Structured Data on Wikimedia Commons

Workshop, WikidataCon 2019
Sandra Fauconnier
sfauconnier@wikimedia.org
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)
- fisherman (Q331432)
- water (Q283)
... so that they can be found in many more languages!

Depicts (P180)

 Islahaddow, CC-BY-SA-4.0, based on Fisherman by the Volta River by Alimdaihli, CC-BY-SA 4.0
This is not meant to replace templates and categories!

Eventual replacement is a community decision.
Structured data is CC0.

Files are available under licenses specified on their description page. All structured data from the file and property namespaces is available under the Creative Commons CC0 License; all unstructured text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and the Privacy Policy.
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

Moulay El Yazid Mosque in Marrakech,
Pitchoon76, CC BY-SA 4.0
Structured data tab:
1. Multilingual captions

**Captions**

- **English**
  A horse-drawn carriage in front of the Moulay El Yazid Mosque in Marrakech, Morocco

- **Nederlands**
  Een koets met paard voor de Moskee van de Kasbah in Marrakesh, Marokko

- **Short description of what you see in the file**
- **Text only (no links or wiki markup)**
- **Becomes part of the search index**

https://commons.wikimedia.org/wiki/Commons:File_captions

---

Moulay El Yazid Mosque in Marrakech, Pitchoon76, CC BY-SA 4.0
2. Depicts

- You are using Wikidata items here :-)
- What is shown in the file?
- You can mark things as prominent
- And add qualifiers

https://commons.wikimedia.org/wiki/Commons:Depicts

Moulay El Yazid Mosque in Marrakech,
Pitchoon76, CC BY-SA 4.0
3. Other statements

- Various other things you can say about a file
- You use Wikidata properties and statements here :-)  
  https://commons.wikimedia.org/wiki/Commons:Statements

Moulay El Yazid Mosque in Marrakech, Pitchoon76, CC BY-SA 4.0
Multilinguality

Wikidata is the key here!

You can help by adding more multilingual labels and descriptions to Wikidata items.
Searching (October 2019). All files with Depicts statements: https://w.wiki/4Qh
What's next?

- More **datatypes**
  - URL
  - geo coordinate
  - time/date
  - identifier
  - ...
- **SPARQL querying**
- **Lua support** (under review)
- **Machine vision** (AI-based suggestions for Depicts)
New tools to work with SDC more easily
ISA
by Wiki In Africa, Eugene Egbe, and Histropedia
Add Depicts statements to files in an easy and fun way
(you can do small competitions too)
https://commons.wikimedia.org/wiki/ Commons:ISA_Tool
Participate in the WikidataCon ISA Challenge!

https://www.wikidata.org/wiki/Wikidata:WikidataCon_2019/Program/ISA_Challenge
Introducing ... the ISA game

Play to add information, descriptions or details to images on Wikimedia Commons

ISA is a fun game aimed at enhancing the metadata of images that have been contributed to Wikimedia Commons via the various Wiki Loves competitions. It is a pilot project for Structured Data on Commons.

To celebrate 5 years of Wiki Loves Africa, ISA was built to describe in detail all visual aspects of images that have been contributed to the Wiki Loves Africa contests. It has been specifically modelled to be useful to all of the Wiki Loves X competitions, and beyond. The tool is aimed to increase the usefulness of Wiki Loves X images to illustrate concepts across the Wikimedia projects.

ISA is both an acronym for Information Structured Acceleration, and a chishona language word for 'put' or 'place'.

Scan the qr code to check out ISA on toolforge:
http://bit.ly/PlayISA

Wiki Loves X Organisers ... scan to create your own WLX campaign:
SDC
by Magnus Manske
Add Depicts statements to files in a category
(works like Cat-a-lot)
https://commons.wikimedia.org/wiki/User:Magnus_Manske/sdc_tool.js
AC/DC
by Lucas Werkmeister
Add (the same) structured data to many files at once (for instance a category)
https://commons.wikimedia.org/wiki/Help:Gadget-ACDC
Statements to add

location

Marrakesh

Add qualifier
Batch uploads and edits?

Pattypan, OpenRefine and QuickStatements don't support SDC (well) yet
PetScan to QuickStatements

by Magnus Manske

Search for a list of files via PetScan. You can add structured data statements there, which will direct you to QuickStatements

https://petscan.wmflabs.org/
<table>
<thead>
<tr>
<th>#</th>
<th>Title</th>
<th>Page ID</th>
<th>Namespace</th>
<th>Size (bytes)</th>
<th>Last change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sharpening Steel.jpg</td>
<td>120043</td>
<td>File</td>
<td>345</td>
<td>20181224171941</td>
</tr>
<tr>
<td>2</td>
<td>Shul.jpg</td>
<td>2069394</td>
<td>File</td>
<td>165</td>
<td>20181224203838</td>
</tr>
<tr>
<td>3</td>
<td>Chaira01.jpg</td>
<td>4393070</td>
<td>File</td>
<td>409</td>
<td>20190523203624</td>
</tr>
<tr>
<td>4</td>
<td>Sharpening steel.jpg</td>
<td>7817963</td>
<td>File</td>
<td>307</td>
<td>20190916171954</td>
</tr>
<tr>
<td>5</td>
<td>US Navy 031204-N-17111-001 Mess Management Specialist 3rd Class Benjamin AlmodovarPacheco from Ceiba, Puerto Rico, watches carefully as culinary arts instructor Chef Sam Glass trains him on the best way to sharpen a butcher's.jpg</td>
<td>8202598</td>
<td>File</td>
<td>1024</td>
<td>20191019122614</td>
</tr>
<tr>
<td>6</td>
<td>Schlachter.jpg</td>
<td>11059431</td>
<td>File</td>
<td>1057</td>
<td>20191017180237</td>
</tr>
<tr>
<td>7</td>
<td>Filleting-01.jpg</td>
<td>18561347</td>
<td>File</td>
<td>922</td>
<td>20191018151500</td>
</tr>
<tr>
<td>8</td>
<td>Wetzstahl (fcm).jpg</td>
<td>21505661</td>
<td>File</td>
<td>509</td>
<td>20190906032626</td>
</tr>
<tr>
<td>9</td>
<td>Fusil d'aiguisage.jpg</td>
<td>22542719</td>
<td>File</td>
<td>315</td>
<td>20181226073433</td>
</tr>
<tr>
<td>10</td>
<td>Wetzstahl.jpg</td>
<td>23985710</td>
<td>File</td>
<td>682</td>
<td>20190629010914</td>
</tr>
<tr>
<td>11</td>
<td>Fusil-aiguisoir-A01.jpg</td>
<td>24942494</td>
<td>File</td>
<td>302</td>
<td>20181226111711</td>
</tr>
<tr>
<td>12</td>
<td>Cucina 009.jpg</td>
<td>31510217</td>
<td>File</td>
<td>956</td>
<td>20191019102128</td>
</tr>
<tr>
<td>13</td>
<td>Acciaino - Musei del cibo - Prosciutto - 0011.tif</td>
<td>47541139</td>
<td>File</td>
<td>426</td>
<td>20191007145107</td>
</tr>
<tr>
<td>14</td>
<td>Acciaino - Musei del cibo - Prosciutto - 0142.tif</td>
<td>47543100</td>
<td>File</td>
<td>509</td>
<td>20191007145107</td>
</tr>
</tbody>
</table>

https://petscan.wmflabs.org/?psid=12589701&al_commands=P180%3AQ520049
<table>
<thead>
<tr>
<th></th>
<th>init</th>
<th>Statement</th>
<th>depicts [P180] : honing steel [Q520049]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>init</td>
<td>Sharpening Steel.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>2</td>
<td>init</td>
<td>Shtul.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>3</td>
<td>init</td>
<td>Chaira01.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>4</td>
<td>init</td>
<td>Sharpening steel.jpeg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>5</td>
<td>init</td>
<td>US Navy 031204-N-17111-001 Mess Management Specialist 3rd Class Benjamin AlmodovarPacheco from Ceiba, Puerto Rico, watches carefully as culinary arts instructor Chef Sam Glass trains him on the best way to sharpen a butcher's.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>6</td>
<td>init</td>
<td>Schlachter.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>7</td>
<td>init</td>
<td>Fileleting-01.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>9</td>
<td>init</td>
<td>Fusil d'aiguisage.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
<tr>
<td>10</td>
<td>init</td>
<td>Wetzstahl.jpg</td>
<td>Statement depicts [P180] : honing steel [Q520049]</td>
</tr>
</tbody>
</table>
How you can help

- Add structured data to files :-)  
- Help figure out the data model!  
  [https://w.wiki/9Z8](https://w.wiki/9Z8)  
- Translate Wikidata labels, and add descriptions, in your language!  
- Create new Wikidata items to describe files on Commons!  
- Develop tools!
Commons:Structured data/Modeling

Category (+): Structured Data on Commons (−) (+) () () (+)

From Wikimedia Commons, the free media repository
< Commons:Structured data
Wikidata item not found.

Please help! – Work in progress

You can help by working on guidelines and providing examples on each of the subpages listed below, or by creating additional subpages. Visit the talk page to discuss this process in general.

This page contains an overview of how to model information (metadata) about files on Wikimedia Commons in Structured Data.

The current templates are used as a starting point of what to model.

How to model specific types of metadata

Here, we look at specific types of metadata for a file: for instance, how to model the various dates and times that are related to a file; how to model the various kinds of authorship?

- Description Work in progress
- Date Work in progress
- Source Work in progress
- Author Work in progress

https://commons.wikimedia.org/wiki/Commons:Structured_data/Modeling
Let's try it!

- Let's add structured data to existing Commons files!
  - As diverse as possible
- Are there things you can't describe yet? Let's discuss!
- Resources to check
  - c:Help:File captions
  - c:Help:Depicts
  - c:Help:Statements
  - https://w.wiki/9Z8 (data modeling)
Diverse files...

- **By type**
  - Digital-born photos
  - Digitized (old) photos
  - 3D files
  - Illustrations
  - Vector graphics
  - Audio files
  - Videos
  - ...

- **By 'content' showing/depicting**
  - People
  - Animals
  - Buildings
  - Artworks
  - Digitized maps, manuscripts, books
  - ...

THANK YOU!
Sandra Fauconnier
sandraf@wikimedia.org
Parking lot
(can be used to illustrate more in-depth questions and answers)
Property for indicating copyright of creative works: P6216
Copyright status

Already used on more than 300,000 creative works on Wikidata

Two possible values
- Public domain
- Copyrighted
(Each work can have multiple of these, and also both values combined)

To be qualified (enhanced) with additional information

WORK IN PROGRESS!
Help:Copyrights on Wikidata
https://www.wikidata.org/wiki/Help:Copyrights
or https://w.wiki/3VD

Kudos to Jarekt and Hanno Lans!

- For creative works (Wikidata)
- And digital representations of those works (Wikimedia Commons)
files are special
(of course!)
Main advice for SDC from the GLAM sector:

Strictly distinguish between

1. Metadata of the file

2. Metadata of the creative work that is represented in the file
Creative works and their digital representations
Integration '76 (Q66305367)
sculpture by Hans Dieter Bohnet in Bonn, Germany
Integration, 76 | Integration 1976

- In more languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Label</th>
<th>Description</th>
<th>Also known as</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Integration '76</td>
<td>sculpture by Hans Dieter Bohnet in Bonn, Germany</td>
<td>Integration, 76 Integration, 1976</td>
</tr>
<tr>
<td>Dutch</td>
<td>Integration '76</td>
<td>sculptuur van Hans Dieter Bohnet in Bonn, Duitsland</td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>No label defined</td>
<td>No description defined</td>
<td></td>
</tr>
<tr>
<td>French</td>
<td>No label defined</td>
<td>No description defined</td>
<td></td>
</tr>
</tbody>
</table>

Statements

- instance of
  - sculpture

One creative work ...
https://www.wikidata.org/wiki/Q66305367

... can have many digital representations
https://commons.wikimedia.org/wiki/Category:Integration_%E2%80%9976
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>creator</strong></td>
<td>Hans Dieter Bohnet</td>
<td>Hans Weingartz</td>
</tr>
<tr>
<td><strong>inception</strong></td>
<td>1976/1986</td>
<td>8 April 2006</td>
</tr>
<tr>
<td><strong>depicts</strong></td>
<td>sphere</td>
<td>Integration '76</td>
</tr>
<tr>
<td><strong>copyright status</strong></td>
<td>copyrighted</td>
<td>copyrighted</td>
</tr>
<tr>
<td><strong>license</strong></td>
<td>-</td>
<td>CC BY-SA 2.0 Germany</td>
</tr>
</tbody>
</table>
Faithful digital representations of two-dimensional artworks ("digital surrogates")

Jan Brueghel the Elder: The Last Judgement (1602) - collection Statens Museum for Kunst - Public Domain