

# Introducción a la edición digital académica en XML-TEI con TEIScribe.

María Gimena del Rio, Elena González Blanco García y Clara Martínez Cantón.

Cita:

María Gimena del Rio, Elena González Blanco García y Clara Martínez Cantón (Noviembre, 2016). *Introducción a la edición digital académica en XML-TEI con TEIScribe. Humanidades Digitales: Construcciones locales en contextos globales. AAHD.*

Dirección estable: <https://www.aacademica.org/gimena.delrio.riande/57>

ARK: <https://n2t.net/ark:/13683/pdea/GAu>



Esta obra está bajo una licencia de Creative Commons.  
Para ver una copia de esta licencia, visite  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>.

*Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.*

I Congreso Internacional

Humanidades Digitales.

Construcciones locales en contextos globales



# Introducción a la edición digital académica

Elena González-Blanco (egonzalezblanco@flog.uned.es)

Universidad Nacional de Educación a Distancia (UNED), Spain

Gimena Del Río Riande (gdelrio@conicet.gov.ar)

Seminario de Edición y Crítica Textual (SECRIT-IIBICRIT CONICET), Argentina

Clara I. Martínez Cantón (cimartinez@flog.uned.es)

Universidad Nacional de Educación a Distancia (UNED), Spain



**POSTDATA**

Poetry Standardization  
and Linked Open Data

**UNED**

**LINHD**  
LABORATORIO DE INNOVACIÓN  
EN HUMANIDADES DIGITALES


**erc**  
European Research Council  
Established by the European Commission  
Supporting top researchers  
from anywhere in the world



**European  
Research  
Council**

# Edición Digital Académica

- Plan / Concepto/ Flujo de trabajo
- Modelización de los datos
- Estándar → TEI
- Arquitectura/estructura
- Open source/ licencias
- Aspecto “user friendly”
- Búsquedas a texto completo / Índices
- Visualización
- Accesibilidad a largo plazo
- Identificadores persistentes
- Documentación

- 
- Planificación y decisiones
  - Modelar los datos (estructura, vocabularios)

- 
- Edición

- 
- Visualización
  - Preservación

# Entorno Virtual de Investigación (VRE)



POSTDATA

Poetry Standardization  
and Linked Open Data



**TextGrid**

Virtual Research Environment  
for the Humanities

# ¿Qué ofrece una edición digital?



7 **D**Wæt ic swefna cýrt seegan wylle hætt mege mette  
8 to midre nihte syðþan reord berend reste wunedon.  
9 þuhte me þæt ic ge sawe syllicre treow onhyt  
10 lædan leohte be wunden beama beorhtost eall þæt  
11 beacen wæx be goten mid golde gimmaþ flodon fegere  
12 æt foldan sceatum . swylce þær siþe wæron uppe  
13 onþam eaxle ge sƿanne be **heoldon** þær engel dryht  
14 neƿ ealle fegere þurh forð ge sceaft newæx ðærhuru  
15 fracodeƿ gealga . ælcine þær **beheoldon** halige gastaƿ  
16 men ofer moldan ƿeall þeof matre ge sceaft .  
17 Syllic wæx se riġe beam ƿic fynnu fih for wounded  
18 mid wommū ge seah ic wuldreƿ treow. wædum geweor  
19 ðode wynnu scinan ge gƿred mid goðde gimmaþ hæfdon  
20 be wriġene weorðlice wealdes treow . hƿæðre ic  
21 þurh þæt gold ongytan meahƿ earmra ærgewin  
22 þæt hit ærest ongan fwetan on þa swiðran healfe  
23 eallic wæx mid furgum gedrefed . Forht icwæx for  
24 þære fegran ġe ƿlðe ge seah ic þæt fure beacen.

# ¿Qué aporta?



POSTDATA

Poetry Standardization  
and Linked Open Data

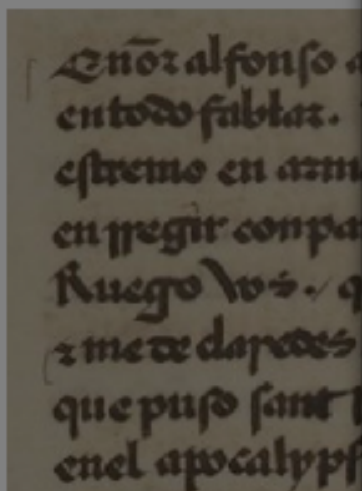
DIALOGO\_MEDIEVAL HOME DOCUMENTACIÓN

Esta pregunta m  
Colunga, de la Ó  
Álvarez de Villas  
oscuras del Apoc



Frey Pedro de Colunga

Se trata de una pieza de arte ma



## Informations

### Poem:

- isometrismo: isométrico
- subtype: pregunta
- asonancia: consonante
- isoestrofismo: isoestrófico
- type: poema
- unisonancia: singular
- rhyme: ABBAACCA
- met: 12, 12, 12, 12, 12, 12, 12, 12

### Stanza 1:

- type: octava
- rhyme: ABBAACCA
- met: 12, 12, 12, 12, 12, 12, 12, 12

### Stanza 2:

- type: octava
- rhyme: ABBAACCA
- met: 12, 12, 12, 12, 12, 12, 12, 12

### Stanza 3:

- type: octava
- rhyme: ABBAACCA
- met: 12, 12, 12, 12, 12, 12, 12, 12

ey Pedro de  
no Alfonso  
unas figuras  
anera

abio perfeto,

a

eto

ro decreto

ión

o.

▼ DIGITALE EDITION

- Convolut I
- Convolut II**
- Convolut III
- Convolut IV
- Convolut V
- Convolut VI
- Convolut VII
- Convolut VIII
- Convolut IX
- Convolut X
- Convolut XI
- Convolut XII
- Convolut XIII

► SUCH

**Metadaten**

Convolut II, Seite 5  
 Alte Zählung:  
 Conv. II, Halbbogen I, 1  
 Akademieausgabe:  
 XXI, 161,02-162,32

Auf dem Manuskript:  
 Neue Zählung: 5  
 Alte Zählung: 2. Conv. I, 1.  
 Signatur Kant: No. 1.  
 Andere Signatur: 2

Seite drucken

Feensteranzeige:

- Linkes Fenster
- Rechtes Fenster
- Beide Fenster

Convolut II P\_5

TRANSCRIPTION

2 Von den metaphysischen Anfangsgründen der Naturwissen-  
 3 schaft (*philosophia naturalis*) \*1 geht die Tendenz i<der Philosophie> auf die  
 Physik

4 Naturforschung (*Physica*) als einem System der empirischen  
 5 Naturkunde, als auf welche jene abzweckt; den̄ was man die  
 6 (mit Newton) die mathematische Anfangsgründe der Naturwissenschaft nen̄  
 7 ist zwar auch die \*2 gleichfalls auf Principien *a priori* ist aber nicht  
 8 ein Erkenntnis aus Begriffen sondern aus der Construction derselben  
 9 (den jenen correspondirenden reinen Anschauungen) und macht keinen  
 10 Theil der Naturphilosophie aus sondern ist \*3 hiebey  $\emptyset$  ~~überwicht~~ nur ein  
 11  $\emptyset$  ~~ie~~ wie wohl vortreffliches ja unentbehrliches Instrument zu dieser  
 12 ihrem Zwecke zu gelangen;  $\emptyset$  ~~welche~~ welcher die empirische Naturforschung  
 13 (*Physica*) ist.

14 Es ist aber unmöglich aus blos empirischen Erkenntni Begriffen  
 15 ein System zu z̄imern. Es wird jederzeit ein  $\emptyset$  ~~zusam̄en~~ gestoppeltes  $\emptyset$   
 16 zusam̄engestoppeltes

17 Aggregat #bleiben von Beobachtungen dieser oder jener Eigenschaft der  
 18 Materie bleiben was zwar ansehnlich aber doch im̄er nur fragmen-  
 19 tarisch wachsen kañ und \*5 w man still stehen kañ wo man will  
 20 weil es an der Idee fehlt mangelt welche ein innerlich #begründetes  
 21 und zugleich sich selbst #begrenzendes Gantze ausmacht; ein solches aber  
 22 alse kañ nicht anders als nach Begriffen *a priori* zu Stande kom-  
 23 men.

23 Gleichwohl verlangt der philosophische Naturforscher daß dieser Schritt  
 24 nämlich von den metaph. Anf. Gr. d. Nat. W. zur Physik geschehe  
 25 und weñ er auch nur einen kleinen Theil dieses Feldes auszufüllen  
 26 hoffen kañ er doch das atomische  $\emptyset$  ~~Fe~~ Erchwerk der Eintheilung der

Convolut II P\_5

TRANSCRIPTION | FAKSIMILE



```

5 <TEI xmlns="http://www.tei-c.org/ns/1.0">
6 <teiHeader>
7 <fileDesc>
8 <titleStmnt><title>LHM: Lyrik des hohen Mittelalters. Eine elektronische
9 <title>Reinmar</title>
10 <author>Manuel Braun</author><author>Sonja Glauch</author><author>Florian Kragl</author></titleStmnt>
11 <editionStmnt><edition>Version 0.1</edition></editionStmnt><publicationStmnt><p>Beispieldatei im Rahmen der Projektentwicklung.</p><p>29.10.2013</p></publicationStmnt>
12 <sourceDesc><p>Ms#252;nchen UB cod. ms. 2#176; 731</p><p>born digital</p></sourceDesc>
13 </fileDesc>
14 </teiHeader>
15 <text>
16 <body>
17 <div type="strophe" n="Reinm E 364"><div type="strophenlabel">I</div><div type="vers" n="0">
18 <l>
19 <milestone unit="oword" n="0"/><milestone unit="rword" n="0"/><hi rend="initiale size=2 color=rot">M</hi>ir
20 <milestone unit="oword" n="1"/><milestone unit="rword" n="1"/><choice><orig>#x017F;</orig><reg type="n1">s</reg></choice>t
21 <milestone unit="oword" n="2"/><milestone unit="rword" n="2"/><choice><orig>#x017F;</orig><reg type="n1">s</reg></choice>ehen
22 <milestone unit="oword" n="3"/><milestone unit="rword" n="3"/>al
23 <milestone unit="oword" n="4"/><milestone unit="rword" n="4"/>ei
24 <milestone unit="oword" n="5"/><milestone unit="rword" n="5"/>ki
25
26 </l>
27 </div>

```

**Reinm E 365** = MF 145,9

**II**

Minne, die der werelde ir fraude meret,  
 seht, die brahte in traumes wis die frauen min,  
 da min lip an slafe was gekeret  
 unde ersach sich an die besten wunne sin.  
 do sach ich ir liechten tugende, ir werden schin,  
 schon unde auch für alle wip geheret,  
 nūwen daz ein lützel was verseret  
 ir vil frauenrichez mündelin.

1 werelde ] werlede

3 slāfen MF/MT 4 der MF/MT 8 vrōuden rīchez rōtez MF/MT

4 Ob sich an etw. ersehen *Dat. oder Akk. verlangt, ist nach Maßgabe der Wbb. nicht zu entscheiden* (vgl. *Le I, Sp. 57; Le I, Sp. 670; BMZ III/2, S. 276f.; keine weiteren Belege in MHDBDB*).

**Reinm E 364** = MF 145,1



Würzburger Liederhandschrift, Hausbuch des Michael de Leone (München UB cod. ms. 2° 731), fol. 190vb



**I**

Mir ist geschehen als eime kindeline,  
 daz sin schōnez bilde in eime glase besach  
 unde greif dar nach sin selbes schine  
 so lange, unz daz sin hant den spiegel gar zebrach.  
 do wart al sin wunne ein leitlich ungemach.  
 also gedahte ich immer vro ze sine,  
 do ich gesach die lieben frauen mine,  
 von der mir herzeleides vil geschach.

Mor C 100  
Mor C, 39

2 gesach MF/MT 4 sō vil, biz daz ez den MF/MT 6 dāhte MF/MT 8 mir bī liebe leides vil MF/MT

# ¿Qué es lo que llamamos texto?

## Requerimientos funcionales para registros bibliográficos



POSTDATA

Poetry Standardization  
and Linked Open Data



# ¿Qué es digital en la edición?

Requerimientos funcionales para registros bibliográficos



POSTDATA

Poetry Standardization  
and Linked Open Data



# Partir de una expresión. Llegar a una manifestación digital



POSTDATA

Poetry Standardization  
and Linked Open Data

## Digitalización

Paz a España, paz segura.  
Canten abiertos los campos  
de Extremadura.

Lean los niños, las flores,  
y entre las negras encinas  
todos los trabajadores.

¡Lejos tanta noche oscura!  
¡Para siempre en primavera,  
la tierra de Extremadura!

Yo os traigo en este pregón,  
el libro, fuente de gracia,  
que ilumina el corazón.

Agua tranquila, agua pura,  
agua que riegue la sangre  
del alma de Extremadura.

Libros de <sup>amor, luz,</sup> ~~para~~, ensueño.  
Alta vida y viento en calma  
en todo el pueblo extremeño.

## Transcripción

Paz a España, paz segura.  
Canten abiertos los campos  
dichosos de Extremadura. !  
Lean los niños, las flores,  
y entre las negras encinas  
todos los trabajadores. !  
¡Lejos tanta noche oscura!  
¡Para siempre en primavera,  
la tierra de Extremadura! !  
Yo os traigo en este pregón,  
el libro, fuente de gracia,  
que ilumina el corazón. !  
Agua tranquila, agua pura,  
agua que riegue la sangre  
del alma de Extremadura. !  
Libros de amor, luz, ensueño.  
Alta vida y viento en calma  
en todo el pueblo extremeño

## Codificación

```
<poema>
  <estrofa>
    <verso>Paz a <topon>España</topon>, paz segura.</verso>
    <verso>Canten abiertos los campos</verso>
    <verso>dichosos de <topon>Extremadura</topon>.</verso>
  </estrofa>
  <estrofa>
    <verso>Lean los niños, las flores,</verso>
    <verso>y entre las negras encinas</verso>
    <verso>todos los trabajadores.</verso>
  </estrofa>
  <estrofa>
    <verso>¡Lejos tanta noche oscura!</verso>
    <verso>¡Para siempre en primavera,</verso>
    <verso>la tierra de <topon>Extremadura</topon>!</verso>
  </estrofa>
  <estrofa>
    <verso>Yo os traigo en este pregón,</verso>
    <verso>el libro, fuente de gracia,</verso>
    <verso>que ilumina el corazón.</verso>
  </estrofa>
  <estrofa>
    <verso>Agua tranquila, agua pura,</verso>
    <verso>agua que riegue la sangre</verso>
    <verso>del alma de <topon>Extremadura</topon>.</verso>
  </estrofa>
  <estrofa>
    <verso>Libros de
      <sust><borr>luz</borr><adi>amor</adi></sust>,
      <sust><borr>para</borr><adi>luz</adi></sust>,
      ensueño.</verso>
    <verso>Alta vida y viento en calma</verso>
    <verso>en todo el pueblo extremeño</verso>
  </estrofa>
</poema>
```

# ¿Qué es un lenguaje de marcado?

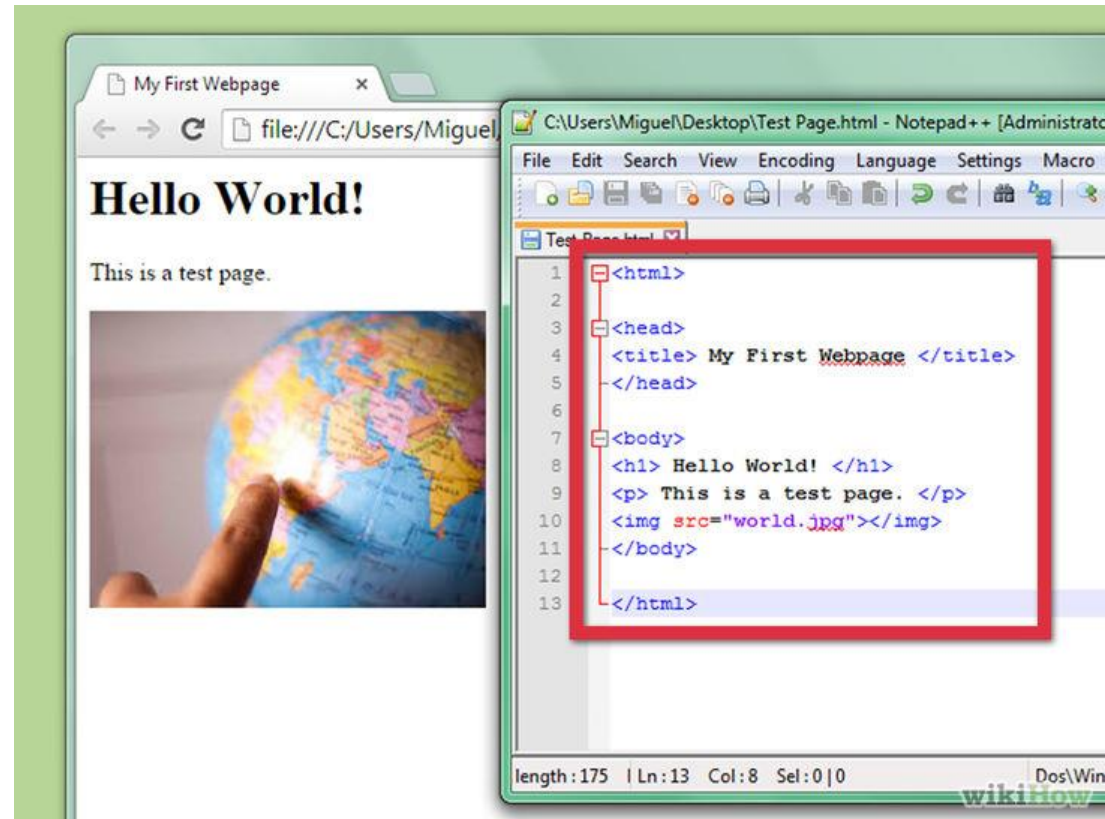


POSTDATA

Poetry Standardization  
and Linked Open Data

Dado que los ordenadores no pueden, todavía, leer el “lenguaje natural”, los lenguajes de marcado permiten a los ordenadores comprender la estructura de un texto, organizarlo, etc.

¿Un ejemplo? ¡HTML!



# ¿Qué es marcar?



POSTDATA

Poetry Standardization  
and Linked Open Data

- Aislar una porción de texto con un significado semántico específico.
- La estructura consiste en la 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.

# Lenguajes de marcado



POSTDATA

Poetry Standardization  
and Linked Open Data

- Se desarrollan para tratar con documentos (información predominantemente textual)
- Describen las partes lógicas de un tipo de documento particular, sin límite en el nivel de profundidad o detalle

```
<lg type="estrofa" subtype="octava">
```

```
<l n="1" rhyme="ista">Vi estar hermosa vista</l>
```

```
<l n="2" rhyme="or">tres colores en una flor,</l>
```

```
<l n="3" rhyme="ista">e avían grand conquista</l>
```

```
<l n="4" rhyme="or">por qual era la mejor,</l>
```

```
<l n="5" rhyme="or">e pedieron judgador</l>
```

```
<l n="6" rhyme="ez">qual levaría el prez,</l>
```

```
<l n="7" rhyme="ez">tomaron por su juez</l>
```

```
<l n="8" rhyme="or">que fuesse don Amor.</l>
```

```
</lg>
```

# Lenguajes de marcado

Son lenguajes porque tienen

- Vocabulario: Para denominar a cada elemento estructural del documento de cualquier nivel de profundidad, a cada atributo de estos elementos, a cada valor de estos atributos (para cierto tipo de atributo), etc.
- Sintaxis (o gramática): Para indicar qué construcciones estructurales están permitidas: qué debe o puede contener cada elemento, si un elemento o atributo es obligatorio, si un elemento puede o no repetirse, qué tipo de valor admite un atributo, etc.



# ¿Qué es XML?

- XML = Extensible Markup Language
- XML es un heredero, más sencillo y eficaz, de su antecesor SGML (Standard Generalized Markup Language)
- Se publicó en febrero de [1998](#) por el W3C Consortium.
- La documentación oficial está disponible en: <http://www.w3.org/XML/>
- Última versión: [XML 1.0 \(5ª edición\)](#)
- Tutoriales en líneas: [Introducción a XML elaborada por TEI](#)

# Estructura de una etiqueta



POSTDATA

Poetry Standardization  
and Linked Open Data

**Elemento:** son las unidades de información semántica compuestas de una marca de apertura **<elemento>** y de cierre **</elemento>**. La etiqueta es el identificador general. Puede contener texto, otros elementos o estar vacío **<elemento/>**.

**Atributo:** añaden una característica adicional al elemento y deben siempre llevar un valor. Pueden ser generales, es decir, utilizables en todos los elementos TEI (**@xml:id**, **@xml:lang**), o específicos de un elemento.

**Valor:** rellena el valor o contenido del atributo.

# Un doc XML



POSTDATA

Poetry Standardization  
and Linked Open Data

Tiene que estar bien formado, es decir estar expresado correctamente según la gramática de XML



# ¿Qué significa bien formado?

1. Tiene que existir un único elemento raíz que contenga todo el documento
2. Cada subárbol debe estar correctamente anidado, es decir, no se pueden solapar
3. Los nombres de las etiquetas son *case/sensitive*:  
`<de> ≠ <De> ≠ <DE>`
4. Una etiqueta de apertura `<p>` debe tener siempre una etiqueta de cierre `</p>` (salvo en los casos de etiquetas vacías `<vacía/>` )
5. No debe haber espacios en blanco tras `<` o antes de `</`  
`<p>párrafo</p>`  
Y no `< p>párrafo</ p>` ni `<p >párrafo</p >`
6. El valor de los atributos ha de estar siempre entrecomillado  
`<p type="text">algo de texto</p>`

Tiene que ser válido

- Para ello, necesita un modelo que fije la estructura básica y las posibilidades de anidación del documento. Las DTD o esquemas (RelaxNG, W3CSchema) establecen la sintaxis que debe tener el documento XML.



# ¿Qué significa válido?

Un documento XML es válido cuando cumple las reglas establecidas en un **esquema externo**.

El esquema especifica:

1. El nombre del elemento raíz
2. El nombre de todos los elementos que se pueden utilizar
3. Los nombres, tipo de datos y (ocasionalmente) los valores por defecto de los atributos
4. Reglas acerca de cómo se pueden anidar los elementos
5. Y otras cosillas dependiendo del lenguaje del esquema

# ¿Para qué un esquema?



POSTDATA

Poetry Standardization  
and Linked Open Data

1. Asegura que en los documentos solo se utilizan los elementos, atributos y entidades predefinidas
2. Establece las reglas estructurales del tipo 'toda receta debe incluir una lista de ingredientes'
3. Asegura que la misma cosa se llame de la misma manera siempre





POSTDATA

Poetry Standardization  
and Linked Open Data

Ahora veamos cómo TEI se expresa a  
través del XML



# Text Encoding Initiative.

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



POSTDATA

Poetry Standardization  
and Linked Open Data



< Text Encoding Initiative >

[Home](#) [Guidelines](#) [Activities](#) [Tools](#) [Membership](#) [Support](#) [About](#) [News](#)

[Home](#)

Entire site

Search

## TEI-C News

[Wiki page for P6 ideas](#)

*Posted on: 2014-10-24*

[Announcing TAPAS launch](#)

*Posted on: 2014-10-24*

[2014 TEI Conference, Evanston, Illinois](#)

*Posted on: 2014-10-23*

[Elections: Deadline VERY soon!](#)

*Posted on: 2014-10-22*

[REMINDER: TEI Board/Council voting deadline is October 24](#)

*Posted on: 2014-10-20*

[TEI Tweet Chat](#)

*Posted on: 2014-10-06*

## Other News

[The iSchool at the University of Illinois is recruiting](#)

*Posted on: 2014-10-29*

[Job: Director of Academic IT Services, University of Oxford](#)

*Posted on: 2014-10-29*

## TEI: Text Encoding Initiative

The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation. In addition to the Guidelines themselves, the Consortium provides a variety of [resources](#) and [training events](#) for learning TEI, information on [projects using the TEI](#), a [bibliography of TEI-related publications](#), and [software](#) developed for or adapted to the TEI.

The TEI Consortium is a nonprofit membership organization composed of academic institutions, research projects, and individual scholars from around the world. Members contribute financially to the Consortium and elect representatives to its Council and Board of Directors.

Want to become active in the TEI community? [Become a TEI Member](#), join a [special interest group](#), sign up for the [TEI-L mailing list](#), and come to our [annual conferences and members' meetings](#).

*Last recorded change to this page: 2013-03-12 • For corrections or updates, contact [web@tei-c.org](mailto:web@tei-c.org)*

# Text Encoding Initiative

- 1987: primera reunión (Vassar College, NY) con el fin de crear unas guías directrices para la codificación de textos electrónicos
- 1988 – 1990: TEI P1
- 1990 – 1992: TEI P2 y creación de grupos de trabajo por •  
1993 – 1994: TEI P3, primera versión oficial de las *Guidelines*
- 2000: creación del Consortium TEI
- 2002: TEI P4, primera versión en XML (de la P3)
- 2007: TEI P5
- 2011-2015: TEI P5 en sus versiones
- 2015: TAPAS project, TEI Simple

- Guidelines
- Lista de proyectos que utilizan TEI
- Una Wiki
- Herramientas
- Lista de discusión

# Text Encoding Initiative

- Iniciativa de codificación de textos en Humanidades, Cs. Sociales y Lingüística.
- Se describe en las TEI *Guidelines: el manual de uso*.
- Se expresa a través del lenguaje XML.
- Tiene un carácter modular y es personalizable
- Es código abierto (GitHub), independiente de cualquier programa o plataforma.

# Algunos problemas



POSTDATA

Poetry Standardization  
and Linked Open Data

- Traducción parcial en español de las *Guidelines*.
- Modelo todavía poco conocido, utilizado sobretudo en países anglosajones, aunque también hay fuertes iniciativas en Francia, Alemania, Holanda, entre otros.
- Se tiende todavía a delegar el trabajo de codificar a los informáticos.
- Poca oferta de formación inicial y avanzada.

# Qué podemos marcar

- Estructura textual (párrafos, listas, tablas...)
- Lagunas, caracteres especiales, saltos de línea, saltos de página, abreviaturas
- Comentarios, glosas
- Errores, cancelaciones
- Lecturas diferentes, adiciones
- Poemas, versos, rimas

Todo eso podemos expresar en XML según las directrices de la Text Encoding Initiative.

# ¿Qué más?



POSTDATA

Poetry Standardization  
and Linked Open Data

- Divisiones estructurales del texto
- Información semántica
- Elementos tipográficos puntuales
- Otra información de interés:
  - personas, lugares, fechas
  - cuestiones gramaticales
  - imágenes, etc.

Todo depende de la información que busquemos recuperar a posteriori y procesar informáticamente.

# Guidelines



POSTDATA

Poetry Standardization  
and Linked Open Data

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 Cs. Sociales (aunque no exclusivamente), y en particular la representación de fuentes primarias para su explotación y análisis. 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/>



- Marco de trabajo flexible basado en módulos  
(un total de 21; 4 obligatorios)
- Una selección de más de 500 etiquetas
- Una estructura en clases de modelo y atributos
- Módulos: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>

# 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)

# Recordemos. Tipos de elementos

**Un elemento puede contener**, según se especifique en el esquema :

**1. Otros elementos**, subelementos o elementos hijos .

Ej.: `<text><body><lg></l>`

**2. Texto.**

Ej.: `<author>Prescod, Paul</author>`

**3. Contenido mixto**, o sea, una mezcla de elementos hijos y texto.

Ej.: `<title>Charles F. Goldfarb's<enfasis>XML</enfasis>  
Handbook</title>`

**4. Nada**, si es un elemento vacío. Ej.: `<Cartel  
fuente="imagen1.jpg"/>`

# Tipos de elementos



POSTDATA

Poetry Standardization  
and Linked Open Data

Un elemento puede ser, según se especifique en el esquema,

- 1. Obligatorio**
- 2. Opcional**
- 3. Prohibido**



**MANOS A LA OBRA**



# Crear un XML-TEI Lite 5 y asociar esquema



POSTDATA

Poetry Standardization  
and Linked Open Data

File Tools Document Window Help



Untitled1.xml x

Associate Schema...

```
1 <?xml version="1.0" encoding="UTF-8"?>
2
3
```

# Semáforo verde

- Para crear un XML válido tengo que asociarle un esquema también válido. Solo así obtendré un documento bien formado

(LUZ ROJA vs. LUZ VERDE)



**EJEMPLO**



[ID0151] PN9-22a (119v) (4, 8)

## **Del Cardinal de mendoçe**

Dama mi gran querer  
En tanto grado me toca  
Que no me puedo valer  
Mi biuir por vos se apoca

Apocafe mi biuir  
Por amar demefiado  
No me aprouecha el feruir  
Ne me aprouecha el cuydado  
Voime del todo a perder  
La vida mia se apoca  
Esto causa mi querer  
Que in tanto grado me tocca

# Metadatos: <teiHeader>



POSTDATA

Poetry Standardization  
and Linked Open Data

- Los metadatos son datos sobre datos: título, fuente original, etc.
- Los metadatos son los datos que permiten identificar, describir, analizar y localizar los recursos
- Datos sobre la historia de revisión del archivo

# Ejemplo <teiHeader>



POSTDATA

Poetry Standardization  
and Linked Open Data

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3   xsi:schemaLocation="http://www.tei-c.org/ns/1.0 metrica%20esquema%20jul14.xsd"
4   xmlns="http://www.tei-c.org/ns/1.0">
5   <teiHeader>
6     <fileDesc>
7       <titleStmt>
8         <title>Arbolito naranjado</title>
9       </titleStmt>
10      <publicationStmt>
11        <publisher>Clara</publisher>
12      </publicationStmt>
13      <sourceDesc>
14        <bibl>
15          <title>Cancionero popular de Jujuy</title>
16        </bibl>
17      </sourceDesc>
18    </fileDesc>
19  </teiHeader>
```



Crear un <teiHeader> con ID027XDN

# EJERCICIO 1

# Ejercicio 1 <teiHeader>

- Generar un doc xml y asociarle el esquema **myTei.rnc**
- Editar poema ID027XDN usando las etiquetas necesarias

# Abrir ID015XGF



POSTDATA

Poetry Standardization  
and Linked Open Data

ID Dutton\_ID ID0151  
MS Manuscript mp4a  
SER Poem\_serial 4  
FOLIO Folio 5<v>-6<r>  
STSIG Stanza\_signature 4,8  
SYLL Syllable\_signature 1x4@8,1x8@8  
FIRST First\_line dama mi gran querer  
SECND Second\_line en tanto grado me toca  
LAST Last\_line quen tanto grado me toca.

Todos los datos de Repertorio métrico de la poesía cancioneril del siglo XV, Alcalá de Henares, Universidad de Alcalá de Henares, 1999, de Ana María Gómez Bravo

<http://poemetca.lnhd.es/>

Presentación crítica del texto sobre MP4: Gimena del Rio Riande

# Completamos allí la información

Nombre	Ejemplo	Descripción
ID Dutton_ID	ID0151	Identificador en Brian Dutton en su <i>Catálogo-Índice de la Poesía Cancioneril del Siglo XV</i>
MS Manuscript	mp4a	Identificador del manuscrito
SER Poem_serial	4	Posición dentro del manuscrito
FOLIO Folio	5<v>-6<r>	Folios dentro del manuscrito
STSIG Stanza_signature	4,8	Estructura de cada estrofa: número de versos por cada estrofa. En este caso 4 versos la primera estrofa, 8 versos la segunda.
SYLL Syllable_signature	1x4@8,1x8@8	Estructura del poema, definiendo qué tipo de versos (versos de cuántas sílabas) forman cada estrofa o tirada de versos. En este caso, una estrofa de 4 versos octosílabos y una estrofa de 8 versos octosílabos
FIRST First_line	dama mi gran querer	Íncipit o primer verso
SECND Second_line	en tanto grado me toca	Segundo verso
LAST Last_line	quen tanto grado me toca	Último verso

```
26     </bibl>
27 ▾ <bibl n="0151">
28     <idno>texid Dutton 0151</idno>
29     <title>Dutton</title>
30 </bibl>
31 ▾ <bibl>
32     <title>Repertorio métrico de la poesía cancioneril del siglo XV</title>
33     <author>Ana María Gómez Bravo</author>
34     <date>1999</date>
35     <pubPlace>Alcalá de Henares</pubPlace>
36     <publisher>Universidad de Alcalá de Henares</publisher>
37 </bibl>
38 ▾ <bibl>
39     <edition>Edición crítica Gimena del Río</edition>
40 </bibl>
41 </listBibl>
42 </sourceDesc>
43 </fileDesc>
44 </teiHeader>
```





<text> y <body>

**EJEMPLO**

# <text> <body> módulo verso

```
<text>  
  <body>  
    <lg>  
      <l></l>  
    </lg>  
  </body>  
</text>  
</TEI>
```

```
<lg type="soneto">  
  <lg type="cuarteto" n="01" rhyme="abba">  
    <1>Escrito está en mi alma vuestro gesto, </1>  
    <1>y cuanto yo escribir de vos deseo; </1>  
    <1>vos sola lo escribisteis, yo lo leo </1>  
    <1>tan solo, que aun de vos me guardo en esto.</1>  
</lg>
```

# Ficha ID015XGF



POSTDATA

Poetry Standardization  
and Linked Open Data

```
45 <text>
46   <body>
47     <lg type="poema">
48       <lg type="copla" ana="xyxy">
49         <l>Dama, mi gran querer</l>
50         <l>en tanto grado me toca</l>
51         <l>que no me puedo valer,</l>
52         <l>mi vevir por vos se apoca.</l>
53       </lg>
```

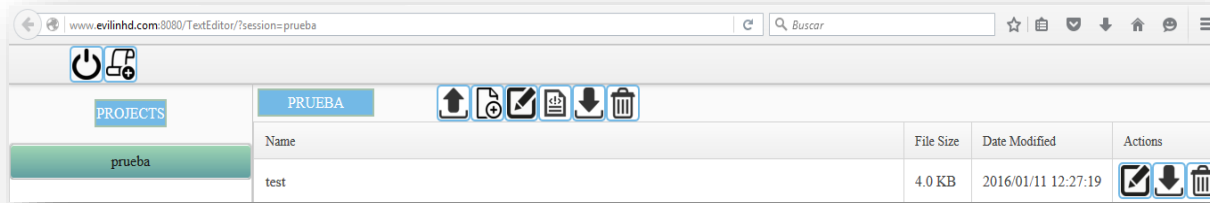


Crear un <text> con ID027XDN

## **EJERCICIO 2**






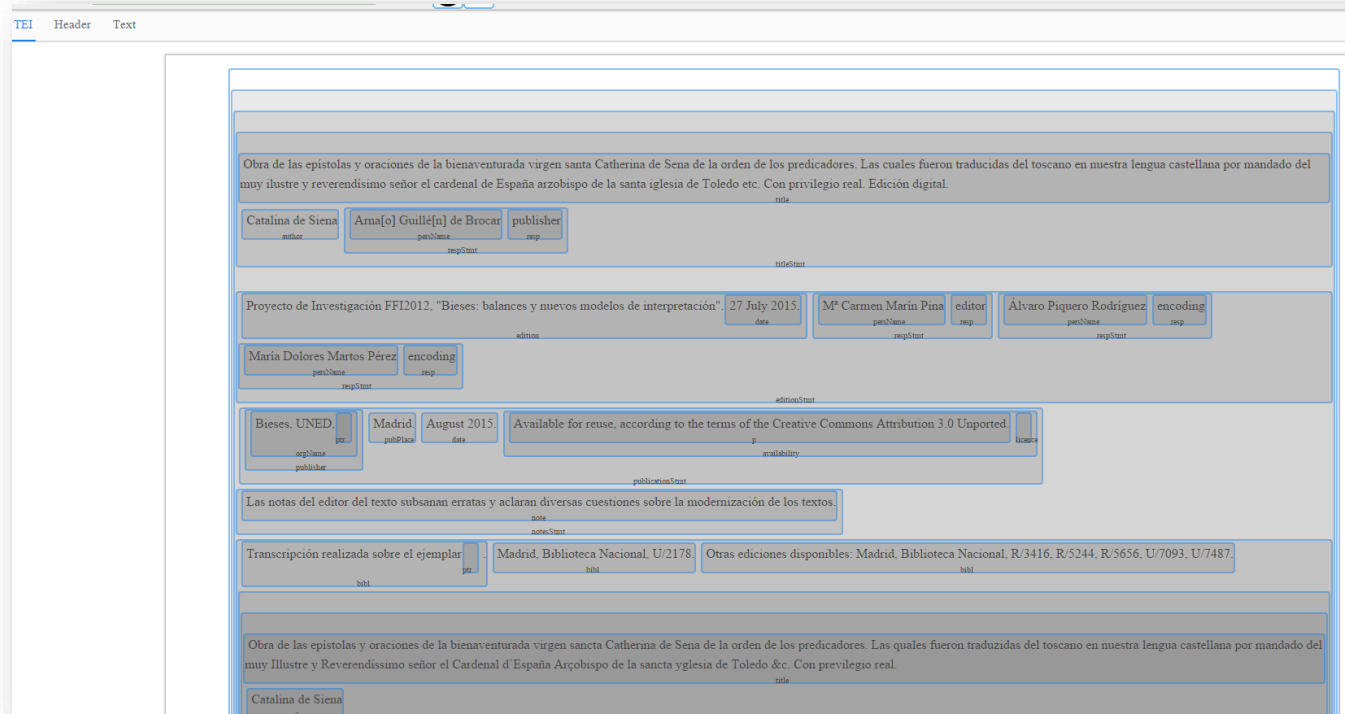
**EVI-LINHD**



www.evilnhd.com:8080/TextEditor/?session=prueba

PRUEBA

Name	File Size	Date Modified	Actions
test	4.0 KB	2016/01/11 12:27:19	  



TEI Header Text

Obra de las epistolas y oraciones de la bienaventurada virgen santa Catherina de Sena de la orden de los predicadores. Las cuales fueron traducidas del toscano en nuestra lengua castellana por mandado del muy ilustre y reverendissimo señor el cardenal de España arzobispo de la santa iglesia de Toledo etc. Con privilegio real. Edición digital.

Catalina de Siena **Arna[o] Guillé[n] de Brocar** publisher

Proyecto de Investigación FFI2012. "Bieses: balances y nuevos modelos de interpretación". 27 July 2015. M<sup>a</sup> Carmen Marin Pina editor Álvaro Piquero Rodriguez encoding

Maria Dolores Martos Pérez encoding

Bieses, UNED. Madrid, August 2015. Available for reuse, according to the terms of the Creative Commons Attribution 3.0 Unported.

Las notas del editor del texto subsanan erratas y aclaran diversas cuestiones sobre la modernización de los textos.

Transcripción realizada sobre el ejemplar Madrid, Biblioteca Nacional, U/2178. Otras ediciones disponibles: Madrid, Biblioteca Nacional, R/3416, R/5244, R/5656, U/7093, U/7487.

Obra de las epistolas y oraciones de la bienaventurada virgen sancta Catherina de Sena de la orden de los predicadores. Las cuales fueron traduzidas del toscano en nuestra lengua castellana por mandado del muy Ilustre y Reverendissimo señor el Cardenal d'España Arçobispo de la sancta yglesia de Toledo &c. Con privilegio real.

Catalina de Siena

# Textos taller



COLECCIONES

textos\_taller

TEXTOS\_TALLER



Nombre	Tamaño	Fecha de modificación	Acciones
poema1	4,0 KB	2016/11/04 14:15:38	   
ejemplo1	4,0 KB	2016/10/29 10:54:20	   
damademigran.xml	4,0 KB	2016/11/04 14:17:07	   

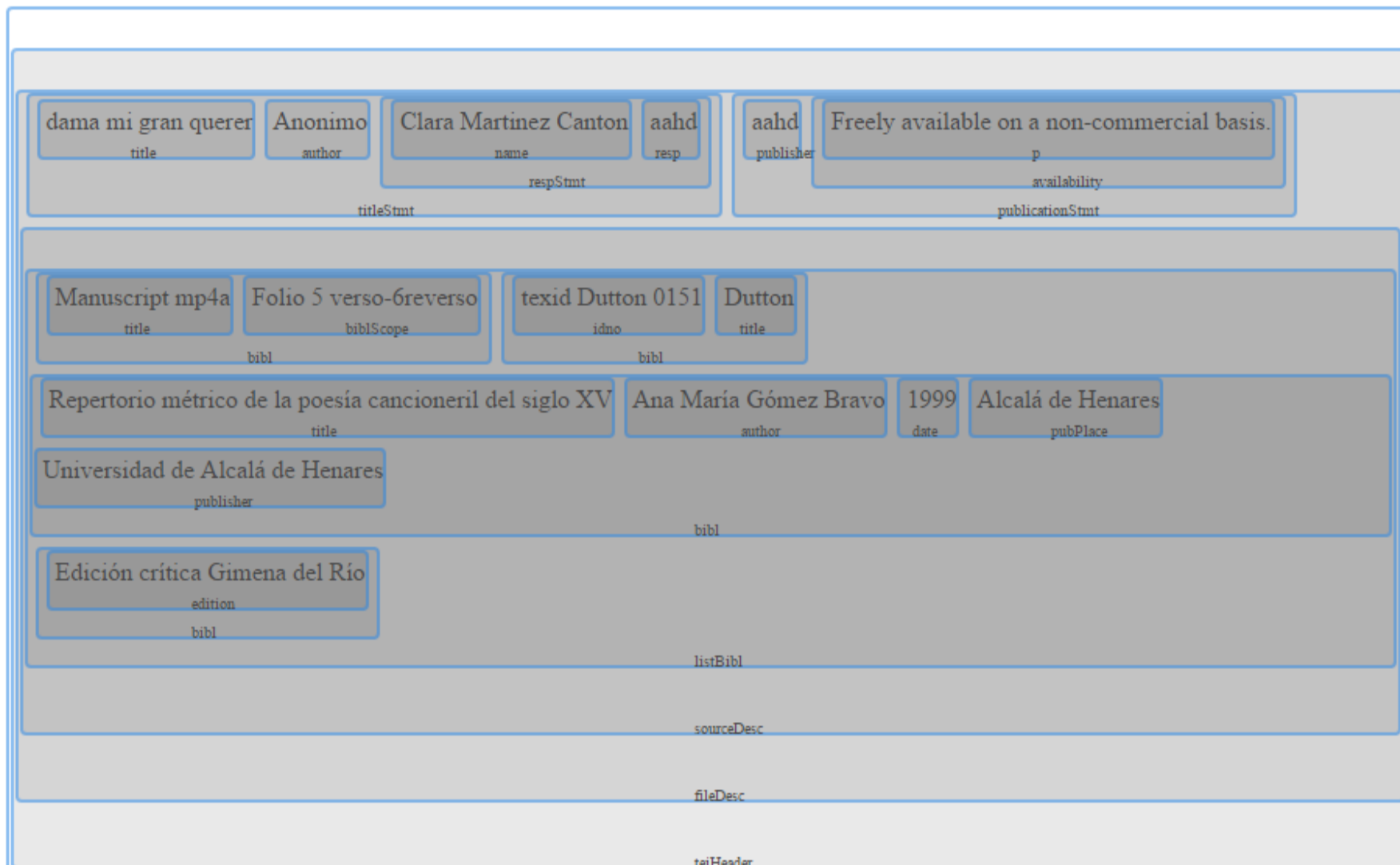


# Modificar



POSTDATA

Poetry Standardization  
and Linked Open Data



# Visualizar



POSTDATA

Poetry Standardization  
and Linked Open Data

TEXTOS\_TALLER



Nombre	Tamaño	Fecha de modificación	Acciones
poema1	4,0 KB	2016/11/04 14:15:38	
ejemplo1	4,0 KB	2016/10/29 10:54:20	
damademigran.xml	4,0 KB	2016/11/04 14:17:07	

Exportar fichero a TeiSimple



---

Dama, mi gran querer  
en tanto grado me toca  
que no me puedo valer,  
mi vevir por vos se apoca.  
Apócase el mi vevir  
por amar demasiado,  
no m' aprovecha el servir  
ni m' aprovecha el cuidado,  
voyme del todo a perder,  
la vida mía se apoca,  
cáusalo mucho querer  
qu' en tanto grado me toca.



POSTDATA

Poetry Standardization  
and Linked Open Data

**Elena González-Blanco** [egonzalezblanco@flog.uned.es](mailto:egonzalezblanco@flog.uned.es)

**Gimena del Rio Riande** [gdelrio.riande@gmail.com](mailto:gdelrio.riande@gmail.com)

**Clara I. Martínez Cantón** [cimartinez@flog.uned.es](mailto:cimartinez@flog.uned.es)



**& Equipo LINHD**

<http://linhd.uned.es>

<http://postdata.linhd.es>

@linhduned



**European Research Council**

Established by the European Commission

**Supporting top researchers  
from anywhere in the world**