New Frontier:
Using Wikidata on Commons

By Jarek Tuszyński
(user:jarekt)
October 2016
Infobox templates on Commons

• Purpose:
  – **Preserve metadata** - Many files come from institutions which store rich metadata. When uploading files to Commons we want to preserve it all.
  – **Internationalization (I18n)** - Commons serves Wikimedia projects using many languages. Our goal is to show pages in the language of the viewer.
  – **Machine readability**

• Metadata is stored in variety of templates, for example
  – {{Artwork}} – for artworks, artifacts and other objects
  – {{Creator}} – for authors
  – {{Institution}} – for galleries, museums and other institutions
  – We also have {{Book}}, {{Authority control}}, {{Location}} (for coordinates) and many more
Preserving metadata during upload
Internationalization

• The same page as seen by English and German speakers
• Language determined from user preferences or from Wikipedia they come from
I18n templates

{{{Artwork}}}  {{{Creator}}}  {{{ISOdate}}}  {{{City}}}  

{{{NationAndOccupation}}}  

{{{Authority Control}}}  

{{{Technique}}}  

{{{Institution}}}  

{{{location}}}  

Translated using {{LangSwitch}} template

Translated on Translatewiki.net
Trouble brewing

• The system become very esoteric, with some templates becoming too complicated to maintain.
• Many of i18n templates are used on a lot of pages, like {{Dir}} used on 41M out of 47M pages.
• We were often tripping the expansion limit “safety fuse”
• We were rescued by two 2013 developments:
  – Introduction of Lua programing language
  – Introduction of Wikidata
Timeline

• **January 2013** - Wikidata phase I went live on Wikipedia projects allowing storage of interwiki links in a single place as Wikidata "sitelinks".

• **March 2013** - Lua is introduced on all Wikimedia projects, including Commons. Many templates were rewritten in Lua.

• **September 2013** - Wikidata phase I went live on Commons, with little notice since most interwiki links on Commons were from Commons categories to Wikipedia articles and back then there was a semi-consensus that sitelinks should not be between pages in different namespaces.

• **September 2015** - [Arbitrary access](#) to Wikidata properties was enabled on English Wikipedia allowing calls like
{#property:P36|from=Q183}}

• **April 2016** - [Arbitrary access](#) was enabled on Commons, which finally made Wikidata useful to Commons.
New Frontier

• Wikidata can simplify two of the purposes of infoboxes:
  – **Preserve metadata** – Much of metadata about people, places, institutions, books, artworks, etc. can be stored on Wikidata, which serves as central repository shared by all the projects.
    • Benefit for Wikidata – there is a wealth of metadata stored on Commons
    • Benefit to Commons - more users will bring better quality and simplify maintenance
  – **Internationalization** – Wikidata can be also used by codes translating terms that have associated Wikidata q-codes

• Since April much work was done on rewriting some of the infobox templates using Lua and accessing Wikidata.
Stages of moving metadata to Wikidata

1. Link Commons pages with Wikidata
   - Store Commons page name in Wikidata properties
   - Store Wikidata q-code on Commons page

2. Rewrite Infobox in Lua to allow pulling metadata from Wikidata

3. Reconcile Commons and Wikidata metadata, through automatic comparison:
   - The same → remove from Commons, as redundant
   - Missing from Wikidata → move to Wikidata
   - Present in both but different → reconcile by human
Rewriting Templates in Lua

• Commons templates often rely on a lot of other templates, which again rely on even more templates
• One challenge of moving to Lua is to minimize number of calls to other templates
• Luckily many of the most basic templates, are already translated
## Progress

<table>
<thead>
<tr>
<th>Template</th>
<th>Lua</th>
<th>Wikidata link</th>
<th>Reconcile</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>{{City}}</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>{{Authority control}}</td>
<td>100%</td>
<td>138k (98%)</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>{{Creator}}</td>
<td>80%</td>
<td>22k (90%)</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>{{Institution}}</td>
<td>0%</td>
<td>4k (97%)</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>{{Artwork}}</td>
<td>5%</td>
<td>60k (6%)</td>
<td>0%</td>
<td><em>Sum of all Paintings</em> project added metadata for 148k paintings and 11k sculptures</td>
</tr>
<tr>
<td>{{Book}}</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>{{Object Location}}</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>
{{City}} template

• {{City}} template serves as a giant lookup table matching 1890 names of cities and countries, like “Москва” or “Moscow”, to 1315 location templates, like {{Moscow}}.

• Location templates return name of location in user’s language with a link to user’s native Wikipedia. All names and links for hundreds of languages were stored in each template.

• We matched all location templates to Wikidata and rewrote them so they pull all the information required directly from Wikidata.

• {{City}} still works as a lookup table but also calls Wikidata directly.
{{Authority control}}

- Used 140k times on some Category pages and most Creator and Institution pages
- We matched most instances to Wikidata and modified each template by adding “Wikidata” parameter
- Pages using both local and Wikidata parameters use color to identify the source
- We reconciled and removed local parameters from 138k pages. 3k remains unmatched.

Matching values

| Authority control |
| VIAF: 39447335 | ISNI: 0000 0001 0889 7770 | Botanist: P. Karsten/P.Karst. (Q31828 P428 “P. Karsten” S143 Q565) |
| GND: 127704353 | SUDOC: 064959767 | BNF: 12322463v |

Commons value Wikidata value

Code which can be used with quick statements tool to add Commons value to Wikidata
{{Creator}}

- {{Creator}} template allows creation of a *creator template* with basic information about people. We have almost 25k of them.

- We matched 90% of them to Wikidata and wrote Lua code that closely mimics current template:

  Local parameters:

  ![Paul Cézanne (1839–1906)](image)

<table>
<thead>
<tr>
<th>Description</th>
<th>French painter and engraver</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth/death</td>
<td>19 January 1839</td>
</tr>
<tr>
<td>Location of birth/death</td>
<td>Aix-en-Provence</td>
</tr>
<tr>
<td>Work location</td>
<td>Paris, Auvers-sur-Oise, Aix-en-Provence, Marseille</td>
</tr>
<tr>
<td>Authority control</td>
<td>• VIAF: 3937403620 • ISNI: 0000 0001 2128 737970 • ULAN: 5000047930 • LCCN: n79055446 • NLA: 3502628689 • WorldCat</td>
</tr>
</tbody>
</table>

  Wikidata based infobox:

  ![Paul Cézanne (1839–1906)](image)

<table>
<thead>
<tr>
<th>Description</th>
<th>French painter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of birth/death</td>
<td>19 January 1839</td>
</tr>
<tr>
<td>Location of birth/death</td>
<td>Aix-en-Provence</td>
</tr>
<tr>
<td>Authority control</td>
<td>• VIAF: 3937403620 • ISNI: 0000 0001 2128 737970 • ULAN: 5000047930 • LCCN: n79055446 • NLA: 3502628689 • WorldCat</td>
</tr>
</tbody>
</table>
Commons / Wikidata Linking (current)

- Sitelinks work for 1-to-1 linking with Wikipedia
- Commons Can have several pages that need to access Wikidata properties: category, creator, institution or gallery pages, but only one sitelink is allowed.

- Current solution:
  - Store Commons page names in Wikidata properties
  - Store Wikidata q-code on Commons page
  - Not use sitelinks which can be pointing to any page

- Issues:
  - Two places to store the same information means a lot of maintenance to keep them in synch
  - Commons moving, redirecting or deleting pages does not update Wikidata
  - Wikidata merges, even deletes, pages without updating Commons
  - Wikidata maintenance tools do not work with linking through properties
• Create Wikidata page for every Category, Creator or Institution page on Wikidata. Connect to Commons through sitelinks. Link them to each other through properties.
  – Issue 1: that is a lot of new pages on Wikidata and interconnections to maintain
  – Issue 2: Where to keep properties? Usually only article pages on Wikidata have properties.

• Allow Commons pages to detect when they are used as a property of some Wikidata page
  – Issue: not the way Wikidata is designed to operate. The properties would become a little like sitelinks