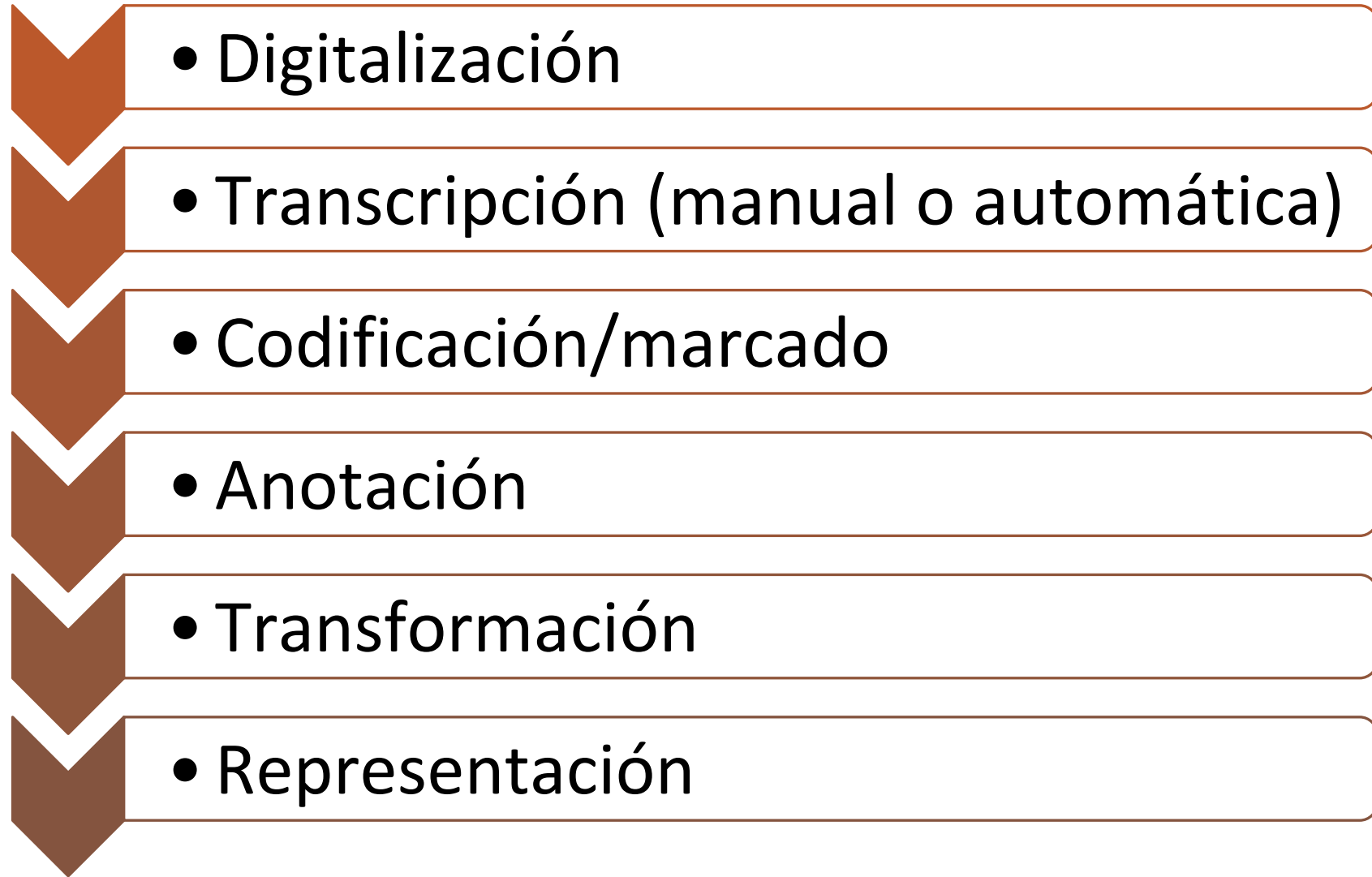


Taller Hackathon: Edición digital colaborativa con TEI del Martín Fierro

DRA. GIMENA DEL RIO RIANDE (CONSEJO NACIONAL DE INVESTIGACIONES CIENTÍFICAS, ARGENTINA) – DRA. SUSANNA ALLÉS TORRENT (UNIVERSITY OF MIAMI)

Las tareas

- 
- Digitalización
 - Transcripción (manual o automática)
 - Codificación/marcado
 - Anotación
 - Transformación
 - Representación

Los componentes

1. Datos
2. Interfaces
3. Scripts



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

Ediciones con TEI

American memory: <http://memory.loc.gov/ammem/index.html>

Biblioteca Digital del Pensamiento Novohispano: <http://www.bdpn.unam.mx/>

Vincent Van Gogh: <http://vangoghletters.org>

Voices of the Holocaust: <http://voices.iit.edu/>

Manuscritos de Montesquieu:

<https://www.unicaen.fr/services/puc/sources/Montesquieu/>

Diálogo Medieval:

http://sade.textgrid.de/exist/apps/SADE/Dialogo_Medieval/index.html

The World of Dante: <http://www.worldofdante.org/>



THE WORLD OF DANTE

INFERNO * PURGATORIO * PARADISO

[Gallery](#) / [Maps](#) / [Music](#) / [Timeline](#) / [Search](#) / [Teacher Resources](#) / [About](#)

PURGATORIO CANTO 2

Già era 'l sole a l'orizzonte giunto
lo cui merid'ian cerchio coverchia
Ierusalèm col suo più alto punto;

e la notte, che opposita a lui cerchia,
uscita di Gange fuor con le Bilance,
che le caggion di man quando soverchia;

sì che le bianche e le vermiglie guance,
là dov' i' era, de la bella Aurora
per troppa etate divenivan rance.

Noi eravam lunghezzo mare ancora,
come gente che pensa a suo cammino,
che va col cuore e col corpo dimora.

Ed ecco, qual, sorpreso dal mattino,
per li grossi vapor Marte rosseggia
giù nel ponente sovra 'l suol marino,

cotal m'apparve, s'io ancor lo veggia,
un lume per lo mar venir sì ratto,
che 'l muover suo nessun volar pareggia.

Dal qual com' io un poco ebbi ritratto

2.1

2.4

2.7

2.10

2.13

2.16

2.19

PURGATORIO CANTO 2

By now the sun was crossing the horizon
of the meridian whose highest point
covers Jerusalem; and from the Ganges,

night, circling opposite the sun, was moving
together with the Scales that, when the length
of dark defeats the day, desert night's hands;

so that, above the shore that I had reached,
the fair Aurora's white and scarlet cheeks
were, as Aurora aged, becoming orange.

We still were by the sea, like those who think
about the journey they will undertake,
who go in heart but in the body stay.

And just as Mars, when it is overcome
by the invading mists of dawn, glows red
above the waters' plain, low in the west,

so there appeared to me—and may I see it
again—a light that crossed the sea: so swift,
there is no flight of bird to equal it.

When, for a moment, I'd withdrawn my eyes

Purg. 2

« [Prev](#) [Next](#) »

- **People:** (PE)

2.61 Virgilio

2.76 Casella

2.91 Casella

2.119 Catone

+ **Places:** (PL)

+ **Creatures:** (C)

+ **Deities:** (D)

- **Structures:** (S)

+ **Images:** (I)

+ **Music:** (M)

EL MERCADO DE TEXTOS EN TEI



¿Qué son las tecnologías de marcado?

Digitalización

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,} ~~amor, luz,~~ 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>,
    <verso>ensueño.</verso>
    <verso>Alta vida y viento en calma</verso>
    <verso>en todo el pueblo extremeño.</verso>
  </estrofa>
</poema>
```


Lenguajes de marcado

Describen las partes lógicas (estructura lógica y semántica) de un tipo de documento (o conjunto de datos) particular, sin límite en el nivel de profundidad o detalle

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.

```
<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>
```



-
- **Text Encoding Initiative:** Iniciativa de codificación de textos, especialmente en Humanidades, Ciencias Sociales y Lingüística.
 - Mantenido hoy por el **Consortio TEI** "*organización internacional cuya misión es el desarrollo y mantenimiento de directrices para la codificación digital de textos literarios y lingüísticos que publica las TEI Guidelines for Electronic Text Encoding and Interchange, un estándar internacional e interdisciplinar ampliamente usado por bibliotecas, museos, editores y eruditos para representar todo tipo de material textual para la investigación y la enseñanza en línea*"
 - Se describe en las **TEI Guidelines**: el manual de uso
 - Tiene un **carácter modular** y es **personalizable**

Objetivos

- Promover la creación, el intercambio y la integración de los datos textuales informatizados.
- Editar textos de cualquier tipología, género y disciplina (Ciencias Humanas y Sociales), en cualquier lengua y de cualquier período cronológico.

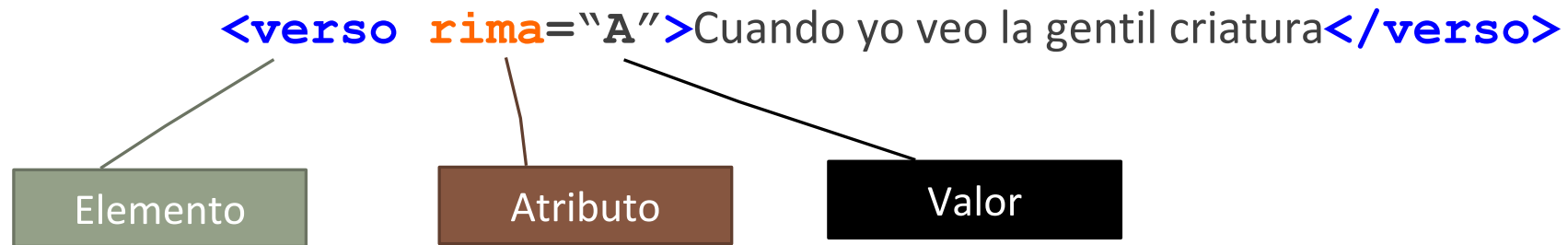
TEI Consortium: <http://www.tei-c.org>

- [Guidelines](#)
- [Lista de proyectos](#) que utilizan TEI
- [Wiki](#)
- [Herramientas](#)
- [Lista de discusión](#)
- Curso de TEI en español: <https://tthub.io/materiales/introduccion-a-tei/>

Módulos de TEI

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`

Tipos de etiquetas: elementos, atributos y valores



Elemento: son las unidades de información semántica compuestas de una marca de apertura `<verso>` y de cierre `</verso>`. El nombre del elemento constituye su identificador general.

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

Valor: rellena el valor o contenido del atributo.

Un doc XML-TEI

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



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 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>`

Esquema

- Todo documento TEI necesita un esquema que establezca una sintaxis precisa. TEI ofrece algunos modelos, como una versión light (`teilight.rng`) o una versión con todos los elementos (`tei_all.rng`) (que es poco recomendable).
- La buena práctica sugiere que cada proyecto debe crear a su medida un esquema, adoptando los módulos que crea conveniente, eliminando y añadiendo elementos, eliminando y concretando atributos y valores. Así, la labor de marcado será mucho más fácil y consistente.

¿Para qué un esquema?

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

ROMA

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

Gracias a esta aplicación, podremos crear el fichero ODD = One Document Does it All, que será esencial para nuestro trabajo.

Sería algo así como... :

- Roma: “el supermercado”
- Esquema: “los ingredientes”
- ODD: “la receta”

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

The screenshot shows the TEI Roma web interface. At the top, there is a navigation bar with buttons for 'New', 'Customize', 'Language', 'Modules', 'Add Elements', 'Change Classes', 'Schema', 'Documentation', 'Save Customization', and 'Sanity Checker'. The 'Modules' button is currently selected. Below the navigation bar, there are two main panels. The left panel is titled 'List of TEI Modules' and contains a table with columns for 'Module name', 'A short description', and 'Changes'. The right panel is titled 'List of selected Modules' and shows a list of modules that have been selected for customization, with 'remove' links next to each.

TEI Roma: generating validators for the TEI

You are currently working on **My TEI Extension**

Modules

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#) [Sanity Checker](#)

List of TEI Modules		
Module name	A short description	Changes
add analysis	Simple analytic mechanisms	
add certainty	Certainty and uncertainty	
add core	Elements common to all TEI documents	
add corpus	Corpus texts	
add dictionaries	Dictionaries	
add drama	Performance texts	
add figures	Tables, formulae, notated music, and figures	
add galle	Character and glyph documentation	
add header	The TEI Header	
add iso-fs	Feature structures	
add linking	Linking, segmentation and alignment	
add msdescription	Manuscript Description	
add namesdates	Names and dates	
add nets	Graphs, networks, and trees	
add spoken	Transcribed Speech	
add tagdocs	Documentation of TEI modules	
add textcrit	Critical Apparatus	
add textstructure	Default text structure	
add transcr	Transcription of primary sources	
add verse	Verse structures	

List of selected Modules

[remove core](#)

[remove tel](#)

[remove header](#)

[remove textstructure](#)

Más información en: <https://tthub.io/materiales/introduccion-a-tei/u7/4-aplicacion-roma/>


Estructura de un documento TEI

<TEI>: Elemento raíz habitual. Tiene dos hijos:

1. <teiHeader> Metadatos sobre el texto codificado (autoría y responsabilidad, información bibliográfica, descripción del manuscrito, historial de revisiones, etc.)
2. <text> Características estructurales del documento (secciones, encabezados, párrafos, citas, resaltado, etc.)

Estructura de un documento TEI

```
<TEI>
  <teiHeader>
    <!-- obligatorio -->
  </teiHeader>
  <text>
    <front>
      <!-- opcional -->
    </front>
    <group>
      <text>
        <body>
          <!-- obligatorio -->
        </body>
      </text>
      <text>
        <body>
          <!-- obligatorio -->
        </body>
      </text>
      <text>
        <body>
          <!-- obligatorio -->
        </body>
      </text>
    </group>
    <back>
      <!-- opcional -->
    </back>
  </text>
</TEI>
```



A vertical red double-headed arrow is positioned to the left of the <group> element in the XML code. The word "REFERIBLE" is written vertically along the arrow, indicating that the <group> element is a referable unit within the document structure.

TEI Header

`<teiHeader>`: Tiene cuatro hijos y numerosos descendientes

`<fileDesc>` (**obligatorio**): descripción bibliográfica del archivo electrónico

`<titleStmt>` (**obligatorio**): título de la obra y responsables de su contenido intelectual

`<editionStmt>`: edición de un texto

`<extent>`: tamaño del texto electrónico, en unidades adecuadas

`<publicationStmt>` (**obligatorio**): publicación y distribución del texto electrónico

`<seriesStmt>`: serie, en su caso, a la que pertenece la publicación

`<notesStmt>`: notas que informan sobre un texto adicional al indicado en otras partes de la descripción bibliográfica

`<sourceDesc>` (**obligatorio**): descripción bibliográfica del o los textos fuente a partir de los cuales se ha generado el texto electrónico.

`<encodingDesc>`: documenta la relación entre un texto electrónico y la fuente original, y las prácticas de codificación y transcripción seguidas

`<profileDesc>`: descripción de aspectos no bibliográficos de un texto (idiomas usados, materia, género, etc.)

`<revisionDesc>`: historia de las revisiones, control de cambios

TEI Text

`<text>`: texto de la obra, individual (una única obra) o compuesto (una colección de obras, p.ej. una antología). Tiene tres hijos:

`<front>` (opcional): materia inicial (portada, dedicatoria, tabla de contenidos, prólogo, prefacio, agradecimientos, resumen, etc.)

`<body>`: **cuerpo de la obra**.

`<back>` (opcional): materia final (apéndice, glosario, notas, bibliografía, índice, colofón, etc.). En realidad se distingue de la inicial sólo por su situación; pueden contener exactamente los mismos tipos de componentes)

Ejercicio 1. Esquema en Roma



Oxygen.exe

-----START-LICENSE-KEY-----

Registration_Name=DH course

Company=DH course

Category=Academic-Classroom

Component=XML-Editor, XSLT-Debugger, Saxon-SA

Version=19

Number_of_Licenses=2

Date=01-18-2018

Subscription=366

SGN=MC0CFFb6+Ejx0gKXvWV/QhKRyczD6B/6AhUAK5zH1QVW+UqbqAxYiDUfibF6HFI\=

-----END-LICENSE-KEY-----

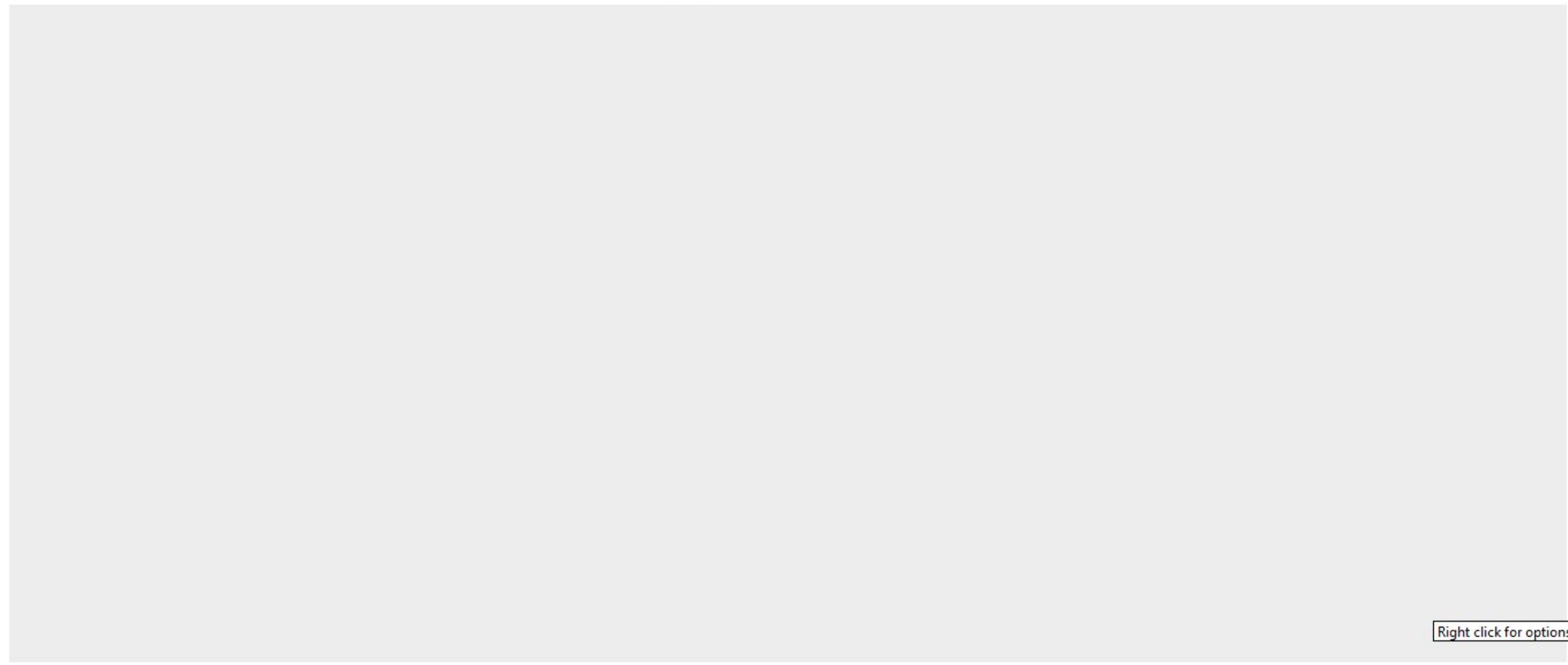
Ejercicio 2. Crear xml doc y asociar esquema



- 1. ABRIR OXYGEN EDITOR
- 2. Click en File > **New** o directamente en el folio en blanco del margen superior izquierdo

Cómo seguimos...

- 3. New XML document
- 4. CUSTOMIZE!
- 5. Click en la carpeta amarilla: Browse for local files
- 6. Abrir ahí el esquema relaxng



Right click for options



XPath 2.0 Execute XPath on 'Current File'



New

Search or select a document type

Recently used

- JTEI Article [TEI JTEI]
- haiku [Haiku]
- postcard_basic [Postcard Basic]

New Document

Global templates

Framework templates

- FO
- Haiku
- DITA
- DITA Map
- DITaval
- Daisy
- DocBook 4

Name: JTEI Article
Extension: xml
Category: TEI JTEI

< Back Customize > Create Cancel

CLICK en NEW DOCUMENT



New

Search or select a document type

Type filter text

- Recently used
 - XML Document
 - RELAX NG Schema - XML
 - jTEI Article [TEI JTEI]
 - haiku [Haiku]
 - postcard_basic [Postcard Basic]
- New Document
 - XML Document
 - XSLT Stylesheet
 - Ant Build
 - XML Schema
 - Document Type Definition
 - RELAX NG Schema - XML
 - RELAX NG Schema - Compact

Name: XML Document
Extension: xml

< Back Customize > Create Cancel



CLICK en CUSTOMIZE





XPath 2.0 Execute XPath on 'Current File'



New

Search or select a document type

Type filter text

- Recently used
 - XML Document
 - RELAX NG Schema - XML
 - jTEI Article [TEI JTEI]
 - haiku [Haiku]
 - postcard_basic [Postcard Basic]
- New Document
 - XML Document
 - XSLT Stylesheet
 - Ant Build
 - XML Schema
 - Document Type Definition
 - RELAX NG Schema - XML
 - RELAX NG Schema - Compact

Name: XML Document
Extension: xml

< Back Customize > Cancel





New

Customize editor

Name: XML Document

Schema URL: file:/C:/Users/Elena/Desktop/haiku/schemas/haiku.rng

Schema type: RelaxNG XML syntax

Public ID:

Namespace: http://www.tei-c.org/ns/1.0

Prefix:

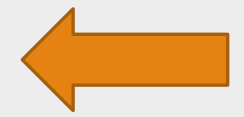
Root Element: TEI

Description
(TEI document) contains a single TEI-conformant document, combining a single TEI header with one or more members of the model.resourceLike class. Multiple TEI elements may be combined to form a teiCorpus element. [4. Default Text Structure 15.1. Varieties of Composite Text]

Add optional content
 Add first Choice particle

< Back Customize > Create Cancel

- Browse for local file
- Browse for remote file
- Browse for archived file
- Browse Data Source Explorer
- Search for file



Ir al archivo .rng y CLICK en **CREATE**





Untitled6.xml x

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="file:/F:/Curso%20Cuba/Taller%20feria/myTEI.rnc" type="application/relax-ng-compact-syntax"?>
3 <TEI xmlns="http://www.tei-c.org/ns/1.0">
4   <teiHeader>
5     <fileDesc>
6       <titleStmt>
7         <title></title>
8       </titleStmt>
9       <publicationStmt></publicationStmt>
10      <sourceDesc></sourceDesc>
11    </fileDesc>
12  </teiHeader>
13  <text></text>
14 </TEI>

```

Right click for options

E [Jing] element "publicationStmt" incomplete; expected element "authority", "distributor", "p" or "publisher"

Text Grid Author



Etiquetas a utilizar

<teiHeader>

<fileDesc>

<titleStmt>

<author>

<editor>

<publicationStmt>

<publisher>

<sourceDesc>

<bibl>

<idno>

<body>

<div>

<lg>

<l>

<choice>

<orig>

<corr>

@xml:id

@type

Codificar este texto

Capítulo I - Cantor y Gaucho

1

Aquí me pongo a cantar

Al compás de la vigüela,

Que el hombre que lo desvela

Una pena extraordinaria

Como la ave solitaria

Con el cantar se consuela.

2

Pido a los Santos del Cielo

Que ayuden mi pensamiento;

Les pido en este momento

Que voy a cantar mi historia

Me refresquen la memoria

Y aclaren mi entendimiento.

Transformación de XML a XHTML

Ahora vamos a hacer una transformación simple (puede ser, por ejemplo, a XHTML o epub)



XPath 2.0 Execute XPath on 'Current File'

Untitled7.xml x joycepostcard.xml x

```

TEI
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0" rend="full" version="5.0">
3 <teiHeader>
4 <fileDesc>
5 <titleStmt>
6 <title>Postcard by James Joyce</title>
7 <author>James Joyce</author>
8 <respStmt>
9 <resp></resp>
10 <name>Gimena</name>
11 </respStmt>
12 </titleStmt>
13 <publicationStmt>
14 <p>[Publication info]</p>
15 </publicationStmt>
16 <sourceDesc>
17 <p>
18 </p>
19 </sourceDesc>
20 </fileDesc>
21 <profileDesc>
22 <correspDesc>
23 <correspAction type="sent">
24 <date when="1908-01-04"></date>
25 <name>James Joyce</name>
26 </correspAction>
27 <correspAction type="received">
28 <name>Elkin Mathews</name>
29 </correspAction>
    
```

Text Grid Author

Configure Transformation Scenario(s)

Type filter text

Association	Scenario	Type
TEI P5 (5)		
<input checked="" type="checkbox"/>	TEI P5 XHTML	XML with XSLT
<input type="checkbox"/>	TEI P5 PDF	XML with XSLT
<input type="checkbox"/>	TEI P5 EPUB	ANT
<input type="checkbox"/>	TEI P5 DOCX	ANT
<input type="checkbox"/>	TEI P5 ODT	ANT

Association follows selection

New Edit Duplicate Remove

Associated scenarios

- TEI P5 XHTML X

Save and close Apply associated (1) Cancel



Tutoriales

[Learn TEI](#)

[TEI WIKI](#)

[TEI by Example](#)

TTHUB. Text Technologies Hub. Tecnologías del texto y edición digital. “Lecciones”.
<https://tthub.io/aprende/>

Algunas herramientas

Lista de herramientas:

Interedtion Wiki. Existing Tools

[TEI Tools](#)

Herramientas de transformación:

[Oxgarage](#)

Visualización y plataformas de publicación:

[TEI Viewer](#)

[TEI Boilerplate](#)

[V-Machine](#)

Algunas herramientas de edición

TEI Boilerplate: <http://dcl.slis.indiana.edu/teibp/index.html>

Collatex <http://collatex.net/>

Juxta Commons <http://juxtacommons.org/>

Classical Text Editor <http://cte.oeaw.ac.at/>

Edictor <http://humanidadesdigitais.org/edictor/>

Versioning Machine 4.0 <http://v-machine.org/>

Otras herramientas

LaTeX: <http://www.latex-project.org/>

Scripto: <http://scripto.org/>

Textual Communities:

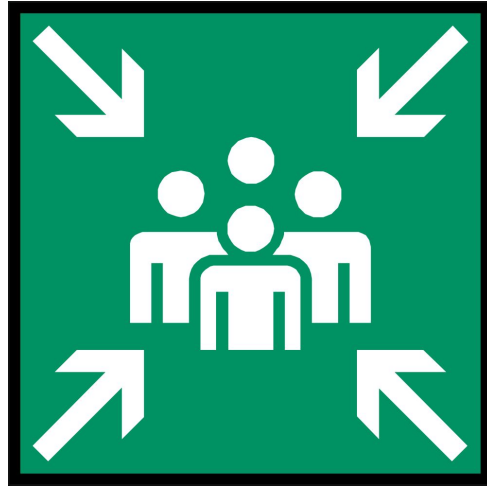
<http://www.textualcommunities.usask.ca/es/web/textual-community>

Open Annotation: <http://www.openannotation.org/>

Shared Canvas, <http://www.shared-canvas.org/>

Philologic: <https://sites.google.com/site/philologic3/home>

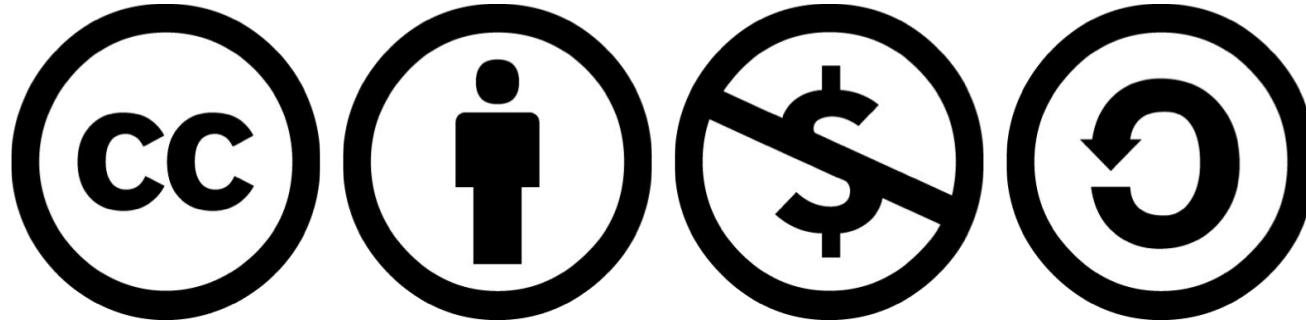
From the page: <http://beta.fromthepage.com/>



Gracias!

Gimena del Rio Riando @gimenadelr

Susanna Allés Torrent @salleto



Compartido con una Licencia Creative Commons 4.0 Internacional.
Compartir igual^[1].

^[1] Atribución — Usted debe darle crédito a esta obra de manera adecuada, proporcionando un enlace a la licencia, e indicando si se han realizado cambios. En 4.0, debe indicar si ha modificado el material y mantener una indicación de las modificaciones anteriores. En 3.0 y versiones anteriores a la licencia, la indicación de los cambios sólo se requiere si se crea una obra derivada.

. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo del licenciante.

NoComercial — Usted no puede hacer uso del material con fines comerciales.

<https://creativecommons.org/>