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 works of art's photographs. The whole archive, counting more than 8000 total items, has been formed by Iginio Supino - he was the first funder of the History of Art Institute of the University of Bologna - for research and didactical purposes.

Respect to the whole Supino fund, the *Partizione Antica*, counting about 3000 items, has always been recognized as a peculiar subset, because of its aesthetic homogeneity and its previous provenance: all the photographs are albumens, casted 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 which `xml:id` value is "`PA_creator`": 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 through paths offered by the annotations on the verso and by the entities reported.

To realize this project have been used some tools in support listed below:

- [Transkribus](#): for the handwritten text recognition and the image-text encoding
- Oxygen: to validate XML/TEI files and perform Xquery
- Github: to store material of the project at https://github.com/enri-ca/DTH_PA[1]
- [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/TEI 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 item and text are identified uniquely with an `xml:id=""` attribute composed by recalling the original title in `teiHeader/fileDesc/titleStmt/title[@type='main']` and adding a suffix ("`_ph`" or "`_tx`").

The images have been originally encoded with the `<facsimile>` element, with partition of space in `<surface>` element and sub-partitions in `<graphic>` and `<zone>` elements; the zone are also characterized by the attributes `@rendition` (with values

`"printspace|Graphic|TextRegion|Line"`) and `@subtype` (with value `"paragraph"`) as shown in the complete XML/TEI files in the Github repository[2]: nevertheless, for the illustrative purposes of this project, a simpler version of the file, extracted from Transkribus without zone and regions, has been used[3].

Normalizations rules

Some normalizations of the files extracted from Transkribus have been needed in order to have a correct visualization in EVT[4] such as:

- `.JPG` strings are erased
- `@n="1"` and `@n="2"` are transformed in `@n="r"` and `@n="v"`
- `<unclear>` elements have been completed with the attribute-value pair `@reason="illegible`
- `<quote>` elements have been transformed in `<q>`

Other interventions are needed to structure the `<div>` blocks as explained in 2. Structural features (1.2.1 and 1.2.2).

1.1. Outline of a single `<TEI>` photograph

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

1.2. Outline of the collective files

1.2.1 `<teiCorpus>` with several `<TEI>`

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0" version="3.3.0">
    <teiHeader>
         [metadata]
    </teiHeader>
    <TEI xml:id="id_ph1">
         <teiHeader>
              [metadata]
         </teiHeader>
         <text xml:id="id_tx1"></text>
    </TEI>
    <TEI xml:id="id_ph2">
         <teiHeader>
```

```
            [metadata]
        </teiHeader>
        <text xml:id="id_tx2"></text>
    </TEI>
</teiCorpus>
```

1.2.2 `<TEI>` with several `<text>`s

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
    <teiHeader>
        [metadata]
    </teiHeader>
    <text xml:id="id_tx1"></text>
    <text xml:id="id_tx2"></text>
</TEI>
```

1.3. Physical boundaries

1.4.1. Recto-verso breaks `<pb>`

```
<pb xml:id="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>`

```
<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. `<div>` blocks

`<div>` blocks have been used to structure the different blocks of information carried by the photographs. Inevitably, logical and physical structure influence each other reciprocally, so it is common to have an image and possibly a caption or other typed information on the recto, and handwritten annotations on the verso.

2.1.1 Photo rectos: `<div>` `@type="recto"` and `@type="caption"`

The photograph recto, whenever it brings some texts, has been encoded within a `<div>` element with "`recto`" as `@type` attribute value; the eventual caption has been encoded within a second level `<div>` element with "`caption`" as `@type` attribute value.

```
<pb xml:id="id_r" n="r"/>
<div type="recto">
      <div type="caption"><div>
</div>
```

2.1.2 Photo versos: `<div type="annotation">` and `<p hand="late">`

The photograph verso has been characterized with a `<div>` element with `"annotation"` as `@type` attribute value. Within this `<div>` element, `<head>` and `<p>` elements have been used.
In the case a different medium of writing has been detected, the `<p>` element has been completed with the `@hand="late"`.

```
<pb xml:id="id_v" n="v"/>
<div type="annotation">
      <head></head>
      <p></p>
      <p hand="late"></p>
</div>
```

2.2. Quotations `<q>`

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

2.3. Figure and drawing blocks `<figure>`

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

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

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>
                  <listBibl></listBibl>
                  <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>
      <date>2022</date>
      <availability>
         <licence target="http://creativecommons.org/licenses/by-
         sa/3.0/">Project distributed under a Creative Commons
         Attribution-ShareAlike 3.0 License. All the images and the data
         are published for illustrative and didactical purposes only. For
         any request of use, please refer to: Biblioteca delle Arti.
         Sezione di Arti visive "I. B. Supino" (images) and Fondazione
         "Federico Zeri" (OA and F records).
         </licence>
      </availability>
   </publicationStmt>
```

4.3. The `<msDesc>` element

The <msDesc> element is structured as follow:

```
[TEI/teiHeader/fileDesc/SourceDesc/msDesc]
 <msDesc>
      <msIdentifier></msIdentifier>
      <msContents></msContents>
      <physDesc></physDesc>
      <history></history>
 </msDesc>
```

4.3.1. The `<msIdentifier>` element

```
[TEI/teiHeader/fileDesc/SourceDesc/msDesc/msIdentifier]
   <msIdentifier>
        <settlement>Bologna</settlement>
        <repository>Biblioteca delle Arti. Sezione di Arti visive "I.
        B. Supino". Alma Mater Studiorum Università di
        Bologna</repository>
   </msIdentifier>
```

In the case of the single `<TEI>` file, the `<msIdentifier>` element could report the `<idno>` element with call number or inventory identifier from F records.

4.3.2. The `<msContent>` element

```
[TEI/teiHeader/fileDesc/SourceDesc/msDesc/msContents]
   <msContents>
      <msItem>
            <author>Supino, Igino Benvenuto</author>
            <title>Fototeca Supino - Partizione Antica</title>
      </msItem>
   </msContents>
```

In the case of the single `<TEI>` file, the `<author>` and `<title>` elements should report the single photograph metadata from F records.

4.3.3. Physical description

```
[TEI/teiHeader/fileDesc/SourceDesc/msDesc/physDesc]
 <physDesc>
     <objectDesc>
           <p>3260 albumen photographs collected and handritten
           annotated on the verso by an unknown Art Historian or
           Collector in the second half of the XIXth century.
           </p>
      </objectDesc>
   </physDesc>
```

In the case of the single `<TEI>` file, the `<pysDesc>` element should contain higher structured description reporting the single photograph metadata from F records such as `<support>`, `<extent>`, `<dimensions>` elements.

4.3.4. Provenance/history

The element `<history>` provide a mean to date the manuscript and to describe its provenance up to the present repository. What follow is a possible example of completion and should be revised.

```
[TEI/teiHeader/fileDesc/SourceDesc/msDesc/history]
   <history>
      <p>For their peculiarity in the overall Supino fund, the 3260
      photographs in the series called Partizione Antica, were
      transferred in 1990 to the Historical Archive of the University
      of Bologna, reorganized on a topographical basis, digitized, and
      cataloged with the Sebina Multimedia software. In 2006 they
      rejoined the rest of the Supino Fund at the Arts Department.</p>
      <p>Between 2019 and 2020 the entire Partizione Antica series was
      cataloged by the Federico Zeri Foundation.</p>
      <p>A sample of 14 recto-verso photographs, is the object of this
      pilot digital edition.</p>
</history>
```

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 and F records.

4.3.1. People, organizations, objects, places

```
[TEI/teiHeader/profileDesc/particDesc|settingDesc]
```
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.

```
[TEI/teiHeader/profileDesc/particDesc|settingDesc]
<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.

```
[TEI/teiHeader/profileDesc/particDesc/listOrg]
      <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.

```
[TEI/teiHeader/profileDesc/particDesc/listObject/object]
<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.

```
[TEI/teiHeader/profileDesc/particDesc/listObject/object]
<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.

```
[TEI/teiHeader/profileDesc/particDesc/listObject/object]
<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.

4.3.4. Bibliographic references

The bibliographic references used by the "PA_creator" to document its annotations have been encoded within a `<bibl>` element recalling by the value of the `@source` attribute, the `@xml:id` of the `<bibl>` element listed in the `<listBibl>` element

```
[TEI/teiHeader/fileDesc/SourceDesc/listBibl]
    <listBibl>
        <bibl xml:id="bibl1_id">
        </bibl>
    <listBibl>
```

5. `<xenoData>`

`<xenoData>` element has been added to allow the enhancement of the digital edition with entities linked open data in rdf.
Predicates and authorities used for the rdf statements, comes from the following namespaces declared in the `<TEI>` element:

```
xmlns:dbo="http://dbpedia.org/ontology/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:geonames="https://www.geonames.org/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:viaf="http://viaf.org/viaf/"
xmlns:wikidata="https://www.wikidata.org/wiki/"
```

Authorities have been used specially for:
- Wikidata: for monuments, works of art or other uncovered entities
- Geonames: for places
- VIAF: for persons and photographer

In addition, for bibliographic entities, oclc id has been added when possible.