Structured Data on Wikimedia Commons

How to get started!
August 2019
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)
... so that they can be found in many more languages!

Depicts (P180)

- ripples (Q15143885)
- fishing net (Q271598)
- canoe (Q171529)
- fisherman (Q331432)
This is **not** meant to replace templates and categories!
Eventual replacement is a community decision.
The project Structured Data on Wikimedia Commons (2017–19) converts the free media files on Wikimedia Commons to a structured and machine-readable format, so that they become easier to view, search, edit, organize and re-use. To achieve that, the Commons backend is migrated to Wikibase, the same technology as used for Wikidata.

New features:
- It is now possible to add multilingual captions and Depicts statements to files on Wikimedia Commons.
- See Commons:File captions and Commons:Depicts for documentation. This documentation is work in progress, and your help is welcome.

Other updates:
- GLAM pilot projects are in progress.
- How should structured data about Commons files be modeled? Input is collected in this table for Wikidata properties. Some already exist, some still need creating.
- Help collect interesting Commons files, as input for the data modelling process! Continuous input is welcome.
- There is a Structured Commons community focus group, for Commons and Wikidata contributors.
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
Structured data tab:

'NEW' icon by Csquest99, CC BY SA 3.0 Unported
### Captions

<table>
<thead>
<tr>
<th>Language</th>
<th>Caption</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Sugar cubes (2018)</td>
<td></td>
</tr>
<tr>
<td>Nederlands</td>
<td>Suikeridentjes (2018)</td>
<td></td>
</tr>
<tr>
<td>Deutsch</td>
<td>Würfelzucker (2018)</td>
<td></td>
</tr>
<tr>
<td>русский</td>
<td>Кусковой сахар (2018)</td>
<td></td>
</tr>
<tr>
<td>français</td>
<td>Sucre en morceaux (2018)</td>
<td></td>
</tr>
<tr>
<td>polski</td>
<td>Kostki cukru (2018)</td>
<td></td>
</tr>
</tbody>
</table>

See fewer languages
Depicts

File information

Structured data

Items portrayed in this file

depicts

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

sugar cube

- quantity: 12
- color: white

Dietmar Rabich / Wikimedia Commons / "Würfelzucker -- 2018 -- 3564" / CC BY-SA 4.0
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

https://commons.wikimedia.org/wiki/Category:Integration_%E2%80%9976
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 surgery | 0 references | add reference | add value |

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

... can have many digital representations
https://commons.wikimedia.org/wiki/Category:Integration_%E2%80%9976
<table>
<thead>
<tr>
<th>Statements</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>instance of</td>
<td>sculpture</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

**creator**
- Hans Dieter Bohnet
- Hans Weingartz

**inception**
- 1976/1986
- 8 April 2006

**depicts**
- sphere
- Integration '76
- Langer Eugen

**copyright status**
- copyrighted
- copyrighted

**license**
- CC BY-SA 2.0 Germany
Faithful digital representations of two-dimensional artworks ("digital surrogates")

Jan Brueghel the Elder: The Last Judgement (1602) - collection Statens Museum for Kunst - Public Domain
<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>creator</td>
<td>Jan Brueghel the Elder</td>
</tr>
<tr>
<td>inception</td>
<td>1602</td>
</tr>
<tr>
<td>depicts</td>
<td>angel; devil; heaven;...</td>
</tr>
<tr>
<td>digital representation of</td>
<td>n/a</td>
</tr>
<tr>
<td>copyright status</td>
<td>Public Domain</td>
</tr>
<tr>
<td>license</td>
<td>CC0</td>
</tr>
</tbody>
</table>

**Additional Information**

- [https://www.wikidata.org/wiki/Q20354492](https://www.wikidata.org/wiki/Q20354492)
- [https://commons.wikimedia.org/wiki/File:Jan_Bru
ergel_d.%C3%84._-_The_Last_Judgement_-_KMSs
p180_-_Statens_Museum_for_Kunst.jpg](https://commons.wikimedia.org/wiki/File:Jan_Bru
ergel_d.%C3%84._-_The_Last_Judgement_-_KMSs
p180_-_Statens_Museum_for_Kunst.jpg)

Statens Museum for Kunst?

- [https://commons.wikimedia.org/wiki/File:Jan_Bru
ergel_d.%C3%84._-_The_Last_Judgement_-_KMSs
p180_-_Statens_Museum_for_Kunst.jpg](https://commons.wikimedia.org/wiki/File:Jan_Bru
ergel_d.%C3%84._-_The_Last_Judgement_-_KMSs
p180_-_Statens_Museum_for_Kunst.jpg)

The Last Judgement (painting)

Public Domain
Searching (August 2019). All files with Depicts statements: https://w.wiki/4Qh
What's coming?

- More datatypes
  - URL
  - geo coordinate
  - time/date
  - identifier
  - ...

- [Image]
What's coming?

- **Search** depicts statements (gadget)
- **Depicts of depicts**
- **Depicts and annotations**
- **UploadWizard** for **custom campaigns** (Wiki Loves… style)
- **SPARQL** querying
- **Lua support** (under review)
- **Machine vision** (AI-based suggestions for Depicts)
Batch uploads and edits?

Pattypan, OpenRefine and QuickStatements don't support SDC (well) yet
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: http://bit.ly/CreateISA
How you can help
## Generic properties

The properties below apply to all types of files.

<table>
<thead>
<tr>
<th>Property / Aliases</th>
<th>Currently exists on Wikidata as</th>
<th>Data type</th>
<th>Qualifiers</th>
<th>Remarks</th>
</tr>
</thead>
</table>
| Depicts / Manifestation of / Digital representation of | depicts (P180) | Item | See below | Same property for things 'depicted' in an audio file. It is actually more 'manifestation of' or 'digital representation of'
| Depicted place | | Item | | To be included in 'Depicts' in general? |
| Depicted event | | Item | | For instance, for photos of a specific festival, a play, a musical performance. To be included in 'Depicts' in general? |
| Depicted date | date depicted (P2913) | Time | | |
| Depicted period | | Item | | where a specific date cannot be given; compare time period (P2348), set in period (P2408) |
| Date of file creation / Scan date | inception (P571) (maybe a more specific subproperty is warranted) | Time | | |
| Source website | No, to be created | Item | | The Wikidata item of the website where the file was taken from, e.g. YouTube, Flickr, Europeana, a museum’s website |
| Source URL | Probably to be created. There is reference URL (P854) (reference URL) but that is for references specifically | URL | | The specific URL where the file was found on the source website |
| Institution’s media file identifier | No, to be created (probably 1 identifier property per institution) | Identifier | | For institutions whose media repositories have unique and persistent identifiers for individual media files. Not to be confused with identifiers for creative works. |

https://commons.wikimedia.org/wiki/Commons:Structured_data/Properties_table
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)
● 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
  ○ Properties table
  ○ d:Help:Copyrights

Let's get to work!
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
    - ...

WIKIMANIA STOCKHOLM
THANK YOU!