Partizione Antica from Fototeca Supino: a pilot digital edition

Encoding Model

0. Introduction

The *Partizione Antica*, is a particularly interesting part of the wider Supino's Archival fund specialized in monuments and work of art photographs. The whole archive, counting more than 8000 total items, has been used for research and didactical purposes – Supino was the first funder of the History of Art Institute of the University of Bologna.

The *Partizione Antica*, counting about 3000 items, has always been recognized as a peculiar subset, because of its uniformity and its previous provenance: all the photographs are in same-size paperboard framework, and they report handwritten annotations written from its first unknown collector, possibly a British art historian or art collector working at the end of the 19th century.

From the annotations it is possible to extract some relevant information about the collector: visited places and dates, bibliographic reference used, other scholars or contemporary people cited, but also some interesting information about the art collector's world (paid prices of pieces, their movement, etc.).

A pilot Semantic Digital Edition of some photographs has been developed in order to show the possibilities of facilitating the navigation of the whole archive and the annotations on the verso.

To realize this project have been used some tools in support listed below:
- Transkribus: for the handwritten text recognition and the image-text relationship encoding
- Oxygen: to validate tei-xml files, perform xquery, transform file via xslt
- Github: to store material of the project at https://github.com/enri-ca/DTH_PA
- EVT-Edition Visualization Technology tool to offer a sample of the possible visualization.

1. Structure overview

The present model assumes that each photograph is transcribed within a single XML file warranting both possibilities either to study them as a unit or put them together in a structure for overall analysis. Two options have been offered to group the text:

a. the first one, in a `<teiCorpus>` structure to guard the `<teiHeader>` element for any single item (use "01_CreationCorpus_fromCollection.xquery" file and normalizations rules);
b. the second one, to allow the publication via EVT, structured as a more classic TEI file, with several `<text>` elements corresponding to the encoded items (use "05_CreationGroups_title.xquery" file and normalizations rules).

Each text is identified uniquely with an `xml:id=""` attribute recalling the original title in `teiHeader/fileDesc/titleStmt/title[@type='main']`.

The images have been originally encoded with the `<facsimile>` element, with partition of space in `<surface>` and subpartitions in `<graphic>` and `<zone>`; the zone are also carachterized by the attributes @rendition (values `"printspace|Graphic|TextRegion|Line"` and @subtype `"paragraph"` (see file in the … section): nevertheless, for the illustrative purposes of this project, a simpler version of the file, extracted from Transkribus without zone and regions, has been used.

Normalizations rules

- All the `<body>` elements became `<front>` elements
- All the strings
  - `</p>\n<pb xml:id="`
  - `/>\n<pb xml:id="`

  are transformed in:
  - `</p>\n</front>\n<body>\n<pb xml:id="`
  - `/>\n</front>\n<body>\n<pb xml:id="`
- All the .JPG strings are erased
- All the `@n="1"` and `@n="2"` are transformed in `@n="r"` and `@n="v"`
- All the `<bibl>\n<publisher>TRP document creator: enrica.zani@studio.unibo.it</publisher>\n</bibl>` are transformed in: `<bibl>Fototeca Supino https://arti.sba.unibo.it/it/chi-siamo/fototeca-supino</bibl>`
- All the `<publisher>tranScriptorium</publisher>` are transformed in `<publisher>DHDK Unibo project's</publisher>`

1.1. Outline of a single photograph

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="photo_id">
  <teiHeader>
      [metadata]
  </teiHeader>
  <facsimile></facsimile>
  <text xml:id="text_id"></text>
</TEI>
```

1.2. Outline of the collective file

1.2.1 TeiCorpus

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0" version="3.3.0">
    <teiHeader>
        [metadata]
    </teiHeader>
    <TEI xml:id="photo1id">
        <teiHeader>
            [metadata]
        </teiHeader>
        <text xml:id="text1_id"></text>
    </TEI>
    <TEI xml:id="photo2id">
        <teiHeader>
            [metadata]
        </teiHeader>
        <text xml:id="text2_id"></text>
    </TEI>
</teiCorpus>
```

1.2.2. TEI with several texts

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
     <teiHeader>
          [metadata]
     </teiHeader>
     <text xml:id="text1_id"></text>
     <text xml:id="text2_id"></text>
</TEI>
```

1.3. Overview of the <text> element internal structure

Since each photograph is composed by a recto, with the actual image often accompany by a caption, and a verso reporting the actual annotation, each element <text> has been structured as follow:

a. **<front>**: contains the actual image and all the written information completing or commenting it. The text usually consists in a caption or in caption-like information such as the photograph title (depicted work of art description and location), the author (photographer atelier), a reference number to the negative. **The element is required** because it is always valorized with at least the <pb> element completed with the @xml:id attribute which value recalls the image file name.

b. **<body>**: contains the verso's handwritten annotations. **The element is required**

The **<back>** element is never used.

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="photo_id">
 <teiHeader>
    [metadata]
 </teiHeader>
 <facsimile></facsimile>
 <text xml:id="text_id">
     <front>      </front>
     <body>       </body>
 </text>
</TEI>
```

1.4. Manuscripts physical boundaries

1.4.1. Page breaks

```
<pb xml:id="photo_id_r|v" n="r|v"/>
```
The xml:id attribute contains the file name of the image representing the recto/verso of the photograph, without extension. The @n attribute has been valorized with "r|v" values for recto and verso.

1.4.2. Line breaks

```
<lb facs="#facs_1_l_19" n="N001"/>
```
Since the final web presentation could require the association between lines of text and images, each line in the original source has been market by a line break element <lb/> reporting attributes such as facs and n, useful for their connection with the facsimile image.
Due to the peculiar handwritten nature of the original text, where does not occur any

line break in the middle of a word, it has been possible to avoid choices about hyphenation.

2. Structural features

2.1. Photograph items' rectos

The photograph recto has been encoded within the `<front>` element with the image as the `xml:id` attribute of the `<pb>` element and the eventual caption within a `<p>` element with `"caption"` as `@type` attribute value.

```
<front>
     <pb xml:id="photo_id_r" n="r"/>
     <p type="caption"><p>
</front>
```

2.2. Figure and drawing blocks

Hypothetical drawings, findable in other not yet encoded photograph items, should be treated as follow:

```
<figure>
    <graphic url="myfile.jpg"/>
</figure>
```

2.3. Annotations: paragraphs

Paragraphs have been used to groups set of lines having the structure of paragraphs. In the case of a new "hand" have been detected, the paragraph has been characterized by an attribute "late".

```
<p>Here my paragraph</p>
```

2.4. Quotations

A block of text quoted from somewhere else (typically inscriptions), are encoded within a `<q>` element.

3. Editorial markup

The editorial markup encodes features of the text that are the result of an editorial evaluation.

3.1. Abbreviations and relative expansions

Abbreviations `<abbr>` have been encoded together with their expansion `<expan>` forms within the `<choice>` "over"-element.

3.2. Uncertainty

Due to the handwritten text, more than to support or conservative issues, some unclear passages are present. They have been encoded with `<unclear>` element together with the key attribute-value `@reason="illegible"`.

```
<unclear reason="illegible">Ansor</unclear>
```

3.3. Correction, errors, repetitions, missing text

The sample analyzed has not presented this kind of challenges.

This section is open to further indications in the case of a complete treatment of the archive will require it.

4. Metadata

Metadata are encoded within the `<teiHeader>` element. Here his ideal structural overview:

```
<teiHeader>
     <fileDesc>
          <titleStmt></titleStmt>
          <publicationStmt> </publicationStmt>
          <sourceDesc>
                <msDescription></msDescription>
            </sourceDesc>
        </fileDesc>
        <profileDesc></profileDesc>
        <encodingDesc></encodingDesc>
        <xenoData></xenoData>
    </teiHeader>
```

However, the `<profileDesc>`, the `<SourceDesc>` as well as the `<xenoData>` are stored in separate files and, where possible, included thanks to the element `<xi:include/>`. In those cases, the real outline of the `<teiHeader>` will be then:

```
<teiHeader>
     <fileDesc>
          <titleStmt></titleStmt>
          <publicationStmt> </publicationStmt>
          <xi:include href="03_SourceDesc_nor.xml"/>
        </fileDesc>
    <xi:include href="02_ProfileDesc_nor.xml"/>
    <xi:include href="04_xenoData.xml"/>
</teiHeader>
```

4.1. Title information

For the comprehensive file, the title has been as follow completed with the principal researcher involved:

```
     <title type="main">Partizione Antica from Fototeca Supino: a
pilot digital edition</title>
<principal>Zani, Enrica</principal>>
```

For the single photographs, the title and the xml-tei file name have been composed putting together:

 a. Normalized shelfmark e.g. `PA_Grecia I 106`
 b. Inventory number e.g. `sup 2470`
 c. = **`Grecia1_106_sup2470`**

In this way, the alphabetic order corresponds to the succession of the archival folders and files and allow an easier visualization.

```
[TEI/teiHeader/fileDesc/titleStmt/title[@type="main"]]
     <title type="main">Grecia1_106_sup2470</title>
```

4.2. Publication statement

```
[TEI/teiHeader/fileDesc/publicationStmt]
   <publicationStmt>
     <distributor>Github</distributor>
     <date>2022</date>
     <availability>
        <licence target="http://creativecommons.org/licenses/by-
     sa/3.0/">Distributed under a Creative Commons Attribution-
     ShareAlike 3.0 Unported License
        </licence>
     </availability>
   </publicationStmt>
```

4.3 Semantic layer markup

The verso-annotations report full of references to entities. Due to the already done in-depth cataloguing of structured metadata in other projects about artists, works of art, photographers, places and so on, the semantic markup concentrates on "contemporary to the PA creator" people, places, dates, and organizations, hoping this could help to build new structured knowledge to be used to explore their identity.
However, due to the special focus on history of art and photography of the fund, the `<object>` element has been exploited figuring out the future enrichment of the depicted works of art data from the already structured OA records and offering an equal dignity – eventually recalling the F records data - to the photographic items encoded as object as well via the negative number when present.

4.3.1. People, organizations, objects, places

The description of the different entities is embedded within the `<profileDesc>`, in `<particDesc>` or `<settingDesc>` elements.
Persons, organizations, objects and places have been stored as `<person>`, `<org>`, `<object>`, and `<place>` elements in respectively `<listPerson>`, `<listOrg>`, `<listObject>`, and `<listPlace>` elements; each `<person>`, `<org>`,`<object>`, and

`<place>` are uniquely identified by an `xml:id` attribute which value is used, preceded by # - by means of the `@ref` attribute in `<rs type="person|organization|object|place">`, or `<persName>`|`<org>`|`<object>`|`<placeName>` elements within the text - to refer to the entity.

```
<profileDesc>
   <particDesc>
      <listPerson>
           <person xml:id="pers1_id">
                <persName>pers1_name</persName>
           </person>
      <listPerson>
      <listOrg>
         <org xml:id="org1_id ">
                 <orgName>org1_name</orgName>
            </org>
      <listOrg>
      <listObject>
           <object xml:id="obj1_id">
                <objectName>obj1_name</objectName>
           </object>
      <listObject>
   <particDesc>
   <settingDesc>
      <listPlace>
           <place xml:id="place1_id">
              <placeName>place1_name</placeName>
           </place>
      </listPlace>
   </settingDesc>

</profileDesc>
```

4.3.2. Organizations

`<org>` elements are also characterized by `@type` attribute with possible values: `"collection|museum|photographer|collector";` each type has a dedicated sublist.

```
      <listOrg>
           <head>Museums</head>
            <org xml:id="org1_id ">
                  <orgName>org1_name</orgName>
             </org>
             <head>Collections</head>
             <head>Photographer</head>
             <head>Other</head>
      <listOrg>
```

4.3.3. Objects

`<objects>` have a `xml:id` attribute as unique identifier recalling:

a. for works of art catalogued in the Zeri catalogue, the OA record number added with the OA_ prefix e.g.
```
<object xml:id="OA_108716">
    <objectName>HERMES</objectName>
</object>
```

b. for works of art not catalogued in the Zeri catalogue, a conventional but legible string e.g.
```
<object xml:id="Plate3">
   <objectName>Plate III</objectName>
</object>
```

c. for negative of the photograph, a conventional string composed by neg+{photographer'sName}+{negativeNumber} e.g.
```
<object xml:id="negGiraudon_1452">
   <objectName>1452</objectName>
</object>
```

This encoding will be probably useful for enriching the digital edition with information coming from external sources.