Assembling the digital postcard archive

Lou Burnard Consulting *September 2014*

1 TEI à la carte : customization

In this tutorial, we'll see how TEI markup might be used in an imaginary (but quite plausible) research project. We'll discuss what markup scheme the project will need, create a TEI schema to support it, and do some experimental markup using it.

As you probably don't know, the late Marcel Virgolos has long been recognised as an expert on the history and evolution of the 20th century postcard. Following his untimely demise, we have received permission from his executors to create a digital research archive of some parts of his unique personal collection of 100,000 mostly used postcards of all kinds in order to further the progress of serious cartophilological studies. We've put digital images of a tiny sample from the collection in the folder Cards inside your Work folder.

1.1 Sample text

In this exercise we will start by marking up one postcard from the collection, respecting its organization into recto and verso, and the components of its address, etc.

Images of the recto and verso of the card are in figures 1 and 2 below (and also in your Work/Cards folder)

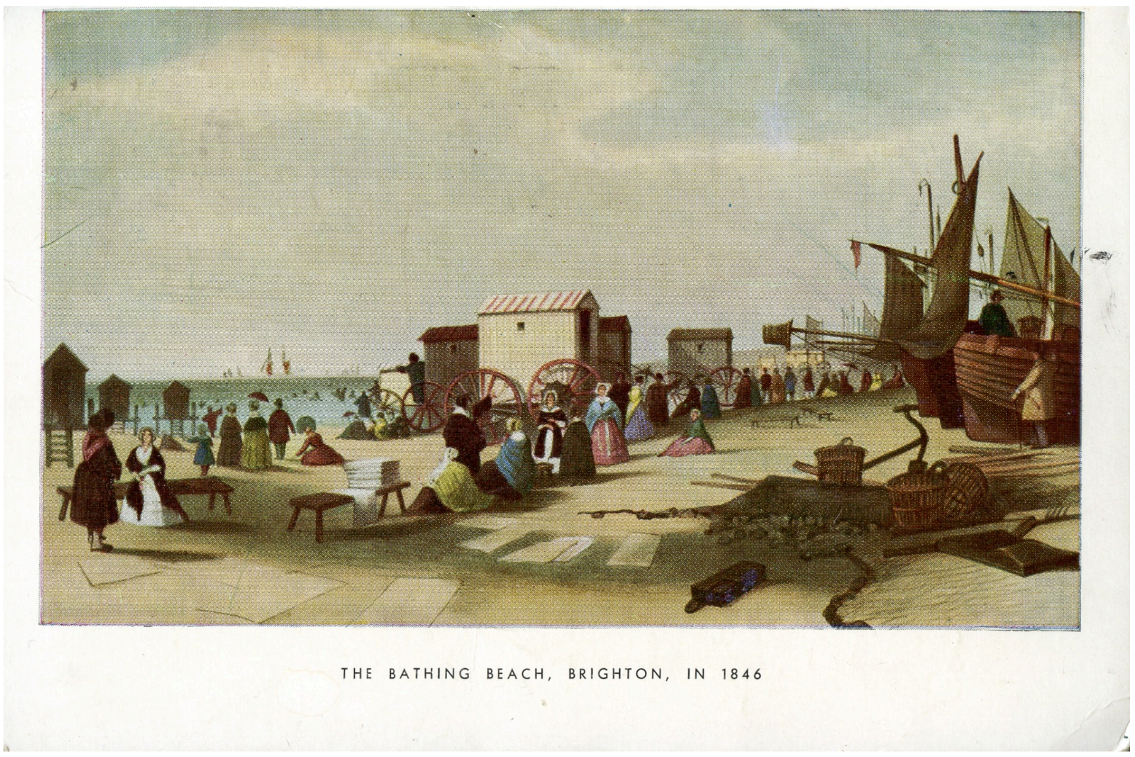
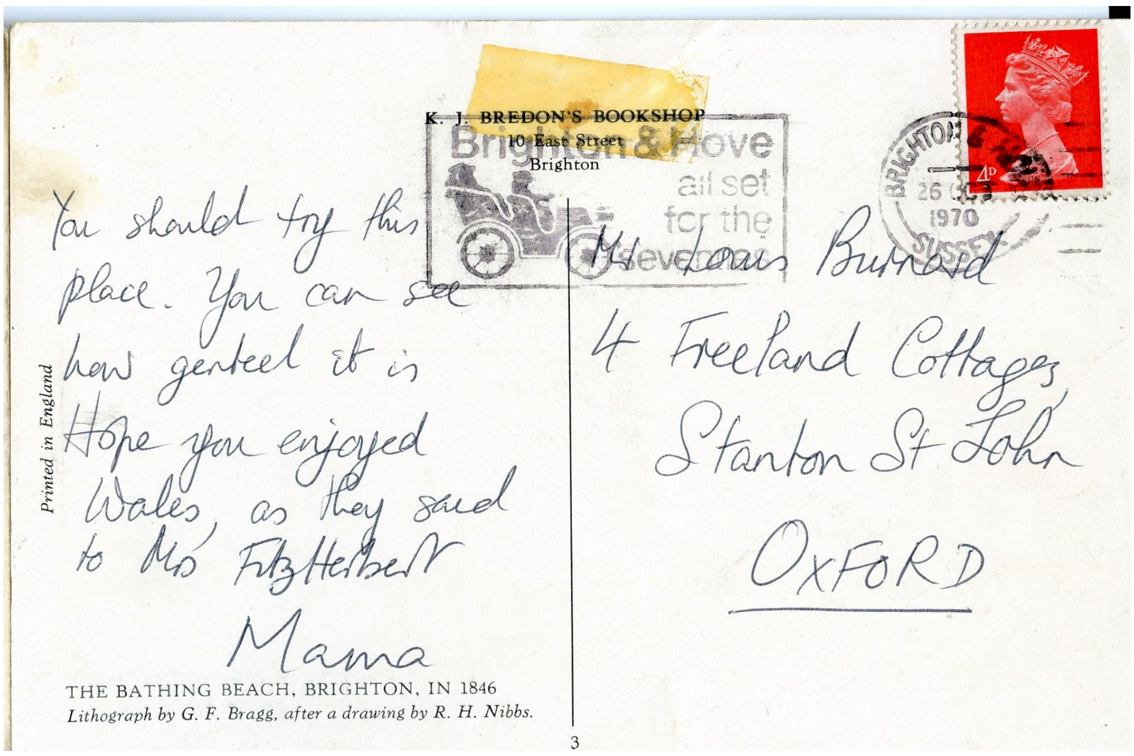


Figure 1: Postcard 197001026_0004 (recto). From the Virgolos Collection

Of course the first part of our research project was spent arguing about how to markup these documents, and if we had more time we'd rerun some of that discussion now. We used Roma to create an appropriate subset of our needs, expressed as a schema called teiCards. As with other customizations, a particular kind of TEI document is used to create and document this schema, which we call an ODD (for One Document Does it all).



K. J. BREDON'S BOOKSHOP
10 East Street
Brighton

Brighton & Hove all set for the veterans

BRIGHTON 26 1970 SUSSEX

4D

Printed in England

You should try this place. You can see how genteel it is
Hope you enjoyed Wales, as they said to Mrs FitzHerbert
Mama

Mr Lou Burnard
4 Freeland Cottages
Stanton St John
OXFORD

THE BATHING BEACH, BRIGHTON, IN 1846
Lithograph by G. F. Bragg, after a drawing by R. H. Nibbs.

3

Figure 2: Postcard 197001026_0004 (verso). From the Virgolos Collection

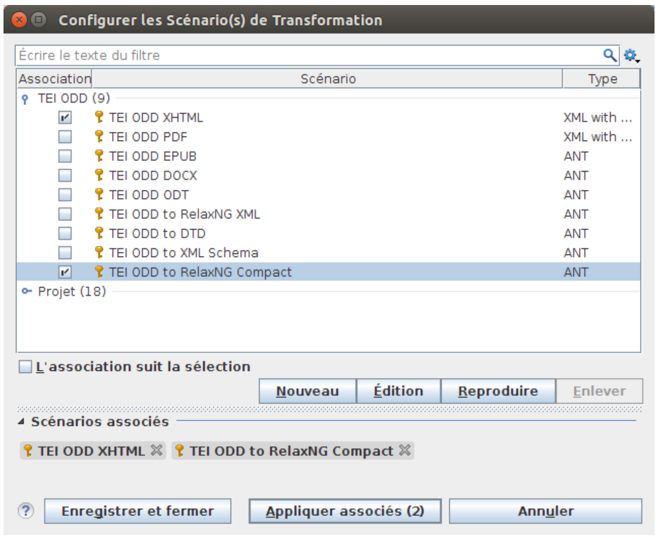
1.2 Processing an ODD

You will learn more about ODDs later with Sebastian. For the moment, we will simply use oXygen to process it.

- Open oXygen
- Open the file teiCards.odd which is available in your Work/Cards folder

As you can see, this is a TEI document like any other, except that it uses some additional elements we have not introduced yet. Fortunately oXygen now includes in its TEI framework ways of processing these additional elements automatically.

- Select Transformation -> ConfigureTransformation Scenario(s) from the Document menu, ou type CTRL-SHIFT-C, or click the spanner + red triangle icon on the tool bar.
- A list of available transformations for the file you are currently editing is displayed. Scroll down to find the nine available for TEI ODD and check the boxes for those which interest us, for example : TEI ODD to XHTML and TEI ODD to RELAXNG Compact

- 

Configurer les Scénario(s) de Transformation

Écrire le texte du filtre

Association	Scénario	Type
TEI ODD (9)		
✓	TEI ODD XHTML	XML with ...
	TEI ODD PDF	XML with ...
	TEI ODD EPUB	ANT
	TEI ODD DOCX	ANT
	TEI ODD ODT	ANT
	TEI ODD to RelaxNG XML	ANT
	TEI ODD to DTD	ANT
	TEI ODD to XML Schema	ANT
✓	TEI ODD to RelaxNG Compact	ANT
Projet (18)		

- Click the button Apply Associated and wait a minute or two.

An HTML file will be generated, and should be displayed by your default browser: as you can see, it is like any other TEI specification. A schema file in RELAXNG compact form, will be created and stored in a subdirectory called out. We will use this in the next step.

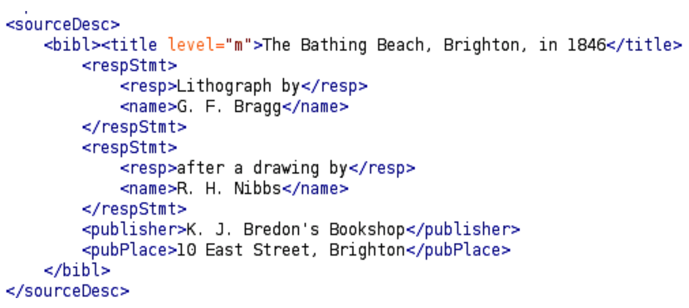
1.3 Create a new document

We will now create a new document which uses the schema generated from this ODD.

- Open oXygen
- Click on the New icon at top left (or select New from the File menu, or type CTRL-N) to open the New document dialogue.
- Choose New Document and then XML Document
- Click on the Customize button at the bottom of the dialog box, to open the Customize Editor dialog box.
- At the far right of the box labelled Schema URL there is a small triangle which you can click to open a drop down menu.
- Select Browse for local file, and navigate to the file teiCards.rnc in your Work/Cards/out folder. Select this file and click the Open button.
- Back in the New document dialogue, press the Create button at the bottom.

oXygen offers you a framework derived from the schema for you to complete. Let's begin by completing the metadata in the TEI Header:

- You need to supply a title (in the`<title>` element): we suggest The Bathing Beach, Brighton, 1846 : digital edition of card 19701026_0004 from the Virgolos collection. For the publication statement, something like 'Unpublished sample, produced as part of the second DiXit Workshop, Graz 2014' would be fine. Don't forget to wrap this piece of text within a `<p>` inside the `<publicationStmt>` element!
- For the source description, if you like, you could include all the bibliographic information from the verso of the card, like this:



```
<sourceDesc>
    <bibl><title level="m">The Bathing Beach, Brighton, in 1846</title>
        <respStmt>
            <resp>Lithograph by</resp>
            <name>G. F. Bragg</name>
        </respStmt>
        <respStmt>
            <resp>after a drawing by</resp>
            <name>R. H. Nibbs</name>
        </respStmt>
        <publisher>K. J. Bredon's Bookshop</publisher>
        <pubPlace>10 East Street, Brighton</pubPlace>
    </bibl>
</sourceDesc>
```

1.4 Adding text to the document

You can, of course, make your own transcription But to save time we propose the following method:

- Check that your cursor is *between* the `<body>` and `</body>` tags in your empty document
- Choose Document -> File -> Insert File from the menu bar (*Not the File Menu on the menu bar, but the one on the Document Menu*)
- Navigate to the file 19701026_0004.txt which you will find in the Cards folder , select it, and click Open.
- Your document is full of red lines but don't panic! we will sort that out in a moment.

1.5 Adding structure to the body of a document

Let's begin by thinking about the structure of this document. As we discussed earlier, every card, including this one, has

- two divisions ... to be tagged with `<<div type="recto">>` and `<<div type="verso">>` respectively
- inside the verso (in this case) there is a further division : one part contains the message, and the other contains an address and other information about the destination of the card, such as the postmark, postage stamp etc.
- Our markup should distinguish all these things. Note that we think this approach will be more useful than simply mimicking the physical layout of the card – we have a picture of it, after all.

Let's go !

- Use the mouse to select all the text you just added, including the initial `<graphic>` element.
- Type CTRL-E (or select XML Refactoring -> Surround with Tags from the Document menu)
- oXygen shows you all the tags available at this point: choose `div` and click Accept.
- In this schema, we must specify a value for the *@type* attribute on every `<div>` element. With the cursor just in front of the > of the `<div>` opening tag type a space to see the list of available attributes. Choose type (it is in bold because it must be supplied) and press RETURN to insert it.
- oXygen displays the predefined list of possible values for this attribute . Choose "recto" and type RETURN to insert it.
- Put the cursor after the word '1846' and type SHIFT-ALT-D (or select XML Refactoring - Split element from the Document menu) to split the element.

- The bit of text ”Beach view...and mats” describes the image but is not actually part of the card. The element provided for such a description is `<figDesc>`. Select that stretch of texts, type CTRL-E and wrap a `<figDesc>` round it. (Don’t worry that this element name doesn’t appear on the list of tags available: you can still type it in).
- The bit of text ”The Bathing ... 1846” is the caption of a graphic. The tag for this is `<head>`. Select that stretch of texts, type CTRL-E and wrap a `<head>` round it.
- The three elements you’ve now marked up (`<graphic>` `<figDesc>` `<head>`) together form a `<figure>`. Select all three of them and wrap a `<figure>` round them. The red lines at the start of your text should now all have gone. Onward!

Now for the verso ... Our aim is to separate the part containing the message (<div type=”message”>) from the part concerned with the sending of the card (<div type=”destination”>). We will use the `<p>` element for paragraphs of transcribed text, `<stamp>` for the various kinds of stamp, and `<address>` and `<addrLine>` for the address and its lines respectively. Many other tags are possible, of course; we are starting simple.

Before we begin, note that we actually have two versions of the verso of the card : one is the image file, and the other is a transcription of it. We will use the attribute *@facs* to associate the image file with its transcription (other, more complex, methods are of course possible).

- Change the value `recto` inside the `<div>` start tag which begins our second `<div>` element into `verso`. Type a space. A list of other available attributes appears.
- Choose facs from the list. Its value should be the string `197001026_004v.jpg`, which is currently the value for the *@url* attribute on a `<graphic >`element. Cut and paste this string into the right place, and then delete the rest of the `<graphic>` element : we don’t need it.
- Now select the whole of the text (i.e. from ”You” to ”OXFORD”) and type CTRL-E to surround it with a single `<div>`, and then (with the text still selected), repeat to surround it all with a `<p>` inside the `<div>`.
- Put the cursor just before the > of the opening tag of the `<div>` that you just inserted, type a space, and select *@type* from the list of attributes available. Press return and you will see what values are predefined for this attribute: choose `message`.
- Almost all the red lines disappear : can you see why we still have a couple associated with those ampersands? The error message at the bottom of the screen may help.
- That’s right : there are just a couple of characters which are magic inside an XML document, and hence need to be treated specially. The pointy bracket is one (for obvious reasons) and the ampersand is the other. Just replace the ampersands with the sequence `&amp;`, so that it now reads `Brighton &amp; Hove`
- The little green square appears! Have we finished? Sorry, no : a document can be valid while still containing many lies ... and lies can also be dangerous : to see why, hit the Indentation button on the toolbar (or type CTRL-SHIFT-P).

Since the line endings are not explicit in the markup, oXygen is not obliged to respect them when formatting the display. To be both honest and explicit, we need to separate the paragraphs in the message from each other and the signature, and we also need to separate the descriptive text about the stamps from the transcribed text.

- Bring back the original display by typing CTRL-Z.

- Split the one paragraph into several, by pressing SHIFT-ALT-D with the cursor positioned after the words ‘is’, ‘Fitzherbert’, ‘Mama’, and ‘vermilion’

- Now put the cursor between the end of the paragraph containing the word ‘Mama’ and the start of the next paragraph, i.e. *between* the `</p>` after ”Mama” and the `<p>` that follows it. Type SHIFT-ALT-D to split the inner division (the `<div type="message">` you just created).

- The word ‘Mama’ isn’t part of the message but a signature, for which the element `<signed>` is provided. You can fix this just by retyping the tags, or, if you prefer, put the cursor inside the `<p>` start-tag, and select Document - XML refactoring - Rename element (ALT-SHIFT-R). Then choose signed from the list of available elements and press Return.

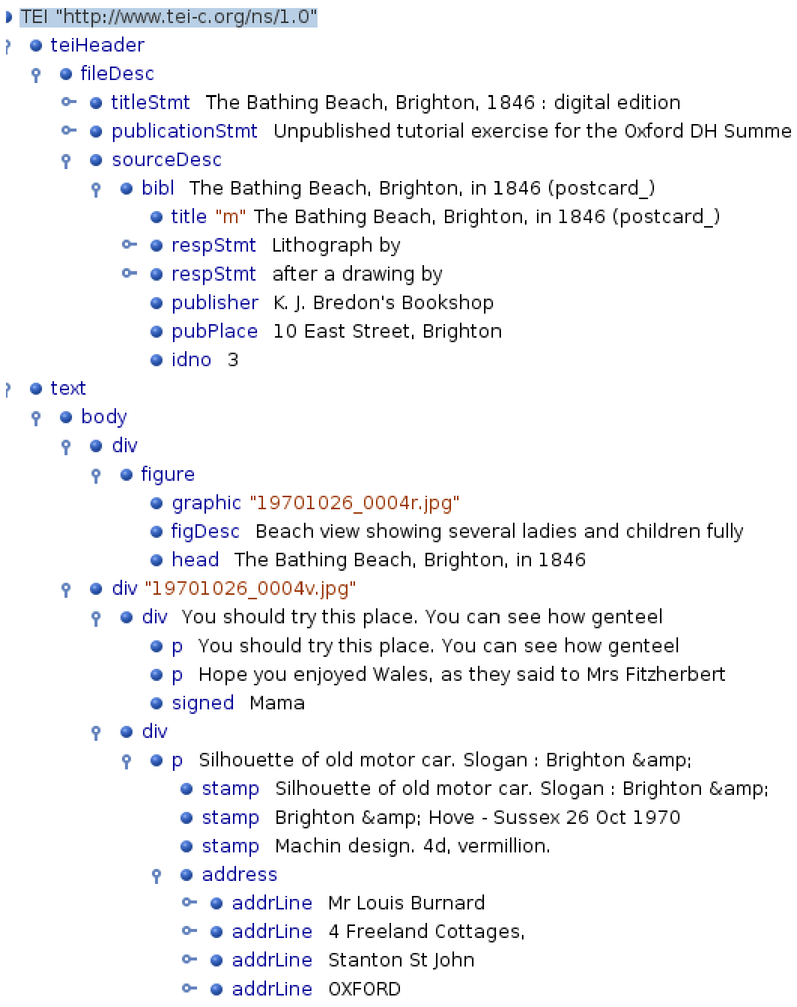
- The remaining `<div>` contains three stamp descriptions and an address, not a message: according to our schema, its *@type* value should be `destination` : make it so.

- Now tag each of the three stamp-descriptions with the `<stamp>` element (we leave it to you whether you do that by making a single `<stamp>` element and splitting it, or by selecting each description and enclosing it with CTRL-E)

- Lastly, let’s deal with the addressee. Select the whole of the address text and (using CTRL-E) tag it as `<address>`. Red line reappears! That’s because the components of an `<address>` must be distinguished : you cannot have just plain text inside this tag. You could markup those components of the address which are names (using `<name>`) and those which make up a street address (using `<street>`); or you could just tag each line of the address separately using `<addrLine>`. You could even do both (put the names etc.*inside* the address lines, if you do)... As is often the way in the TEI, it’s Up To You.

1.6 Reality check

As you’ve already seen, oXygen can display the hierarchic structure of the document – if you look in the Summary window to the left and expand some of the buttons there you should now see something like this :



```
TEI "http://www.tei-c.org/ns/1.0"
  teiHeader
    fileDesc
      titleStmt  The Bathing Beach, Brighton, 1846 : digital edition
      publicationStmt  Unpublished tutorial exercise for the Oxford DH Summe
      sourceDesc
        bibl  The Bathing Beach, Brighton, in 1846 (postcard_)
          title "m" The Bathing Beach, Brighton, in 1846 (postcard_)
          respStmt  Lithograph by
          respStmt  after a drawing by
          publisher  K. J. Bredon's Bookshop
          pubPlace  10 East Street, Brighton
          idno  3
  text
    body
      div
        figure
          graphic "19701026_0004r.jpg"
          figDesc  Beach view showing several ladies and children fully
          head  The Bathing Beach, Brighton, in 1846
      div "19701026_0004v.jpg"
        div  You should try this place. You can see how genteel
          p  You should try this place. You can see how genteel
          p  Hope you enjoyed Wales, as they said to Mrs Fitzherbert
          signed  Mama
        div
          p  Silhouette of old motor car. Slogan : Brighton &amp;
            stamp  Silhouette of old motor car. Slogan : Brighton &amp;
            stamp  Brighton &amp; Hove - Sussex 26 Oct 1970
            stamp  Machin design. 4d, vermillion.
            address
              addrLine  Mr Louis Burnard
              addrLine  4 Freeland Cottages,
              addrLine  Stanton St John
              addrLine  OXFORD
```

.

Have we finished our markup now?

Clearly there is no simple answer to this question. We've marked up the basic structure of the document in such a way that we can identify and display its components in various ways and carry out searches by date or location. But we haven't marked up everything possible within that structure.

As you are probably quite bored with this particular postcard by now, we suggest you take a look at some of the others in the Cards folder. We've s upplied images of the recto and verso for each card, and some preliminary transcription as well. But you will need to check and tag these carefully! See how many cards you can do in the remaining time. and don't forget to save all your encoded versions ! We'll use them later on.