

# SO I WANT TO USE THE TEI... NOW WHAT?

---

Joey Takeda  
October 25<sup>th</sup>, 2018  
UBC DH Pixelating Mixer



HI!

Joey Takeda, MA Student, Dept. of English Language and Literatures (Science and Technology Studies Program)

Undergrad in English Honours and Gender Studies @ Uvic

Digital Humanities programmer for projects such as *The Map of Early Modern London*, *Linked Early Modern Drama Online*, and *The Endings Project*

Specialize in TEI, XHTML, and Digital Humanities project development

<https://github.com/joeytakeda>

This workshop: <https://github.com/joeytakeda/teiworkshop>

# OVERVIEW

1

(Very quick)  
recap of XML  
and encoding

2

Introduction to  
our primary  
source

3

Introduction to  
the TEI  
Guidelines

4

Introduction to  
XML tools

# WEBSITE

<https://joeytakeda.github.io/teiworkshop>



# XML EXPLAINED

# XML EXPLAINED

The two pointy brackets is called an **element**

- E.g. <book> would be called the book element

# XML EXPLAINED

The two pointy brackets is called an **element**

- E.g. <book> would be called the book element

All elements have **start** and **end tags**

- E.g. <book> is the start tag and </book> is the end tag

# XML EXPLAINED

The two pointy brackets is called an **element**

- E.g. <book> would be called the book element

All elements have **start** and **end tags**

- E.g. <book> is the start tag and </book> is the end tag

# XML EXPLAINED

The two pointy brackets is called an **element**

- E.g. <book> would be called the book element

All elements have **start** and **end tags**

- E.g. <book> is the start tag and </book> is the end tag

Elements can also have **attributes** and each attribute must have a **value**

- E.g. <book type=“primary”> has a **type attribute with the value of primary**
- (Think of attributes as you would in everyday life; people don’t have “height” or “age” without a value)

# XML EXPLAINED

# XML EXPLAINED

Elements **cannot** overlap

- <sentence><word>Word1</word></sentence> is right
- <sentence><word>Word1</sentence></word> is wrong

# XML EXPLAINED

## Elements **cannot** overlap

- <sentence><word>Word1</word></sentence> is right
- <sentence><word>Word1</sentence></word> is wrong

## Elements **nest** and use genealogical terms

- I.e this bit of XML

```
<book>
  <chapter></chapter>
</book>
```

Can be described as “chapter is a child of book” OR “book is a parent of chapter”

# XML EXPLAINED

## Elements **cannot** overlap

- <sentence><word>Word1</word></sentence> is right
- <sentence><word>Word1</sentence></word> is wrong

## Elements **nest** and use genealogical terms

- I.e this bit of XML

```
<book>
  <chapter></chapter>
</book>
```

Can be described as “chapter is a child of book” OR “book is a parent of chapter”

## There is **always** a **root** element

- That is, there is always one element that encloses everything

# WHY DO WE ENCODE?

Interpretation

Usability (text mining, et cetera)

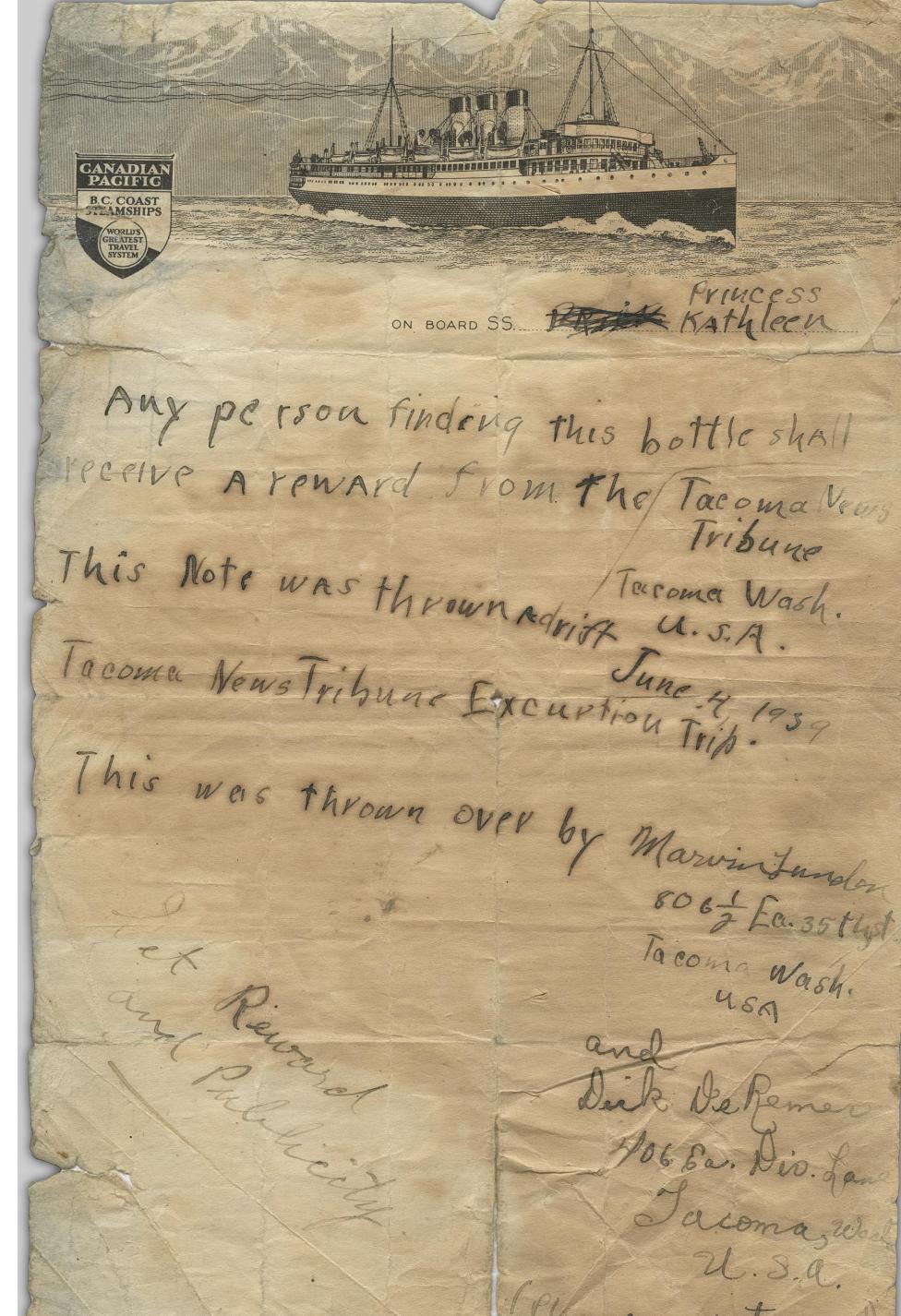
Interoperability

Preservation

# WHAT ARE WE ENCODING TODAY?

Description: A letter thrown off of the Princess Kathleen in a bottle, June 4, 1939 by Marvin Lunder and Dick de Remer of Tacoma Washington, and found in October by W[illiam] B. D. Firmming on Whilby Island.

Source: Original Format: University of British Columbia. Library. Rare Books and Special Collections. The Chung Collection. CC-TX-151-18



# WHAT ARE SOME FEATURES OF THIS TEXT?

Blocks of text (are they paragraphs?)

Lines

Divisions?

Signatures

Opener/closer

Names

Dates

++++

# THE TEI GUIDELINES

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>

# TEI IS FOR DATA AND METADATA

```
<TEI>
  <teiHeader>
  </teiHeader>
  <text>
    </text>
  </TEI>
```

metadata

data

# ELEMENTS

Each element has its own page in the guidelines that includes:

- A definition of the element
- Links to chapters in the TEI guidelines
- What attributes are available on that element
- What elements can contain that element
- What elements can that element contain
- Example uses of that element
- +

# EXAMPLE: THE <MONOG> ELEMENT:

<monogr>		<a href="#">Home</a> <a href="#">C Elements</a>
<monogr> (monographic level) contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item (i.e. as a separate physical object). [3.11.2.1 Analytic, Monographic, and Series Levels]		
<b>Module</b>	core — Elements Available in All TEI Documents	
<b>Attributes</b>	<a href="#">att.global</a> (@xml:id, @n, @xml:lang, @xml:base, @xml:space) ( <a href="#">att.global.rendition</a> (@rend, @style, @rendition)) ( <a href="#">att.global.linking</a> (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) ( <a href="#">att.global.analytic</a> (@ana)) ( <a href="#">att.global.facs</a> (@facs)) ( <a href="#">att.global.change</a> (@change)) ( <a href="#">att.global.responsibility</a> (@cert, @resp)) ( <a href="#">att.global.source</a> (@source))	
<b>Contained by</b>	<a href="#">core: biblStruct</a>	
<b>May contain</b>	<a href="#">core: author</a> <a href="#">biblScope</a> <a href="#">editor</a> <a href="#">imprint</a> <a href="#">meeting</a> <a href="#">note</a> <a href="#">ptr</a> <a href="#">ref</a> <a href="#">respStmt</a> <a href="#">textLang</a> <a href="#">title</a> <a href="#">header: authority</a> <a href="#">availability</a> <a href="#">edition</a> <a href="#">extent</a> <a href="#">funder</a> <a href="#">idno</a> <a href="#">sponsor</a> <a href="#">textcrit: witDetail</a>	
<b>Note</b>	May contain specialized bibliographic elements, in a prescribed order. The <a href="#">monogr</a> element may only occur only within a <a href="#">biblStruct</a> , where its use is mandatory for the description of a monographic-level bibliographic item.	
<b>Example</b>	<pre>&lt;biblStruct&gt;   &lt;analytic&gt;     &lt;author&gt;Chesnutt, David&lt;/author&gt;     &lt;title&gt;Historical Editions in the States&lt;/title&gt;   &lt;/analytic&gt;   &lt;monogr&gt;     &lt;title level="j"&gt;Computers and the Humanities&lt;/title&gt;     &lt;imprint&gt;       &lt;date when="1991-12"&gt;(December, 1991)&lt;/date&gt;       &lt;imprintn&gt;         &lt;biblScope&gt;25, 6&lt;/biblScope&gt;         &lt;biblScope unit="page" from="377" to="380"&gt;377-380&lt;/biblScope&gt;       &lt;/imprintn&gt;     &lt;/monogr&gt;   &lt;/biblStruct&gt;</pre>	<a href="#">Show all</a>

# EXAMPLE: THE <MONOG> ELEMENT:

The diagram illustrates the structure of the TEI XML schema. A large green arrow labeled "Definition" points to the first row of the table, which contains general information about the `<monogr>` element. Another green arrow labeled "Definition chapters" points to the second row, which provides specific details about the attributes and contained elements of the `<monogr>` element.

<monogr>	
<monogr> (monographic level) contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item (i.e. as a separate physical object). [3.11.2.1 Analytic, Monographic, and Series Levels]	
<b>Module</b>	core — Elements Available in All TEI Documents
<b>Attributes</b>	<code>att.global</code> (@xml:id, @n, @xml:lang, @xml:base, @xml:space) ( <code>att.global.rendition</code> (@rend, @style, @rendition)) ( <code>att.global.linking</code> (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) ( <code>att.global.analytic</code> (@ana)) ( <code>att.global.facs</code> (@facs)) ( <code>att.global.change</code> (@change)) ( <code>att.global.responsibility</code> (@cert, @resp)) ( <code>att.global.source</code> (@source))
<b>Contained by</b>	<code>core: biblStruct</code>
<b>May contain</b>	<code>core: author</code> <code>biblScope</code> <code>editor</code> <code>imprint</code> <code>meeting</code> <code>note</code> <code>ptr</code> <code>ref</code> <code>respStmt</code> <code>textLang</code> <code>title</code> <code>header: authority</code> <code>availability</code> <code>edition</code> <code>extent</code> <code>funder</code> <code>idno</code> <code>sponsor</code> <code>textcrit: witDetail</code>
<b>Note</b>	May contain specialized bibliographic elements, in a prescribed order. The <code>monogr</code> element may only occur only within a <code>biblStruct</code> , where its use is mandatory for the description of a monographic-level bibliographic item.
<b>Example</b>	<pre>&lt;biblStruct&gt;   &lt;analytic&gt;     &lt;author&gt;Chesnutt, David&lt;/author&gt;     &lt;title&gt;Historical Editions in the States&lt;/title&gt;   &lt;/analytic&gt;   &lt;monogr&gt;     &lt;title level="j"&gt;Computers and the Humanities&lt;/title&gt;     &lt;imprint&gt;       &lt;date when="1991-12"&gt;(December, 1991)&lt;/date&gt;       &lt;idno&gt;25.6&lt;/idno&gt;       &lt;biblScope&gt;25.6&lt;/biblScope&gt;       &lt;biblScope unit="page" from="377" to="380"&gt;377-380&lt;/biblScope&gt;     &lt;/monogr&gt;   &lt;/biblStruct&gt;</pre> <a href="#">Show all</a>

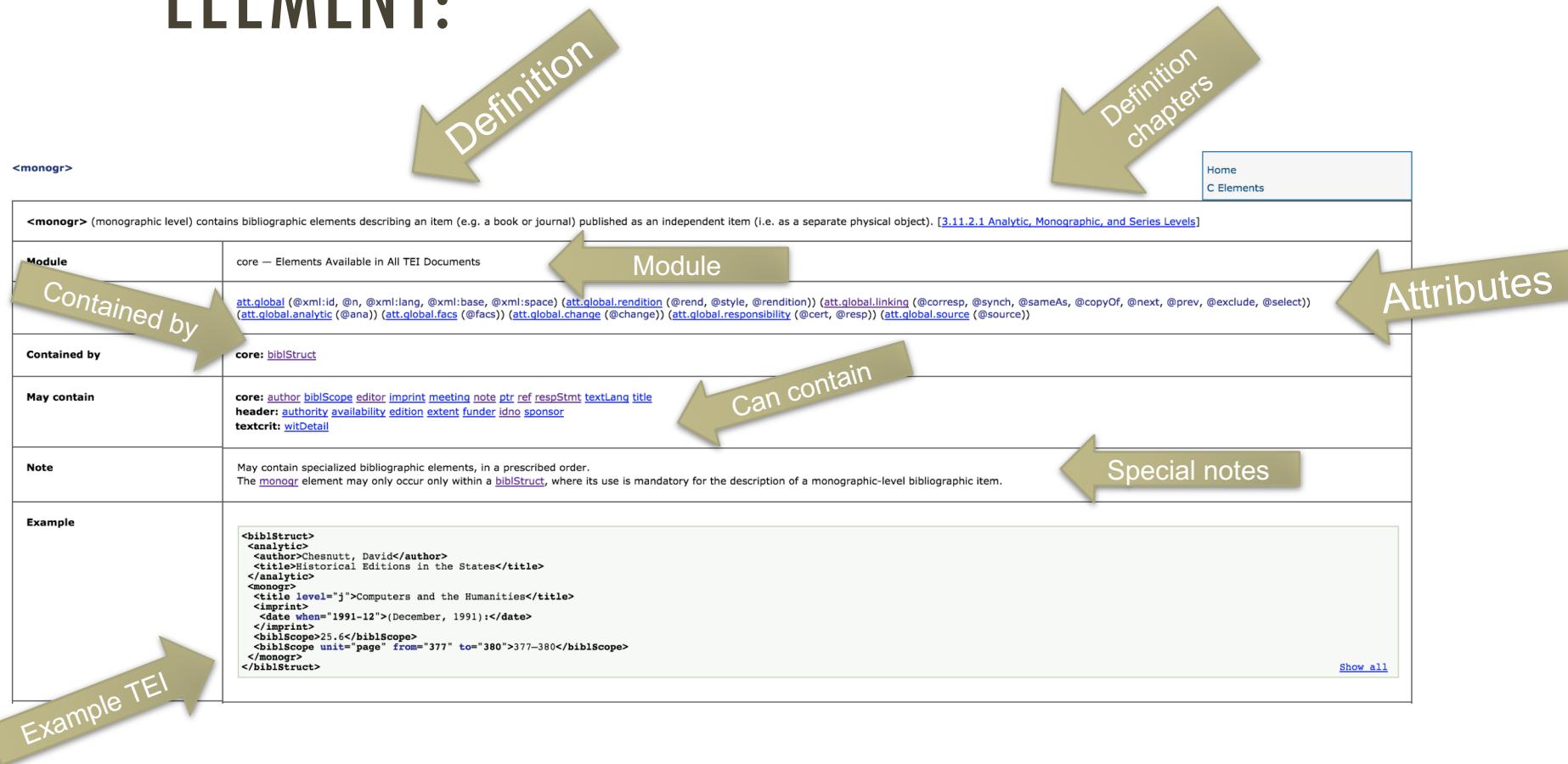
# EXAMPLE: THE <MONOG> ELEMENT:

<monogr>	
<monogr> (monographic level) contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item (i.e. as a separate physical object). [3.11.2.1 Analytic, Monographic, and Series Levels]	
<b>Module</b>	core — Elements Available in All TEI Documents
<b>Attributes</b>	<code>att.global (@xml:id, @n, @xml:lang, @xml:base, @xml:space) (att.global.rendition (@rend, @style, @rendition)) (att.global.linking (@corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select)) (att.global.analytic (@ana)) (att.global.facs (@facs)) (att.global.change (@change)) (att.global.responsibility (@cert, @resp)) (att.global.source (@source))</code>
<b>Contained by</b>	<code>core: biblStruct</code>
<b>May contain</b>	<code>core: author biblScope editor imprint meeting note ptr ref respStmt textLang title</code> <code>header: authority availability edition extent funder idno sponsor</code> <code>textcrit: witDetail</code>
<b>Note</b>	May contain specialized bibliographic elements, in a prescribed order. The <code>monogr</code> element may only occur only within a <code>biblStruct</code> , where its use is mandatory for the description of a monographic-level bibliographic item.
<b>Example</b>	<pre>&lt;biblStruct&gt;   &lt;analytic&gt;     &lt;author&gt;Chesnutt, David&lt;/author&gt;     &lt;title&gt;Historical Editions in the States&lt;/title&gt;   &lt;/analytic&gt;   &lt;monogr&gt;     &lt;title level="j"&gt;Computers and the Humanities&lt;/title&gt;     &lt;imprint&gt;       &lt;date when="1991-12"&gt;(December, 1991)&lt;/date&gt;       &lt;idno type="oclc"&gt;25_6&lt;/idno&gt;       &lt;biblScope&gt;25_6&lt;/biblScope&gt;       &lt;biblScope unit="page" from="377" to="380"&gt;377-380&lt;/biblScope&gt;     &lt;/monogr&gt;   &lt;/biblStruct&gt;</pre> <a href="#">Show all</a>

# EXAMPLE: THE <MONOG> ELEMENT:

<monogr>	
<monogr> (monographic level) contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item (i.e. as a separate physical object). [3.11.2.1 Analytic, Monographic, and Series Levels]	
<b>Module</b>	core — Elements Available in All TEI Documents
<b>Contained by</b>	core: <a href="#">biblStruct</a>
<b>May contain</b>	core: <a href="#">author</a> <a href="#">biblScope</a> <a href="#">editor</a> <a href="#">imprint</a> <a href="#">meeting</a> <a href="#">note</a> <a href="#">ptr</a> <a href="#">ref</a> <a href="#">respStmt</a> <a href="#">textLang</a> <a href="#">title</a> header: <a href="#">authority</a> <a href="#">availability</a> <a href="#">edition</a> <a href="#">extent</a> <a href="#">funder</a> <a href="#">idno</a> <a href="#">sponsor</a> textcrit: <a href="#">witDetail</a>
<b>Note</b>	May contain specialized bibliographic elements, in a prescribed order. The <a href="#">monogr</a> element may only occur only within a <a href="#">biblStruct</a> , where its use is mandatory for the description of a monographic-level bibliographic item.
<b>Example</b>	<pre>&lt;biblStruct&gt;   &lt;analytic&gt;     &lt;author&gt;Chesnutt, David&lt;/author&gt;     &lt;title&gt;Historical Editions in the States&lt;/title&gt;   &lt;/analytic&gt;   &lt;monogr&gt;     &lt;title level="j"&gt;Computers and the Humanities&lt;/title&gt;     &lt;imprint&gt;       &lt;date when="1991-12"&gt;(December, 1991)&lt;/date&gt;       &lt;place&gt;New York&lt;/place&gt;     &lt;biblScope&gt;25.6&lt;/biblScope&gt;     &lt;biblScope unit="page" from="377" to="380"&gt;377-380&lt;/biblScope&gt;   &lt;/monogr&gt; &lt;/biblStruct&gt;</pre> <a href="#">Show all</a>

# EXAMPLE: THE <MONOG> ELEMENT:



# ACTIVITY (5 MIN)

Look up the following elements:

- <p>
- <ab>
- <opener>
- <closer>
- <msDesc>
- <correspDesc>

Will they work for this? How can we use them?

# ENCODING

Open the text editor in one tab, the primary source transcription in another, and, if you want, the image from the UBC Repository

# METADATA

Include the title of this document

Include your name by using the <editor> element after the <title>

Put in a link to the source using a <ptr/>

What else could you do?

# DATA

Copy and paste the content into the document

What are the organizing divisions? Divs? Paragraphs? Floating text? Anonymous blocks?

# WHAT ARE SOME OTHER FEATURES?

Names?

Dates

Notes

Editorial apparatus (if you want)

Hands

Etc

# TOOLS

Download your XML file

And now, in the index page, look for the tools:

TEI Validator

TEI Toolkit

OxGarage

# VALIDATOR

Validate your file

# TOOLKIT

Check your encoding

Convert to a PDF

# OXGARAGE

[Convert to a word document](#)

[Convert to a PDF](#)

# THANKS!

UBC Libraries, esp. Allan Cho for the invitation!

Martin Holmes, Janelle Jenstad, The Endings Project, and Uvic's Humanities Media and Computing Center