthub-repo/ejercicios/blob/master/5.Codificacion-biografia.md>

Doña Catalina de Lancaster

La reina doña Catalina, muger deste rey don Enrique, fue fija de don Iohan de Lencastre, fijo legitimo del rey Aduarte de Inguelaterra, el qual duque caso con doña Costança, fija del rey don Pedro de Castilla e de doña Maria de Padilla.

Fue esta reina alta de cuerpo e muy gruesa, blanca e colorada e rubia. En el talle e meneo del cuerpo tanto pareçia onbre como muger. Fue muy onesta e guardada en su presona e fama, liberal e manifica, pero muy sometida a priuados e muy regida dellos, lo qual, por la mayor parte, es biçio comun de los reyes. No era bien regida en su presona\*; ouo una grande dolença de perlesia, de la qual non quedo bien suelta de la lengua nin libre del cuerpo.

Murió en Valladolid en hedad de çinquenta años, año de mil e quatroçientos e dies e ocho años.

<https://archbio.miami.edu/works/generaciones-y-semblanzas/index/#245-C>

[atalina de Lancaster](#)



# ¿Y después de la codificación? Escenarios de transformación

# ¿Qué hacemos después con esos documentos XML?

Las respuestas son múltiples y ahí van algunas:

- concebir un sitio web en el que publicar el texto
- transformación del texto en múltiples formatos: desde el archivo XML se puede crear, además de la versión HTML para web, una versión en pdf, txt, epub, doc, etc.
- procesar la información codificada para la creación de índices (por ejemplo, extracción de los nombres de personas o de nombres de lugar)
- análisis del texto en función de la codificación (por ejemplo, estudio de un personaje en una obra a partir de sus palabras en discurso directo o indirecto)
- extracción de información para crear redes (por ejemplo, extraer todos los personajes mencionados por un personaje, o en un capítulo, o en una carta)
- etc.

## Dos lecturas recomendadas:

- Susanna Allés Torrent, “[Edición digital y algunas tecnologías aliadas: codificación TEI y transformaciones XSLT](#)”. *Insula. Revista de Letras y Ciencias Humanas* 822 (June 2015): 18–21
- Susanna Allés Torrent, “Crítica textual y edición digital o ¿Dónde está la crítica en las ediciones digitales?” *Studia Aurea* 14 (2020) <https://doi.org/10.5565/rev/studiaaurea.395>

# Lista de marcos o programas para transformar los documentos XML-TEI en “sitios web”:

- [CETEIcean](#)
- [Ed. A Jekyll Theme for Minimal Editions](#)
- [Ediarium](#)
- [Edictor](#)
- [Edition Visualization Technology](#)
- [eLaborate](#)
- [Entorno Virtual de Investigación](#) (LINHD)
- [Kiln](#) (King’s College London)
- [Tapas Project](#)
- [TEI Boilerplate](#)
- [TEI Critical Apparatus Toolbox](#) (Marjorie Burghart)
- [TEI Publisher](#)
- [TEIViewer](#)
- [Textual Communities](#)
- [TextGrid](#)
- [Version Machine](#)

Para una lista más detallada, véase también «[Ways to Publish TEI on the Web](#)» (Wiki TEI).

# Manos a la obra: Transformación de un documento XML-TEI a HTML

## Actividad 6: Transformación con Oxygen

- Abrid el documento que queráis transformar File > Open y elegid el archivo XML-TEI.
- Id a Document > Transformation > Configure Transformation Scenario
- A continuación, se nos abrirá una ventana en la que elegiréis la transformación “TEI P5 XHTML”, y haréis click en “Apply associated (1)”
- Automáticamente se os abrirá vuestro navegador web, y veréis vuestra primera transformación web.

La hoja XSLT que ofrece Oxygen es realmente muy básica pero os podéis hacer una idea de qué aspecto tendría una transformación simple de vuestro archivo. Alternativamente, se podría usar vuestra propia hoja XSLT para darle el diseño que quisierais y acompañarla además de una CSS para mejorar el diseño.

[https://tthub-repo.github.io/ejercicios/9.Ej\\_transformacion](https://tthub-repo.github.io/ejercicios/9.Ej_transformacion)

# Actividad 7: Transformación de un documento XML-TEI con una hoja XSLT personalizada

Descargad esta carpeta zip: <http://tthub.io/UCV/DocsValencia2.zip>

Tened abiertas las “chuletas” / “machetes”:

- <https://www.w3schools.com/css/>
- <https://www.w3schools.com/html/>

# Editor de código abierto

**Virtual Studio Code**

<https://code.visualstudio.com/>





# Visual Code Studio

Código de colores de VS Code:

- Marrón: <elementos>
- Rojo: atributos=
- Azul: “valores”
- Verde: < - - comentarios - - >

Instrucciones de descarga e instalación

<https://bit.ly/VSCodetei>

Plugins de interés: ScholarlyXML  
Tei Publisher

# ¡Gracias!

Podéis contactarnos a: [susanna\\_alles@miami.edu](mailto:susanna_alles@miami.edu) o  
[gdelrio@conicet.gov.ar](mailto:gdelrio@conicet.gov.ar)