

TEI à la carte: a case study

Lou Burnard

Where do you get a schema ?

- Different projects have different requirements... but also overlap
- and there will always be some unexplored areas (that's what research is for)
- which is why the TEI is designed the way it is...

The TEI architecture to the rescue

The TEI offers you a semi-automatic procedure for selecting from hundreds of markup specifications. You can:

- Just choose everything... (probably not a very good idea)
- Work with one of the predefined selections (TEI Lite, TEI Bare...)
- Roll your own, according to the specific needs of your project

[Roma](#) an online tool which makes this task easier

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

For example ...

TEI out of the box is designed to work with traditionally organised books and manuscripts. But suppose we want to work on a slightly different kind of object... a postcard collection, or a monumental inscription? How do we make a TEI schema capable of handling hundreds or thousands of things like this:

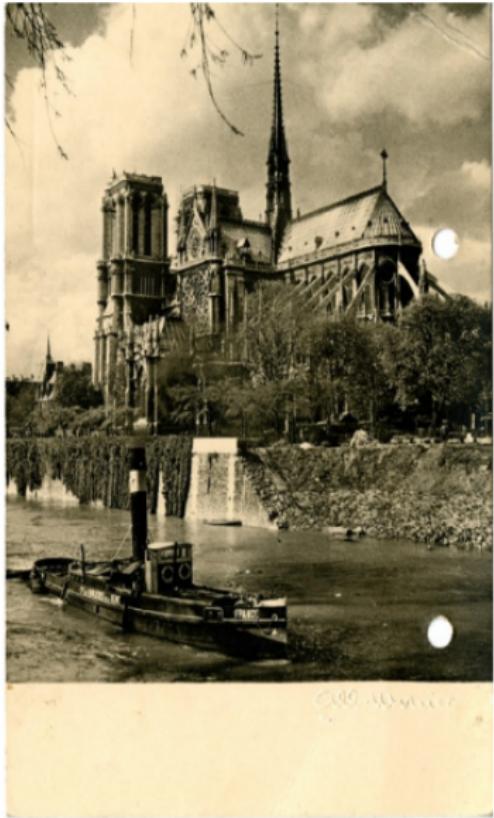
A postcard (front)



A postcard (back)



Another postcard



2/ 193 — PARIS - Notre-Dame. 24.7.53 - 19^H
Mon Amour,
Je ne peux que penser à toi -
il m'est impossible de fixer
mon esprit, sur un autre
sujet (pratique) aussi important
soit-il - il n'y a pas plus
de Commerce que de ministre
d'avis que de travail . et même
ma famille , père et mère . en
ont l'air surpris - Je rêve
littéralement - c'est fou . Je
t'en suffis laisse moi un peu
il faut penser à l'avenir . dont
tu parles . heureux . riche . utile .
Je t'aime et . - je me pense qu'à toi (je)

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Which are the most significant components of these texts?

- the picture
- the postmark
- the printed part
- the message(s) written on them
- the addressee(s)
- subject matter of the picture
- information about the publishing, printing, circulation of the card or other metadata...

Suggestion

We could begin by structuring the card as divisions of various types

Physically:

- **recto**: one side, usually the one with the picture
- **verso**: the other side, usually the one with the message

On these two surfaces, we expect to find various other subsections, such as:

- the message
- information about the sending of the card, notably:
 - the addressee
 - the postmark, stamp, etc.
- data about the publication, sale, collection, etc. of this card

First try at encoding a postcard

```
<carte n="0010">
<recto url="cartes/19800726_001r.jpg"/>
<verso url="cartes/19800726_001v.jpg">
<obliteration>
<date>PM ?? Jul ???</date>
<lieu>EL PASO. TX 799</lieu>
</obliteration>
<message>
<p>26 juill 80</p>
<p>Chère Madame, après New-
York et Washington dont le gigantisme m'a beaucoup séduite, nous avons commencé notre conquête de l'
cola permet de résister -
l'Amérique m'enchanté ! Bientôt, le grand Canyon, le Colorado et San Francisco... En espérant que vo...
<p> Sylvie </p><p>François. </p>
</message>
<destinataire>
Madame Lefrère
4, allée George Rouault
75020 Paris
France
</destinataire>
</verso>
</carte>
```

Commentary

- We didn't use the TEI vocabulary. This means we may have trouble sharing or explaining our data with non-french speakers. Or benefitting from their work.
- We haven't included all the things that might be encoded: for example, corrections in the text, layout of the components, names of people or places referred to, linguistic or historical features, bibliographic data about where the card was printed ...
- We haven't structured (for example) the address, which will make intelligent searching difficult.
- Of course, we can always invent more tags for these things. But isn't it rather a waste of our time if the TEI has already done the job ?

TEI version

- We regard each card as a `<text>` containing two `<div>` elements, one for the recto and one for the verso of each card.
- We markup each functional division of the card as a `<div type=" [function] " >`
- Metadata about the published card will go in a `<bibl>` in the TEI Header
- We markup names of people and places with `<name>` and dates with `<date>`
- We use the attribute `@facs` to associate parts of transcribed text with their digital image, indicated by a `<graphic>` element.
- We use `<address>` for the address; `<stamp>` element for stamps, postmarks, and similar things.
- We may also need `` (for deletions), `<add>` (for additions), `<reg>` (for regularized spellings), `<unclear>` for things we cannot read, `<lb>` for line breaks ...

First try at a TEI version : the header

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>San Antonio River : digital edition of card 19800726_001 from the Virgolos
collection</title>
    </titleStmt>
    <publicationStmt>
      <p>Demonstration at DH OXSS 2013</p>
    </publicationStmt>
    <sourceDesc>
      <bibl>
        <title level="m">San Antonio River (postcard)</title>
        <publisher>School Mart</publisher>
        <pubPlace>1812 South Press, San Antonio, Texas 70210</pubPlace>
        <idno>SA-146-C</idno>
        <note resp="#ed">The San Antonio river, often called the Venice of Texas, winds its
way through the business section of San Antonio. It is very picturesque with its many
bridges and beautifully landscaped banks.</note>
      </bibl>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

First try at a TEI version : the text

```
<text>
  <body>
    <div type="recto">
      <figure>
        <graphic url="../../Graphics/Cartes/19800726_001r.jpg"/>
        <figDesc>View of a stream with a stone bridge and little mexican-style houses. In the foreground a man and a woman are riding in an open boat.</figDesc>
        <head>San Antonio River</head>
      </figure>
    </div>
    <div facs="19800726_001v.jpg"
      type="verso">
      <div type="message">
<!-- ... --&gt;
      &lt;/div&gt;
      &lt;div type="destination"&gt;
        &lt;p&gt;
<!-- stamps --&gt;
        &lt;/p&gt;
        &lt;p&gt;
          &lt;address&gt;
<!-- ... --&gt;
          &lt;/address&gt;
        &lt;/p&gt;
      &lt;/div&gt;
    &lt;/div&gt;
  &lt;/body&gt;
&lt;/text&gt;</pre>
```

First try at a TEI version : the message

```
<div type="message" xml:lang="fr">
  <p>
    <date when="1980-07-26">26 juill 80</date>
  </p>
  <p>Chère Madame, après New-York et Washington dont le gigantisme m'a beaucoup séduite, nous
avons commencé notre conquête de l'Ouest par New Orleans, ville folle en fête perpétuelle.
Il fait une chaleur torride au Texas mais le coca-cola permet de résister – l'Amérique
m'enchanté ! Bientôt, le grand Canyon, le Colorado et San Francisco... </p>
  <p> En espérant que vous passez de bonnes vacances, affectueusement. </p>
  <signed>Sylvie </signed>
  <signed>François </signed>
</div>
```

First try at a TEI version : the destination

```
<div type="destination">
<ab>
  <stamp type="postmark">
    <placeName>El Paso</placeName> - TX 799 -<date notBefore="1980-07-26">
      <unclear>PM JUL</unclear>
    </date>
  </stamp>
  <stamp type="postage">Profil masculin, avec un avion et un radar au second
    plan: <mentioned>US Airmail 21 c.</mentioned>
  </stamp>
</ab>
<ab>
  <address>
    <addrLine>Madame <name>Lefrère</name>
    </addrLine>
    <addrLine>4, allée George Rouault</addrLine>
    <addrLine>75020 Paris</addrLine>
    <addrLine>France</addrLine>
  </address>
</ab>
</div>
```

Why use TEI (or any other common framework)

- re-usability and repurposing of resources
- modular software development
- lower training costs
- ‘frequently answered questions’ — common technical solutions for different application areas

The TEI was designed to support multiple views of the same resource. The TEI is an evolving model of the concerns of Digital Humanities.

A word on TEI Conformance

A document is *TEI Conformant* if and only if it:

- is a well-formed XML document
- can be validated against a *TEI Schema*, that is, a schema derived from the TEI Guidelines
- conforms to the TEI Abstract Model
- uses the *TEI Namespace* (and other namespaces where relevant) correctly
- is documented by means of a TEI Conformant specification (an *ODD file*) which refers to the TEI Guidelines

Standardization should not mean ‘Do what I do’, but rather
‘Explain what you do in terms I can understand’

For the TEI, that explanation is an ODD

- Experimenting in this way, we can develop a vocabulary which is specific to our project
- but which is also understandable outside our project
- because it is defined and documented according to predefined standards

That's what an ODD is (**One Document Does it all**)

How do you define an ODD ?

You need to

- decide on the elements and attributes you need
- specify their content and values
- document any special usage rules or restrictions

ODD allows you to do this by recycling/modifying the existing TEI definitions.

An ODD for postcards

```
<schemaSpec ident="tei_cartes"
  docLang="fr" start="TEI teiCorpus">
  <moduleRef key="tei"/>
  <moduleRef key="textstructure"
    include="TEI body dateline div postscript salute signed text"/>
  <moduleRef key="core"
    include="add addrLine address bibl date del foreign graphic head hi item lb list name p publisher q reg
resp respStmt street teiCorpus title unclear"/>
  <moduleRef key="header"
    include="teiHeader fileDesc titleStmt publicationStmt sourceDesc"/>
  <moduleRef key="figures"
    include="figure figDesc"/>
  <moduleRef key="msdescription"
    include="stamp"/>
  <moduleRef key="transcr"
    include="att.global.facs"/>
  <moduleRef key="namesdates"
    include="persName placeName"/>
  <elementRef key="ab"/>
<!-- ... -->
</schemaSpec>
```

This is just the formal part of the ODD, which defines the schéma.
The rest of the ODD provides human-readable documentation...

ODD processing

An ODD can be processed to generate

- complete documentation for your TEI application in a variety of formats (HTML, DOCX, PDF...)
- one or more schemas (a formal grammar) which can be used to check and validate all your TEI XML files

These transformations are defined in the TEI Guidelines and in the TEI Stylesheets : they are also implemented

- on the web using Roma, and Byzantium
- at the command line, in a Unix environment
- within oXygen

Sebastian will present ODD in more detail later...

Exploration d'un ODD

- Avec oXygen, ouvrez le fichier cartes-tei.odd qui se trouve dans votre dossier Travaux
- Regardez (et modifiez si vous le souhaitez) les parties documentaires et les parties ODD de ce fichier.
- Générez un schéma et créez un document qui se sert de ce schéma en utilisant oXygen
- Dans le dossier Travaux/Cartes, vous trouverez quelques cartes postales, partiellement transcrrites... à vous de les compléter, en vous servant du schéma !
- Si vous souhaitez modifier le schéma (par ex. pour ajouter une balise pas encore disponible), n'hésitez pas à modifier le ODD et générez à nouveau le schéma
- Quand vous aurez fini, enregistrez votre carte balisé dans le dossier Cartes avec les autres