

TEI à la carte: a case study

Lou Burnard

Où trouver son schéma ?

- Chaque projet est différent; mais chaque projet ressemble à un autre quelconque
- La TEI vous propose
 - choisir tout (déconseillé mais simple)
 - choisir une des selections prédéfinies (TEI Lite, TEI Simple...)
 - choisir l'option de rouler-soi-meme

Un cas d'étude : la carte postale

La TEI – la *texte* encoding initiative – s'applique quotidiennement aux livres et manuscrits, aux corpus littéraires et linguistiques.

Supposons que l'on ait obtenu un financement ANR pour un projet de numérisation et de transcription d'une collection gigantesque de cartes postales. Comment va-t-on créer un système TEI apte à gérer des centaines de choses comme celles-ci ?

Une carte (recto)



Une carte (verso)

SA-146-C

26 juil 80

SAN ANTONIO RIVER
San Antonio, Texas

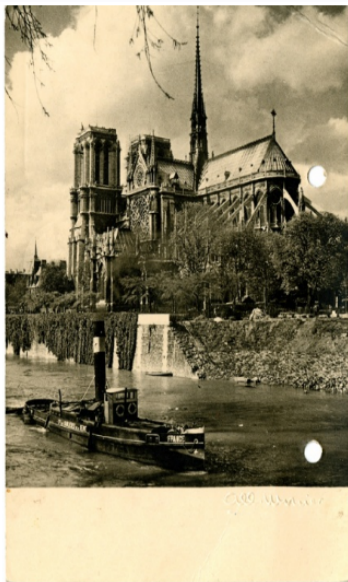
The San Antonio river, often called the "Venice of Texas," winds its way through the business section of San Antonio. It is very picturesque with its many bridges and beautifully landscaped banks.



Chère Madame, après New York
et Washington dont le gigantisme
m'a beaucoup séduit mais
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'en
chanté ! Bientôt, le grand
Canyon, le Colorado et
San Francisco... En espérant
que vous passiez de bonnes
vacances, affectueusement
Sylvie François

Madame Lefrère
4 allée Georges Rouault
75020 PARIS
FRANCE

Encore une carte



2/193 - PARIS - Notre-Dame 26.7.53 - 1944

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 reve-
 littéralement - c'est fou - Je
 t'en supplie laisse moi un peu
 il faut penser à l'avenir - dont
 tu parlais - heureux - riche - utile -
 Je t'aime et - je ne pense qu'à toi Jean

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- **image**
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Quels sont les éléments significatifs qu'il faut absolument distinguer dans notre encodage ?

- image
- oblitération
- partie imprimée
- message(s) écrit(s)
- destinataire(s)
- le(s) sujet(s) représenté(s) sur la carte
- infos sur l'édition, impression, circulation de la carte ...
metadonnées à volonté
- ...

Propositions

Nous commençons en structurant chaque carte en divisions de type variés :

Physiquement :

- **recto** : la surface qui porte l'image (normalement)
- **verso** : la surface qui porte (normalement) le message

Ensuite, nous constatons la possibilité d'avoir indifféremment sur ces deux surfaces d'autres sous-divisions telles que :

- le message
- les marques associés à la communication du document, notamment :
 - le destinataire
 - les oblitérations
- des infos sur l'édition, la vente, la collection etc. de la carte

Première proposition d'encodage

```
<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
cola permet de résister –
l'Amérique m'enchant ! Bientôt, le grand Canyon, le Colorado et San Francisco... En es
<p> Sylvie </p><p>François. </p>
</message>
<destinataire>
Madame Lefrère
4, allée George Rouault
75020 Paris
France
</destinataire>
</verso>
</carte>
```

Commentaire

- On ne s'est pas servi du vocabulaire TEI : donc on aura de mal à expliquer cet encodage à des non-francophones, ou de partager nos travaux avec d'autres.
- On n'a pas mis en évidence tout ce qui pourrait nous intéresser du point de vue patrimonial, par exemple : les ratures, les retours à ligne, les noms de personnes ou de lieux, les aspects linguistiques, les métadonnées...
- On n'a pas structuré (par exemple) le destinataire, ce qui va nous embêter pour faire des recherches efficaces...
- Pour ajouter de telles choses, on doit inventer encore du vocabulaire... ce qui gaspille un peu notre temps, vu que la TEI l'a déjà fait...

Proposition TEI initiale

- On traite chaque carte comme un `<text>` et chaque sous-partie comme un `<div>` composé de `<p>` ou `<ab>`
- On se sert de l'attribut `@type` pour caractériser chaque division
- On va structurer les composants du destinataire en `<address>`, qui regroupe un `<persName>` et des `<addrLine>`
- On aura également besoin de l'attribut `@facs` pour associer une partie de texte avec son image numérisée
- On propose l'élément `<stamp>` pour tout timbrage, oblitération, etc. ; et de l'élément `<bibl>` pour les métadonnées concernant l'édition de la carte avant son usage
- On pourrait distinguer les noms de lieux avec `<placeName>`, les noms de personnes avec `<persName>`, les dates avec `<date>`
- On pourrait se servir de `` (pour les ratures), `<add>` (pour les ajouts), `<reg>` (pour les régularisations), `<hi>` pour les mots ou phrases mises en évidence, `<lb>` pour les retours à la ligne....

Première proposition d'encodage TEI – l'entête

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>San Antonio River : digital edition of card 19800726_001 from the
Virgolos collection</title>
    </titleStmt>
    <publicationStmt>
      <p>Démonstration à une formation TEI</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>
```

Première proposition d'encodage TEI – le texte

```
<text>
  <body>
    <div type="recto">
      <figure>
        <graphic url="../../../Talks/Graphics/Cartes/19800726_001r.jpg"/>
        <figDesc>Vue d'un cours d'eau avec un pont en pierre et de petites
maisons de
style mexicain. Un homme et une femme naviguent un pédalo au
premier
plan.</figDesc>
      <head>San Antonio River</head>
    </figure>
  </div>
  <div facs="../../../Talks/Graphics/Cartes/19800726_001v.jpg"
type="verso">
    <div type="message">
<!-- ... -->
    </div>
    <div type="envoi">
      <ab type="evidences">
<!-- ... -->
      </ab>
      <ab type="destinataire">
<!-- ... -->
      </ab>
    </div>
  </div>
</body>
</text>
```

Première proposition d'encodage TEI – le message

```
<div type="message">
  <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 la Nouvelle Orléans,
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>
```

Première proposition d'encodage TEI – l'envoi

```
<div type="envoi">
  <ab type="evidences">
    <stamp type="cachetPoste">
      <placeName>El Paso</placeName> - TX 799
    -<date when="1983-08-18">18-08-1983</date>
    </stamp>
    <stamp type="timbrePoste">
      <figure>
        <figDesc>Portrait masculin de profil, avec un avion ou un radar au
second
                plan</figDesc>
      </figure>
      <seg>US Airmail 21 c.</seg>
    </stamp>
  </ab>
  <ab type="destinataire">
    <address>
      <addrLine>
        <persName>Madame <surname>Lefrère</surname>
      </persName>
    </addrLine>
    <addrLine>4, allée George Rouault</addrLine>
    <addrLine>75020 Paris</addrLine>
    <addrLine>France</addrLine>
  </address>
</ab>
</div>
```

Pourquoi utiliser la TEI (ou autre standard)?

- faciliter la ré-utilisation des ressources
- réduire les frais de développement et la complexité des outils
- profiter de l'existence des 'frequently answered questions' — des solutions communes ayant des applications diverses

La TEI fut conçue et dessinée pour soutenir plusieurs perspectives sur une même ressource. Elle évolue en modélisant les préoccupations des Humanités numériques.

Etre conforme à la TEI veut dire quoi?

- **être honnêt(e)** : Les éléments XML qui sont déclarés comme appartenant au namespace TEI doivent respecter les définitions TEI de ces éléments
- **être explicite** : Pour valider un document TEI, un ODD est fortement conseillé, parce que cela mettra en évidence toutes les modifications effectuées.

Plus formellement dans un document TEI

- Tout élément appartenant à l'espace de nommage TEI doit être valide par rapport au schéma TEI-ALL.
- Toute autre élément présent doit donc appartenir à un autre namespace

L'objet de ces règles est de faciliter le "blind interchange" des documents. Comme standard, la TEI ne dit pas 'fais ce que moi je fais', mais plutôt 'explique moi ce que tu fais'.



Cette explication s'exprime par un ODD

- On exprime ses besoins/souhaits d'encodage formellement
- Cette expression est (bien sûr) conforme à un standard documentaire prédéfini
- Donc il reste compréhensible, voire utilisable, par d'autres personnes

Voilà l'objectif du système ODD (One Document Does it all)

Traitement d'un ODD

Comme vous avez déjà vu, un ODD est un document XML comme les autres, susceptible donc d'être transformé en

- manuel de projet, en plusieurs formats (HTML, DOCX, PDF...)
- un ou plusieurs schémas pour la validation des fichiers TEI XML files

Les feuilles de style XSLT TEI servent à faire tout les deux. Ils sont disponibles

- sur le web à travers l'interface Roma or Byzantium
- intégrées dans oXygen
- sur ligne de commande (Unix ou Mac)

Exploration d'un ODD TEI

- Avec oXygen, ouvrez le fichier tei-cartes.odd
- Regardez (et modifiez si vous le souhaitez) les parties documentaires de ce fichier.
- Générez un schéma et créez un document qui utilise ce schéma.
- Dans le dossier Cartes, vous trouverez quelques cartes postales, partiellement transcrites... 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