Extending the DSE: LOD support and TEI/IIIF integration in EVT

Paolo Monella
Università di Palermo / Venice Centre for Digital and Public Humanities

Roberto Rosselli Del Turco
Università di Torino

EVT Home: http://evt.labcd.unipi.it/
GitHub: https://github.com/evt-project
Outline

● Introduction
  ○ The birth and development of EVT
  ○ Expanding the DSE: The distributed edition
● LOD (Linked Open Data)
  ○ LOD support in EVT, current and forthcoming
  ○ TEI/LOD integration
● IIIF (International Image Interoperability Framework)
  ○ IIIF support in EVT, current and forthcoming
  ○ TEI/IIIF integration
● The distributed edition via Zenodo and GitHub
The birth and development of EVT

- EVT (Edition Visualization Technology) was born as the edition viewer component of the Digital Vercelli Book project

- Developed to be a general purpose tool, many new features implemented thanks to collaboration with other projects
  - **EVT 1**: a diplomatic edition / single-witness critical edition viewer based on XSLT 2 and Web technologies [final version EVT 1.3 released December 2019]
  - **EVT 2**: a “reboot” of the project based on the MVC design pattern and the AngularJS framework [beta 2 due Spring 2020]
  - **EVT 2+**: porting to Angular 8 under way [don't ask]

- Modular design to improve flexibility and power
- UI/X designed (and tested) to maximize ease of use
Expanding the DSE

- A DSE is a **dynamic object**, a research tool which assists the scholar in data interpretation and analysis.
- This also means that it can engage in dialogue and interaction with other Internet-based resources by:
  - taking advantage of **semantic web** technologies and **Linked Open Data** to enrich the edition content.
  - **text/image** linking, pointing to digital collections of images of manuscripts maintained by external repositories.
  - modelling **intertextual relationships** through canonical text services.
  - becoming itself a **source** of shareable, reusable data.
The distributed edition

● This is an EVT goal dating back quite some time

● It is only recently that all the necessary technologies seems to be falling into place
  ○ LOD repositories
  ○ IIIF framework for images
  ○ CTS and DTS protocols for text
  ○ Zenodo as a safe, long term general purpose repository
  ○ GitHub as a server of live pages

● To use those effectively we need to work both on the general methodology and to experiment with the existing tools

● Great opportunities, but also some risks and complications
LOD
Linked Open Data
Guido da Velate is a archbishop.
Guido da Velate
http://dbpedia.org/resource/Guido_da_Velate

is a
http://purl.org/linguistics/gold/hypernym

archbishop
http://dbpedia.org/resource/Archbishop
LOD

F. Lynam, from Vimeo video Linked Open Data in practice
Using LOD / Creating LOD

- **Using** (referencing) external resources

  <persName ref="http://dbpedia.org/resource/Guido_da_Velate">
  Guidonem</persName>

  Guido da Velate
  http://dbpedia.org/resource/Guido_da_Velate

  - http://dbpedia.org/ontology/deathDate
    1071-8-23
  - http://dbpedia.org/ontology/deathPlace
  - http://dbpedia.org/resource/Alessandria
  - www.w3.org/1999/02/22-rdf-syntax-ns#type
  - http://yago-knowledge.org/resource/clergy/man_109927451
Using LOD / Creating LOD

- **Creating** new LOD resources

```xml
<persName ref="http://dbpedia.org/resource/Guido_da_Velate">Guido da Velate</persName>
The work I published
http://myproject.it/myeditionuri
refers to
www.ancientwisdoms.ac.uk/media/ontology/sawsOntology.owl#refersTo
Guido da Velate
http://dbpedia.org/resource/Guido_da_Velate"
```
Using LOD / Creating LOD

- **Creating** new LOD resources

```
<persName ref="http://dbpedia.org/resource/Guido_da_Velate">Guidonem</persName>
```

“Guidonem”

is an attested name for

[http://xmlns.com/foaf/0.1/name](http://xmlns.com/foaf/0.1/name)

Guido da Velate

Using LOD / Creating LOD

- **Using** (referencing) LOD resources
- **Creating** new LOD resources
<teiHeader>

<listPerson>

<person xml:id="Rothari">

<persName>Rotari</persName>

<sex>M</sex>

<birth>606</birth>

<death>652</death> [...]

</person>

</listPerson>

</teiHeader>

<body>

<persName ref="#Rothari">

<w>rothari</w></persName>

</body>
Using LOD: EVT support 1 (via local list)

<teiHeader>

<listPerson>
<person xml:id="Rothari">
<persName>Rotari</persName>
<sex>M</sex>
<birth>606</birth>
<death>652</death>
</person>
</listPerson>

<body>
<persName ref="#Rothari">
<w>rothari</w>
</persName>

<listLoc>

1. IN NOMINE DAI INCIPIT
2. EDICTVM QVEM RE
3. NOVAUIT DOM ROTH
4. RAIUIR EXCELLENTISSIMO
5. REX GENTI LANGOBAR
6. DORVM CVM PRIMA
7. TOS IUDICES SVOS

8. INCIP PROLOGVS

9. Ego in di nomine rothari

10. Vir exell, rex

11. Des septimo decimù rex geras langobar

12. Doru annuente do propitiante regni mer

</listLoc>
</body>
Using LOD: EVT support 1 (via local list)

<teiHeader>
  <listPerson>
    <person xml:id="Rothari">
      <persName>Rotari</persName>
      <sex>M</sex>
      <birth>606</birth>
      <death>652</death>
      <note>V. anche la voce corrispondente su
      <ref target="http://dbpedia.org/resource/Rothari">DBpedia</ref>.</note>
    </person>
  </listPerson>
</teiHeader>

<body>
  <persName ref="#Rothari">rothari</persName>
</body>
Using LOD: EVT support 1 (via local list)

<teiHeader>

<listPerson>
  <person xml:id="Rothari">
    <persName>Rotari</persName>
    <sex>M</sex>
    <birth>606</birth>
    <death>652</death>
    <note>V. anche la voce corrispondente su
    <ref target="http://dbpedia.org/resource/Rothari">DBpedia</ref>.</note>
  </person>
</listPerson>

<body>
  <persName ref="#Rothari">
    <w>rothari</w>
  </persName>
</body>
Using LOD: EVT support 1 (via local list)

```
<teiHeader>
  <listPerson>
    <person xml:id="Rothari">
      <note>See <ref target="http://dbpedia.org/resource/Rothari">DBpedia</ref> for further information.</note>
    </person>
  </listPerson>

<body>
  <persName ref="#Rothari">
    <w>rothari</w>
  </persName>
```
Using LOD: EVT support 1 (via local list)

```
<teiHeader>

<person xml:id="Rothari">
  <ptr target="http://dbpedia.org/resource/Rothari"/>
  <note>
  </note>
</person>

<body>

<persName ref="#Rothari">
  <w>rothari</w>
</persName>
```

Partial EVT support (showing URL)
Using LOD: EVT support  2 (direct link)

EVT support forthcoming
(currently showing empty info box)

<persName ref="http://dbpedia.org/resource/Rothari">
  <w>rothari</w>
</persName>
Using LOD: EVT support 2 (direct link)

<teiHeader>

<interpGrp type="rhetoric">

<interp

sameAs="https://www.wikidata.org/entity/Q486050">
Anafora
</interp>
[...]

</interpGrp>

<body>

<seg type="r" ana="#anafora">Quotiens</seg>

</body>
### Using LOD: EVT support 1 (local list) vs 2 (direct link)

<table>
<thead>
<tr>
<th></th>
<th>Distributed (LOD)</th>
<th>Durable</th>
<th>Oxygen support</th>
<th>EVT support</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Local list</strong></td>
<td>(redundant)</td>
<td>(dead URIs ahead?)</td>
<td>(suggested values → persName @ref)</td>
<td>(info box popping up)</td>
</tr>
<tr>
<td><code>&lt;listPerson&gt; / &lt;person&gt;</code></td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Direct link to LOD only</strong></td>
<td>✓</td>
<td>?</td>
<td>✗</td>
<td>✗</td>
</tr>
<tr>
<td><code>&lt;persName ref=&quot;http://dbpedia...&quot;&gt;</code></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Creating LOD

- Creating new LOD resources (triples) → External SW needed

Edictum Rothari
http://leges.labcd.unipi.it

refers to
www.ancientwisdoms.ac.uk/media/ontology/sawsOntology.owl#refersTo

Rothari
http://dbpedia.org/page/Rothari
LOD support in EVT

- **Current**
  - Encoder creates **local list** (e.g. `<listPerson>`)
  - Encoder **links** local element to LOD: `<person><ref="http://dbpedia.org/resource/Rothari">`)
  - EVT visualizes **info box from** `<person>`
  - EVT shows **clickable link** to LOD resource
- **Project’s own software can extract** RDF triples from XML
- **Forthcoming**
  - Support **direct link**: text → LOD resource (no local list)
  - **Import info from LOD** resource and visualize in info box
  - **Extract** RDF triples from XML (?)
TEI/LOD Integration

- Current TEI/LOD integration
  - Linking from TEI to LOD resources
    - Via local data structure (<listPerson>, <listPlace>, etc.): minimal overhead
    - Or directly from the text (e.g. <persName>): closer to LOD philosophy, but burdensome if the same link occurs multiple times

- Issues in TEI/LOD integration
  - TEI is flexible, but it currently is document-oriented more than LOD-oriented
    - Re-orienting the TEI Guidelines and Schema (Chiarcos-Ionov 2019)
    - E.g.: allowing for (LOD) external URIs in some attributes
  - Tools needed:
    - Editing: Oxygen and other authoring tools could facilitate LOD linking
    - Visualization: EVT could retrieve and show information retrieved directly from DBpedia displaying it (in part or fully) in a dedicated container
    - Tools to extract and expose semantic information (LOD triples) from TEI
IIIF
International Image Interoperability Framework
IIIF: what and why

- **IIIF**: International Image Interoperability Framework
  - Exchange and integrate image-based resources in Web-based systems

- Use cases:
  - **IIIF image server** ([e-codices](https://e-codices.ch), Biblioteca Vaticana, Biblioteca Ambrosiana)
  - **IIIF collection** (MS images: [Sang. 730](https://e-codices.ch), Edictum Rothari)
  - **IIIF canvas** (each image / MS page: folio 20)
  - TEI linking to the URI of the external IIIF image (e.g. `<pb facs="IIIF_URI_here">`)
IIIF metadata

- Whole collection / MS (Sang. 730)
  - [Link](https://www.e-codices.unifr.ch/metadata/iiif/csg-0730/manifest.json)
  - Presentation API
- Individual image / page (folio 20)
  - [Link](https://www.e-codices.unifr.ch/metadata/iiif/csg-0730/canvas/csg-0730_020.json)
  - Image API
- JSON metadata files reference actual image files:

```json
manifest.json

service:
  @context: "http://iiif.io/api/image/2/context.json"
  @id: "https://www.e-codices.unifr.ch/oris/csg/csg-0730a/csg-0730a_020.jp2"
  profile: "http://iiif.io/api/image/2/level2.json"
```
IIIF Image API by examples

- [https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2](https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2)  
  Image file URI
IIIF Image API by examples

- [Link](https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg)

- whole image (no region cropping)
- full size (no resizing)
- 0° rotation
- default quality (color, gray...)
- jpg format (tif, png...)
IIIF Image API by examples


whole image (no region cropping)
IIIF Image API by examples

- https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/1400,880,2850,600/full/0/default.jpg
Two directions for TEI/IIIF integration

1. **IIIF → TEI**: adding a transcription to an IIIF-based digital facsimile
   
   IIIF manifest.json → TEI XML transcription file (details)

2. **TEI XML → IIIF URI**: embedding IIIF URIs in a TEI-based diplomatic transcription

```xml
<pb facs="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg"/>
```
Two directions for TEI/IIIF integration

1. IIIF → TEI

IIIF manifest.json → TEI XML transcription file (details)

Manifest

```
34: (-)
35: 
  @id: "http://213.21.172.53/canvas/c06027113680a8073"
  @type: "sc:Canvas"
  label: "036_D23sup_r.13v"
  height: 2960
  width: 1480
  images: [-]
  otherContent:

      0:

  @id: "http://213.21.172.53/manifests/public/list/0b6027113680a8073/06027113680a8073.jsen"

  @type: "sc:AnnotationList"
  thumbnail: [-]
  36: (-)
  37: (-)
  56: (-)
```

TEI XML

annotationList

```
"resource": {
  "@id": "http://www1.unipa.it/paolo.monella/reires2019/code/d23sup13r/transcription.xml",
  "@type": "dctypes:Text",
  "format": "application/tei+xml"
}
```
Two directions for TEI/IIIF integration

1. **IIIF → TEI**
   
   **IIIF manifest.json → TEI XML transcription file** (details)

2. **TEI XML file → IIIF URI**

   ```xml
   <pb facs="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg"/>
   ```
Two directions for TEI/IIIF integration

1. IIIF → TEI
   IIIF manifest.json → TEI XML transcription file (details)

2. TEI XML file → IIIF URI
   <pb facets="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg"/>
IIIF support in EVT

Page ➔ Local image

<data:image src="data/images/single/VB_fol_104v.jpg"/>
IIIF support in EVT

Page → IIIF full image

<pb facs="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg"/>
IIIF support in EVT

Page → IIIF image region

<pb facs="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/800,600,3800,5000/full/0/default/jpg"/>
IIIF support in EVT

Region (&lt;lb&gt;, &lt;p&gt;, &lt;div&gt;...) → &lt;zone&gt; (region) → IIIF full image

&lt;surface&gt;
  &lt;zone lrx="1200" lry="423" ulx="360" uly="180"
    xml:id="SG730_hs_20_01"/&gt;
&lt;/surface&gt;

&lt;pb facs="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg"/&gt;
&lt;div facs="#SG730_hs_20_01"&gt;
IIIF support in EVT

Region (<lb>, <p>, <div>...) → <zone> (region) → IIIF full image

<surface>
<zone lrx="1200" lry="423" ulx="360" uly="180"
xml:id="SG730_hs_20_01"/>
</surface>

**Issue:** Tool needed to draw rectangles and get coordinates

- Local image: Ch. Sparks, TEI zoner (see 1, 2, 3)
- Local and IIIF image: Oxygen Image Maps in TEI
- IIIF image (under development): T. Mancinelli, Barberino generator
IIIF/EVT future development

Page → <surface> → IIIF full image

<surface xml:id="image-p2">
<graphic url="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg"/>
</surface>

<pb facs="#image-p2"/>
IIIF/EVT future development

Region → IIIF image region

● EVT can already import image regions through IIIF Image API URIs like this
● However, visualization strategies must be implemented to show them (hotspot, pop-up window within text, text-embedded image, etc.)
IIIF/EVT future development

- EVT2+ (Angular 8 version, in development)
- Presentation API support
- EVT config.json configuration file

```json
{"title": "My Digital Edition", "manifestURL": "https://www.e-codices.unifr.ch/metadata/iiif/csg-0730/manifest.json" }
```
TEI/IIIF integration

- Tool needed to define area coordinates in IIIF-served images
  - A variant of **TEI Zoner** (T. Mancinelli’s **Barberino generator**)
  - Oxygen in Author mode (**Image Map Editor**)
- Concurrent mechanisms to define image portions
  - TEI `<zone lrx="1200" lry="423" ulx="360" uly="180"/>`
- IIIF calculates coords differently than TEI
  - TEI: both rectangles (coords for upper left and lower right, ulx, uly, lrx, lry attributes) and polygons (points attribute)
  - IIIF: only rectangles (coords for upper left vertex, then rectangle dimensions)
- Tool needed to convert from IIIF format into TEI
The distributed edition
The distributed edition

● A concept dating back to the end of the ‘90s

● A view of DSEs not only taking advantage of LOD, but becoming resources on their own

● To do this several requisites must be met
  ○ suitable repositories
  ○ long term sustainability
  ○ data sharing methodology
  ○ clear editorial responsibility

● We now have suitable frameworks ([Zenodo](https://zenodo.org), [OpenAIRE](https://www.openaire.eu), [GitHub](https://github.com)) and sensible dissemination criteria ([FAIR Principles](https://fairdata.org))
“Good things come in small packages”

- A spin-off of the Visionary Cross project and a consequence of its “data-first” approach
  - The Visionary Cross project on Zenodo: [https://zenodo.org/communities/the_visionary_cross/](https://zenodo.org/communities/the_visionary_cross/)

- General plan: making all the edition data available for other researchers to use as they see fit (no intermediation)

- Not a LOD approach, strictly speaking, but one that fits well with the open, distributed edition concept

- It will require work and experimentation (also to make EVT suitable for the task)
Project approach

● Flexible
  ○ direct access to data (also thanks to descriptive metadata)
  ○ access single / group of objects

● Extensible
  ○ add, rearrange, or reuse material without negotiation

● Authoritative
  ○ preserve credit/responsibility for all contributions

● Durable
  ○ permanently discoverable and available
  ○ low/no maintenance
Zenodo record

- Metadata record + file(s)
- RDF (internal and external relations)
- Two kinds of DOIs:
  - “Conceptual” (latest)
  - “Version” (current)
- RESTful files URLs
  - No link rot
An edition is built around records
Conclusion
Conclusion

- **Our approach**
  - having a look at standards, frameworks, protocols currently available (LOD, IIIF) and see how they are supposed to work
  - direct experimentation with available tools (TEI, EVT)

- **The good**
  - the networked / distributed DSE is nearly there
  - current frameworks are sound and work well

- **The bad**
  - got to be careful in choosing the best methodological approach

- **The ugly**
  - when TEI and IIIF/LOD speak different languages a “translator” is needed (i.e. new tools must be implemented)
Credits

- **EVT2js** is being developed by
  - Giulia Cacioli
  - Chiara Di Pietro
  - Angelo Mario Del Grosso
  - Federica Spinelli

- The porting to Angular8 (**EVT2+**) is being done by
  - Giulia Cacioli
  - Chiara Di Pietro
  - Giacomo Cerretini
  - Guglielmo Lischi
  - Federica Spinelli
  - Simone Zenzaro

- Project directed by R. Rosselli Del Turco and Chiara Di Pietro

- Thanks to all former developers!
  - Chiara De Martin
  - Julia Kenny
  - Raffaele Masotti
  - Greta Musu
  - Jacopo Pugliese
  - Rocco Russo
  - Ilaria Tiezzi
Thank you for your attention!