Structured Data on Wikimedia Commons for GLAM-Wiki
SDC Updates

What's new?
What's coming?
Structured Data on Commons
2017-2019

adding metadata on Commons in a structured, multilingual & machine-readable format

making Commons files easier to view, search, edit, organize and re-use
Describing media files with Wikidata...

Depicts (P180)

- ripples (Q15143885)
- fishing net (Q271598)
- canoe (Q171529)
- water (Q283)
- fisherman (Q331432)

Things, not strings

Islahaddow, CC-BY-SA-4.0, based on Fisherman by the Volta River by Alimdaithli, CC-BY-SA 4.0
... so that they can be found in many more languages!
What's new on Commons?

https://w.wiki/UR

- Multilingual file captions
- Depicts statements with Qualifiers
- Other statements than Depicts

Add via file pages, UploadWizard, API
A simple example
Structured data tab:
<table>
<thead>
<tr>
<th>Captions</th>
<th>Language</th>
<th>Description</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Sugar cubes</td>
<td>(2018)</td>
<td></td>
</tr>
<tr>
<td>Nederlands</td>
<td>Suikeridentjes</td>
<td>(2018)</td>
<td></td>
</tr>
<tr>
<td>Deutsch</td>
<td>Würfelzucker</td>
<td>(2018)</td>
<td></td>
</tr>
<tr>
<td>русский</td>
<td>Кусковой сахар</td>
<td>(2018)</td>
<td></td>
</tr>
<tr>
<td>français</td>
<td>Sucre en morceaux</td>
<td>(2018)</td>
<td></td>
</tr>
<tr>
<td>polski</td>
<td>Kostki cukru</td>
<td>(2018)</td>
<td></td>
</tr>
</tbody>
</table>
Depicts

Items portrayed in this file

depicts

Search to add items (house cat, mountain, Taj Mahal, etc.) from Wikidata

sugar cube

- quantity: 12
- color: white
<table>
<thead>
<tr>
<th><strong>creator</strong></th>
<th>Dietmar Rabich</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>license</strong></td>
<td>Creative Commons Attribution-ShareAlike 4.0 International</td>
</tr>
<tr>
<td><strong>Commons quality assessment</strong></td>
<td></td>
</tr>
</tbody>
</table>
Search results

File:Koppie foam grasshopper (Dictyophorus spumans spumans) nymph.jpg
I, the copyright holder of this work, hereby publish it under the following license: English P180 Q10409638 Q33971 (3,554 × 2,369 (3.97 MB)) - 19:19, 30 April 2019

File:Caelifera 1904.jpg
English Grasshoppers P180 Q10409638 (3,088 × 4,111 (8.46 MB)) - 04:12, 14 May 2019

All files with Depicts statements: https://w.wiki/4Qh
In progress

- SPARQL querying
- Lua support under review
- More datatypes
- UploadWizard for custom campaigns (Wiki Loves… style)
- Machine vision (AI-based suggestions for Depicts)

We might also do

- Search depicts statements
- Depicts of depicts
- Depicts and image annotations
GLAM pilot projects

https://w.wiki/4QH

Content partnerships

Art museum
Decorative arts
Maps
Encyclopedic museum

Community initiatives

Tools and prototypes

Digitized books and Wikisource

Hjälp till att översätta bildbeskrivningar från Musikverket

Musikverket behöver din hjälp med att översätta beskrivningar av teaterbilder från sent 1800-tal och tidigt 1900-tal. Logga in med ditt Wikipedia/Wikimedia konto för att hjälpa HU.
First GLAM projects with SDC
Finding GLAMs

Identify GLAMs and the issues preventing them from contributing content
Two parts

Worldwide database of cultural heritage institutions

- Public datasets, crowdsourcing

Supporting GLAMs to share their content

- Case studies, best practices, guides
Problematic data

Workshop with GLAMs
- Historic metadata exposed to new audiences
- Publish recommendations

Using SDC for targeted campaigns?
- Enable shared ownership/custody
- Encourage translations
- Encourage adding context
Historic music

How to interconnect:
- Recording (Audio file)
- Images of the performance
- Sheet music being performed
- OMR of the sheet music

Using sheet music + Copyright data to encourage creation of free new recordings.
Missing metadata structures

Metadata structure for specialised topics:

- Facilitate mapping for GLAM uploads
- Crowdsource missing metadata
- Develop structures for capturing this metadata during digitisation
Metadata Roundtripping

Institutional ingestion of Wikimedia Data
What?

3 participating museums

Two crowdsourcing campaigns

Information from Wikitext, Wikidata and SDC ingested into collection management systems!
Dräkt. Kvinna stående i helfigur med vitt förkläde. Akvarell i storformat av O. Wallgren - Nordiska museet - NMA.0070093 (1).jpg

Depicts:
Add:

<table>
<thead>
<tr>
<th>folk costume</th>
<th>hat</th>
<th>armband</th>
<th>slip-on shoe</th>
<th>fur clothing</th>
</tr>
</thead>
</table>

This tool suggestion feature only contains a small subset of Wikidata items related to fashion. To add other depicts statements you will need to use the usual Wikimedia Commons interface.
- Tools and code is open source
- A PDF file
- Communication and Workshops with museums

Results & What’s next
The ISA tool
ISA is a micro-contribution tool

- It adds structured data to Wikimedia Commons files
- It facilitates the creation, management and analysis of campaigns
- It has a multilingual interface
- It is mobile friendly
Inspired from

Wiki Loves Africa
- 48,466 images collected
- 6,213 unique submitters
- Images viewed 280 millions times
- 225+ events in 23 African countries
In development

- First discussed summer 2018
- Project launched spring 2019
- Involves
  - Wiki in Africa
  - Histropedia
  - Structured Data on Commons
- Phabricator and tools.wmflabs
- http://tools.wmflabs/isa
Campaigns directory

- Tool accessible on desktop and mobile
- Basic stats view within the tool
- Full statistics available on desktop for download
Campaigns are entry points

- Project lead creates campaign
- Campaigns defined primarily by
  - beginning/end time
  - categories of images to work on
  - structured data: depict and/or captions
- Project lead invites participation
Contributions

- Displays images with key extracts of data recorded on Wikimedia Commons
- Contribution to depict and caption data open to logged-in users
- Multilingual display and records
How can you help? (short term)

- Translate the interface *(instructions to be published soon...)*
- Help write the tool wikipage: [[Commons:ISA Tool]]
- Test the tool
Digitization and Wikisource pilot
What?

- Qissa genre of Punjabi literature - private collection
- Document the (ideal) workflow related to digitized works
Workflow

- Collecting the works  Done
- Collecting metadata  Ongoing
- Creating Wikidata items  Done
- Scanning the works  Done
<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
<th>Type</th>
<th>Publisher</th>
<th>Language</th>
<th>Rights Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puran Bhagat</td>
<td>Qadaryer</td>
<td>literary work</td>
<td>Punjabi Qissa</td>
<td>Punjabi</td>
<td>public domain</td>
</tr>
<tr>
<td>Puran Bhagat</td>
<td>Pandit Kavi Kalidas Sahib Guzarwanwalia</td>
<td>literary work</td>
<td>Punjabi Qissa</td>
<td>Punjabi</td>
<td></td>
</tr>
<tr>
<td>Puran Bhagat</td>
<td>Bhai Bhan Singh te Chanan Singh</td>
<td>literary work</td>
<td>Punjabi Qissa</td>
<td>Punjabi</td>
<td></td>
</tr>
<tr>
<td>Puran Bhagat</td>
<td>Bhai Jodh Singh Karmjit Singh Pustkan Wale</td>
<td>literary work</td>
<td>Punjabi Qissa</td>
<td>Punjabi</td>
<td></td>
</tr>
<tr>
<td>Rani Indubati te Raja Roop Chand</td>
<td>Puran Chand Ram Ji Das</td>
<td>literary work</td>
<td>Punjabi Qissa</td>
<td>Punjabi</td>
<td></td>
</tr>
<tr>
<td>Rani Indubati te Raja Roop Chand</td>
<td>Bhai Gurdial Singh and Sons Pustakan Wale</td>
<td>literary work</td>
<td>Punjabi Qissa</td>
<td>Punjabi</td>
<td></td>
</tr>
<tr>
<td>Rooh te Butt</td>
<td>Gurdit Singh</td>
<td>literary work</td>
<td>Punjabi Qissa Amrit Pustak Bhandar</td>
<td>Punjabi</td>
<td></td>
</tr>
<tr>
<td>Roop Basant</td>
<td>Shiv Dyal</td>
<td>literary work</td>
<td>Punjabi Qissa Amrit Pustak Bhandar</td>
<td>Punjabi</td>
<td>public domain</td>
</tr>
<tr>
<td>Sassi Pannu</td>
<td>Hashim</td>
<td>version, edition, or translation</td>
<td>Punjabi Qissa Amrit Pustak Bhandar</td>
<td>Punjabi</td>
<td>public domain</td>
</tr>
<tr>
<td>Sassi Pannu</td>
<td></td>
<td>literary work</td>
<td>Punjabi Qissa</td>
<td>Punjabi</td>
<td></td>
</tr>
<tr>
<td>Shah Bohram te Husan Bano</td>
<td>Imam Bakhsh</td>
<td>literary work</td>
<td>Punjabi Qissa Amrit Pustak Bhandar</td>
<td>Punjabi</td>
<td>public domain</td>
</tr>
</tbody>
</table>
Workflow

- Uploading to Commons: Ongoing
- Adding Structured Data: Ongoing
- Indexing on Wikisource: Next
Open questions?

- Data modeling of published works?
  - Which data lives where?

- How to avoid duplication of data?
Théo Van Rysselberghe - Portret van Marguerite van Mons, Public Domain, collection Museum of Fine Arts Ghent
The Mower (1892), bronze sculpture by Constantin Meunier, photo EvelienBauwens MSKGent013, CC BY-SA 4.0, collection Museum of Fine Arts Ghent
Philippe Wolfers, Design drawing (1902), CC BY-SA 4.0 collection King Baudouin Foundation
Philippe Wolfers, Niké (1902), CC BY-SA 4.0, collection King Baudouin Foundation and Royal Museums of Art and History
MTN25-0006c4-1981-San Salvador de Serantes, CC BY 4.0, © Instituto Geográfico Nacional
Astrolabe of ‘Umar ibn Yusuf ibn ‘Umar ibn ‘Ali ibn Rasul al-Muzaffari, CC0 1.0, collection Metropolitan Museum of Art
Giuliano da Maiano, Francesco di Giorgio, Studiolo from the Ducal Palace in Gubbio, CC0 1.0, collection Metropolitan Museum of Art

Satdeep Gill, Punjabi Qisse 19, CC BY-SA 4.0

Screenshots of ISA tool mockups, Seantime, CC BY-SA 4.0
Screenshot of https://tools.wmflabs.org/smv-description-translations/
THANK YOU!