Introduction

- 3 years project to build a LOD vocabulary and platform, started in 2017
- Migration of TBRC’s dataset (now native RDF, under CC0 license):
  - 7k places
  - 17k persons
  - 27k works (371k texts, 72k volumes, 15M images)
  - 13k unicode texts (formerly TEI files)
- Handle Sanskrit, Tibetan, Chinese and Pali domains
- Extension of vocabulary to cover the needs of our partners in:
  - Philology
  - Lexicography
  - Codicology
- Web annotations (on text, RDF, images):
  - provenance information
  - Crowdsourcing

Web Annotations

Knowledge graph (RDF)

<table>
<thead>
<tr>
<th>bdr:AS001</th>
<th>a</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td>&quot;some person&quot;@en</td>
<td></td>
</tr>
<tr>
<td>hasDescription</td>
<td>bdr:ETXT001</td>
<td></td>
</tr>
</tbody>
</table>

Statement (reified RDF)

<table>
<thead>
<tr>
<th>bdr:ST001</th>
<th>a</th>
<th>rdf:Statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>hasBody</td>
<td>bdr:ETXT001#chars=826-850</td>
<td></td>
</tr>
<tr>
<td>hasEditor</td>
<td>bdr:P32134</td>
<td></td>
</tr>
<tr>
<td>hasIIIFManifest</td>
<td><a href="http://iiif.bdrc.io/578b2-655/1910_01_1923.300/180">http://iiif.bdrc.io/578b2-655/1910_01_1923.300/180</a></td>
<td></td>
</tr>
<tr>
<td>hasTranslator</td>
<td>bdr:P3214</td>
<td></td>
</tr>
<tr>
<td>hasUnicodeText</td>
<td>bdr:ETXT001</td>
<td></td>
</tr>
</tbody>
</table>

Unicode text (bdr:ETXT001)

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

None of our partners have

Conclusion and perspectives

- Migrating our legacy data to native RDF forced us to create a very complete vocabulary to record all details, not just a vocabulary for data exchange
- We hope our vocabulary (the first in the domain) will become a reference and that we can build a working group / review committee to expand
- No existing gazetteer in the domain, so we need to allocate resources for entity matching with our partners, Wikidata, VIAF, Worldcat, etc.
- Many projects interested but only few have the technical resources to jump in
- LOD lexicography (ontoles, etc.) requires important resources we're not sure our partners have
- Once the tools are built, we hope users can create annotations proposing new information, fixes, discussions, and improve the data

Acknowledgments

The BUDA platform
LOD platform and model for Buddhist Studies: Bibliography, Prosopography, Geography

Technical Contact: roux.elie@gmail.com