

Structured Data on Commons for GLAM institutions



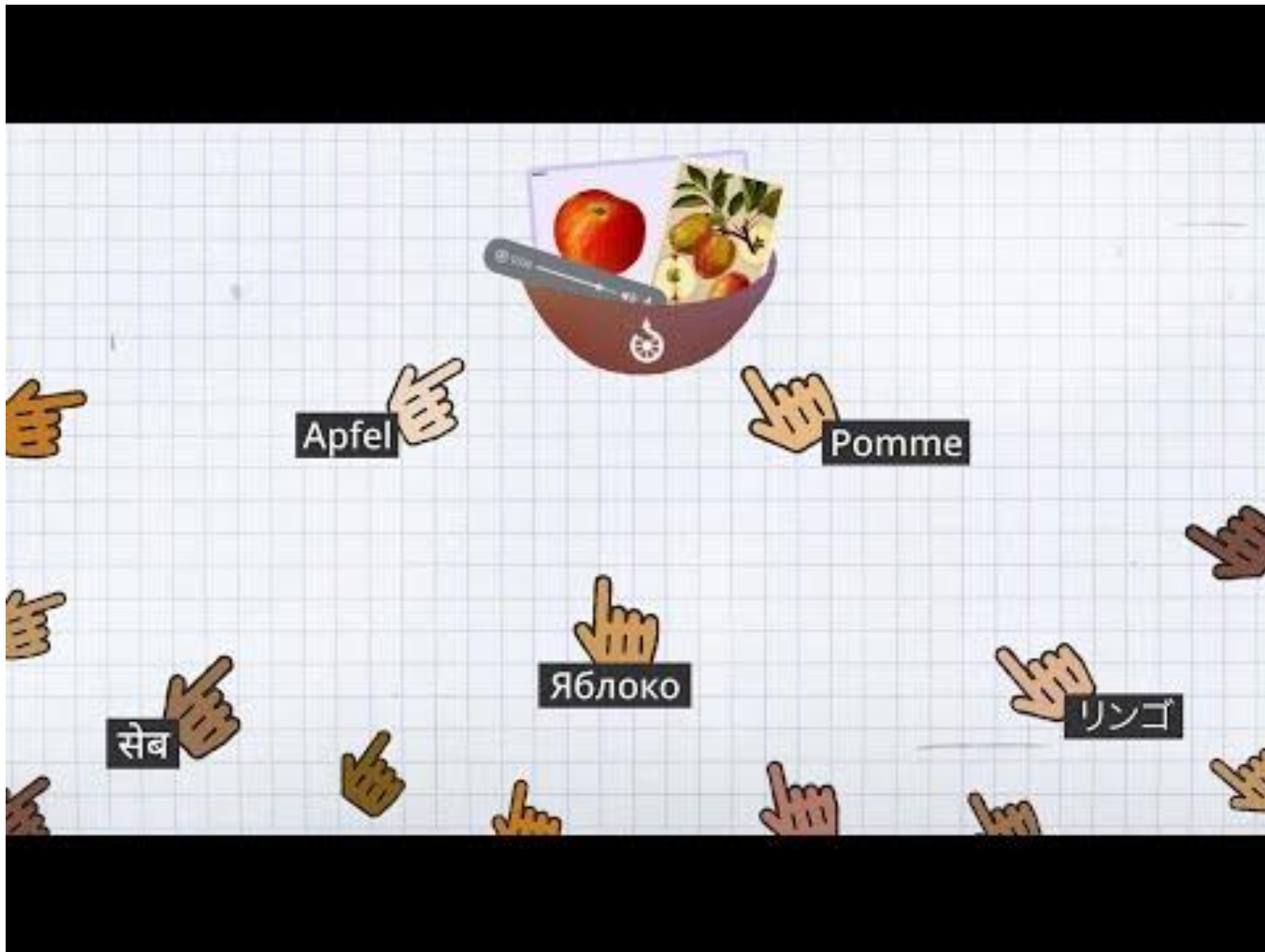
WIKIMEDIA
FOUNDATION

Structured Data on Commons

Presentation and tutorial for GLAM institutions

2021 LD4 Conference on Linked Data
Giovanna Fontenelle, Program Officer,
GLAM and Culture





"Structured Data on Commons: We're making 50 million audiovisual pieces accessible in all languages!"

[w.wiki/YRG](https://www.wiki.org)

Structured Data on Commons

[w.wiki/UR](https://www.wikidata.org/wiki/UR)

What is it?

- A structured data insertion system for Wikimedia Commons
- A multilingual information understood by humans and machines

How it works?

- Adds multilingual captions to files
- Inserts Wikidata information into media files on Commons
 - Linked Open Data for media files

File information

Structured data



Captions

[Edit](#)

English	Sugar cubes (2018)
español	Cubos de azúcar (2018)
français	Sucre en morceaux (2018)
português	Cubos de açúcar (2018)
português do Brasil	Cubos de açúcar (2018)
Deutsch	Würfelfucker (2018)
русский	Кусковой сахар (2018)
polski	Kostki cukru (2018)
Nederlands	Suikerklontjes (2018)
norsk (bokmål)	Sukkerbiter (2018)
italiano	Zollette di zucchero (2018)
svenska	Sockerbitar (2018)
Esperanto	Cukerkuboj (2018)
македонски	Коцки од шеќер (2018)
한국어	각설탕 (2018)
magyar	Kockacukor (2018)
čeština	Kostkový cukr (2018)
中文	糖块 (2018)

A image of sugar cubes: captions in more than 25 languages.

[w.wiki/3ehj](https://www.wiki/3ehj)

file name **File:Mona Lisa, by Leonardo da Vinci, from C2RMF retouched.jpg**

From Wikimedia Commons, the free media repository



image

captions, summary, and other information

File information **Structured data**

structured data statements

Wikidata property

Items portrayed in this file

[Edit](#)

edit items/qualifiers

depicts

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

add items from Wikidata

Wikidata value

Person depicted in Mona Lisa

Prominent

rank

statement group

- expression, gesture or body pose: smile
- expression, gesture or body pose: sitting
- hair color: brown
- eye color: brown
- relative position within image: **pct:11.1992,10,78.8008,82.9518**

Wikidata qualifiers

creator [Edit](#)

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

Leonardo da Vinci Prominent

object has role: painter

Center for Research and Restoration of Museums of France [Mark as prominent](#)

object has role: photographer

add Wikidata statement

[+ Add statement](#)

File:Templo de fuego, Baku, Azerbaiyán, 2016-09-27, DD 33.jpg

From Wikimedia Commons, the free media repository

[File](#)

[File history](#)

[File usage on Commons](#)

[File usage on other wikis](#)

[Metadata](#)

[Reuse this file](#)



File:Templo de fuego, Baku, Azerbaiyán, 2016-09-27, DD 33.jpg

From Wikimedia Commons, the free media repository

[File](#)

[File history](#)

[File usage on Commons](#)

[File usage on other wikis](#)

[Metadata](#)

[Reuse this file](#)



Structured Data on Commons

[w.wiki/UR](https://www.wikidata.org/wiki/UR)

What it does?

- Improves accessibility
- Improves the quality of information
- Offers new ways to find, explore, and use content

このファイルに描写されている項目

[編集](#)

題材:ts

追加する項目を検索 (猫、山、ターザン、マスルなど) etc.)

[ウィキデータを検索](#)

角砂糖cube

[Provi](#) 顕著

数:12y:12
色:白white
形状:直方体d
材料:白砂糖:white sugar

作者:or

[編集](#)

追加する項目を検索 (猫、山、ターザン、マスルなど) etc.)

[ウィキデータを検索](#)

Dietmar Rabich 英語

[Provi](#) 顕著

使用許諾

[編集](#)

追加する項目を検索 (猫、山、ターザン、マスルなど) etc.)

[ウィキデータを検索](#)

CC BY-SA 4.0 国際 Attribution-ShareAlike 4.0 International

[Provi](#) 顕著

コモンズの品質評価assessment

[編集](#)

追加する項目を検索 (猫、山、ターザン、マスルなど) etc.)

[ウィキデータを検索](#)

[ウィキメディアのコモンズの秀逸な画像](#)

[顕著とマークする](#)

[ウィキメディアのコモンズの今日の一枚](#)

[顕著とマークする](#)

Structured Data on Commons

[w.wiki/3erF](https://www.wikidata.org/wiki/3erF) or [w.wiki/3erG](https://www.wikidata.org/wiki/3erG)

What it does **for GLAMs?**

"It benefits cultural heritage data: Wikimedia Commons files can be described with structured, machine-readable and multilingual metadata (Linked Open Data), structured copyright and attribution information, and rich APIs."

Structured Data on Commons

[w.wiki/3erF](https://www.wikidata.org/wiki/3erF) or [w.wiki/3erG](https://www.wikidata.org/wiki/3erG)

What it does **for GLAMs**?

Makes the content from the collections more...

- Translatable
- Accessible
- Machine-readable
- Connectable
- Findable
- Usable

Structured Data on Commons

How to contribute?

- Upload new, freely licensed files to Wikimedia Commons.
 - Multilingual file captions and depicts statements
- Add or enhance structured data to already existing files.

Structured Data on Commons

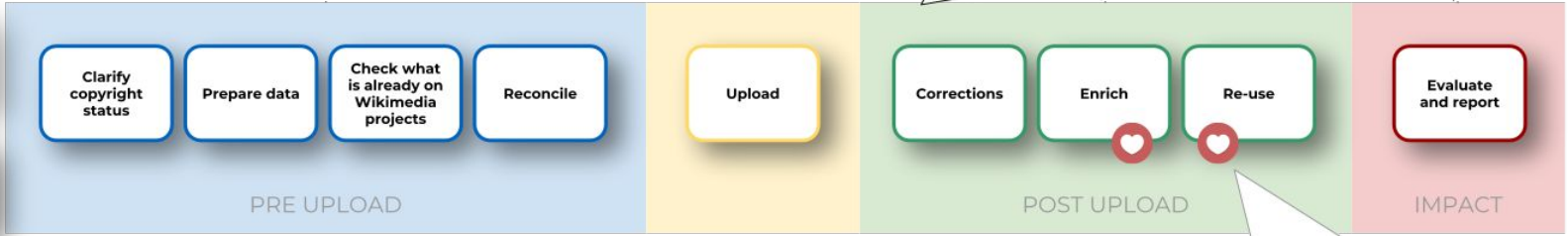
How to get involved?

- Wikimedia community: providing feedback, helping others, adding structured data statements, translating content, helping to decide how to model certain metadata...
- Representatives from GLAMs: sharing content, providing feedback, making use of the tools, and helping with modeling.



First negotiations

Dialogue with Wikimedia communities throughout the entire process



These two steps are sometimes forgotten - or you can do so much more here!

Workflow

for data and media partnerships with Wikidata and with (Structured Data on) Wikimedia Commons



Structured Data on Commons stages

PREPARE and tidy source data and images	RECONCILE with Wikimedia modeling and coverage	INGEST data and media
<ul style="list-style-type: none">• Spreadsheets (external)• OpenRefine	<ul style="list-style-type: none">• OpenRefine (SDC still pending)	<ul style="list-style-type: none">• QuickStatements• Pattypan• url2commons• Pywikibot
ANALYZE, correct, and enrich	RE-USE content intra-wiki and externally	REPORT and measure
<ul style="list-style-type: none">• AC/DC• SD.js• Image Annotator	<ul style="list-style-type: none">• Media Search on Commons• Image Suggestions	<ul style="list-style-type: none">• Wikimedia Commons Query Service (Beta)

Structured Data on Commons

How to contribute?

With tools:

- ISA
- AC/DC
- SDC
- Image Annotator

With batch upload tools:

- QuickStatements + Minefield + PetScan

Structured Data on Commons

Difficulty levels when working with structured data on Commons:

Easy	Medium	Hard	Expert
<p>Built into Commons interface:</p> <ul style="list-style-type: none">• Structured Data tab on the Wikimedia Commons Media Search	<p>Scripts on Commons:</p> <ul style="list-style-type: none">• AC/DC• SD.js	<p>Toolforge and external tools:</p> <ul style="list-style-type: none">• PetScan• Pattypan• Quickstatements• Image Annotator• Minefield• Wikimedia Commons Query Service	<p>Scripting and bots:</p> <ul style="list-style-type: none">• Pywikibot• Direct API interface

AC/DC

- Add to Commons / Descriptive Claims
- Adds a collection of structured data statements to a set of files.
- Needs to be installed:
 - Preferences > Gadgets > Add to Commons / Descriptive Claims (ACDC).
 - It becomes available on your Commons' sidebar, on the "Tools" section.

AC/DC Publish changes

Files to edit

File:2012 Rostelecom Cup 01d 037 Caroline Zhang.JPG ×
File:Caroline Zhang combo spin.JPG × File:Caroline zhang grand prix 2007 lp 2.jpg ×
... File:Caroline Zhang Spin EX 2008 Junior Worlds.jpg × *
File:2007 Skate America Zhang04.jpg × File:2007-GPF-ladies-Zhang002.jpg ×
File:Example.png | File:Example.jpg

Statements to add

Items portrayed in this file
depicts

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

Caroline Zhang Mark as prominent

×
[Add qualifier](#)

Remove all

[+ Add statement](#)

ISA

- Official pilot project for Structured Data on Commons.
- Allows anyone to add structured data to files on Commons in an easy and fun way.
- A tool for enrichment and micro-contributions useful for some GLAMs.
- Adds depicts statements and multilingual captions to files.
- Anyone can create and organize small campaigns and competitions.
 - isa.toolforge.org

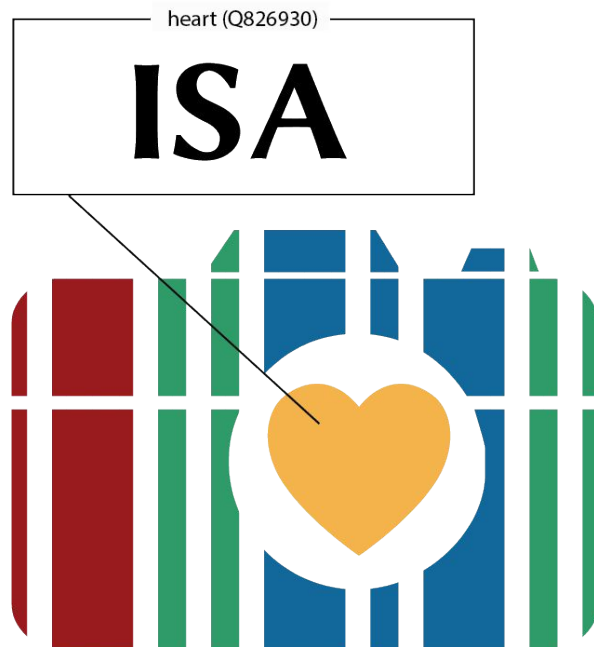


Image Annotator and SDC

Image Annotator

- A tool that runs image annotation campaigns on Wikimedia Commons.
- Adds depicts (P180) statements with the relative position within image (P2677) qualifier.
- Based on the Wikidata Image Positions.

SDC

- User script that allows adding a limited number of structured data statements to files in a category.

Open Refine + SDC

- Approved proposal - mid-2022 ([w.wiki/3DbY](https://www.wikidata.org/wiki/3DbY))
- Fix the current upload problems
- Help with GLAM-Wiki projects
- Allow the Wikidata and Wikimedia Commons, with Structured Data, integration
 - Batch editing SDC
- "There is no reliable, flexible and powerful tool yet to add and modify structured data to very large batches of files on Wikimedia Commons."



QuickStatements

- Wikidata tool
- Spreadsheet, text editor or an external code like Lua (in a template).
- Add and remove statements, labels, descriptions, aliases, properties, lexemes, and qualifiers or sources.
- **Batch edit SDC with the help of Minefield or PetScan:**
 - Change "*Create new command batch for*" from Wikidata to Commons.
 - Use Media IDs (MIDs) instead of Q numbers (QIDs).



PetScan and Minefield



Tools that extract MIDs to be use on QuickStatements.

MIDs: letter "M" + eight numbers

- **PetScan:**
 - Extracts MIDs through files on Commons categories.
- **Minefield:**
 - Convert Commons file page titles to Media IDs.



List of MIDs generated from PetScan

PetScan [Manual](#) [Issues](#) English  [Examples](#) [List](#) 

[Categories](#) [Page properties](#) [Templates&links](#) [Other sources](#) [Wikidata](#) [Output](#)

Language [Commons](#) [Wikispecies](#) [Wikidata](#)

Project

Depth (depth 0 means no subcategories; appending '|' and a number to a category in the fields below will override it for that category)

Categories

Combination Intersection Union

Negative categories

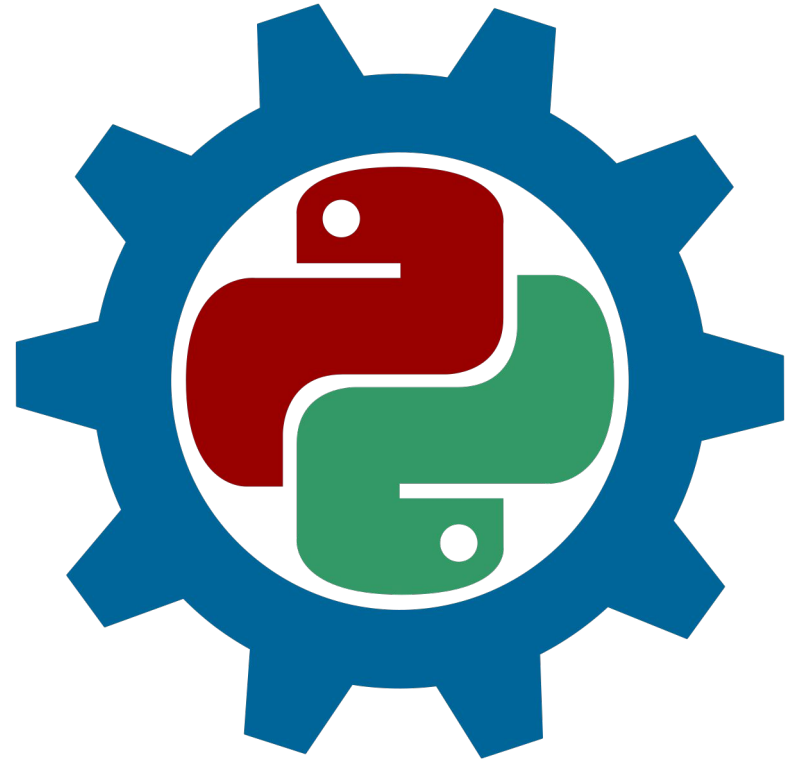
[Do it!](#) [Link to a pre-filled form for *the query you just ran with* and *without* auto-run. PSID is **18866400**. \[Find images for Wikidata.\]\(#\)](#)

List of MIDs generated from PetScan

#	Title	Page ID	Namespace	Size (bytes)	Last change
<input checked="" type="checkbox"/>	1 Metropolitan Museum of Art (19).JPG	36428915	File	533	20201029145319
<input checked="" type="checkbox"/>	2 Metropolitan Museum of Art (22).JPG	36428922	File	533	20201030121929
<input checked="" type="checkbox"/>	3 Plaque Portrait of Benjamin Franklin MET 25577.jpg	57363912	File	4507	20210322205205
<input checked="" type="checkbox"/>	4 Plaque Portrait of Benjamin Franklin MET 25592.jpg	57363913	File	4665	20210322205304
<input checked="" type="checkbox"/>	5 Plaque Portrait of Benjamin Franklin MET 25578.jpg	57363914	File	4451	20210322205213
<input checked="" type="checkbox"/>	6 Portrait of Benjamin Franklin MET 25595.jpg	57363915	File	4254	20210322205618
<input checked="" type="checkbox"/>	7 Plaque Portrait of Benjamin Franklin MET 25580.jpg	57363916	File	4529	20210322205230
<input checked="" type="checkbox"/>	8 Plaque Portrait of Benjamin Franklin MET 25582.jpg	57363917	File	4484	20210322205246
<input checked="" type="checkbox"/>	9 Portrait of Benjamin Franklin MET 25596.jpg	57363918	File	4273	20210322205626
<input checked="" type="checkbox"/>	10 Plaque Portrait of Benjamin Franklin MET 25598.jpg	57363919	File	4214	20210322205642
<input checked="" type="checkbox"/>	11 Portrait of Benjamin Franklin MET 25894 83.2.234.jpg	57363920	File	4365	20210322205609
<input checked="" type="checkbox"/>	12 Medallion of Benjamin Franklin MET 25593.jpg	57363921	File	4358	20210322205313
<input checked="" type="checkbox"/>	13 Plaque Portrait of Benjamin Franklin MET 25581.jpg	57363922	File	4265	20210322205236
<input checked="" type="checkbox"/>	14 Portrait of Benjamin Franklin MET 25894.jpg	57363923	File	4211	20210322205701
<input checked="" type="checkbox"/>	15 Portrait Miniature of Benjamin Franklin MET 25579.jpg	57363924	File	4520	20210322205221
<input checked="" type="checkbox"/>	16 Portrait of Benjamin Franklin MET 25583.jpg	57363925	File	4476	20210322205254
<input checked="" type="checkbox"/>	17 Portrait Miniature of the Marquis de Lafayette MET 25677.jpg	57363926	File	4340	20210322205743
<input checked="" type="checkbox"/>	18 Portrait of Benjamin Franklin MET 25656.jpg	57363927	File	4198	20210322205706
<input checked="" type="checkbox"/>	19 Portrait of Benjamin Franklin MET 25676.jpg	57363928	File	4186	20210322205737

Pywikibot

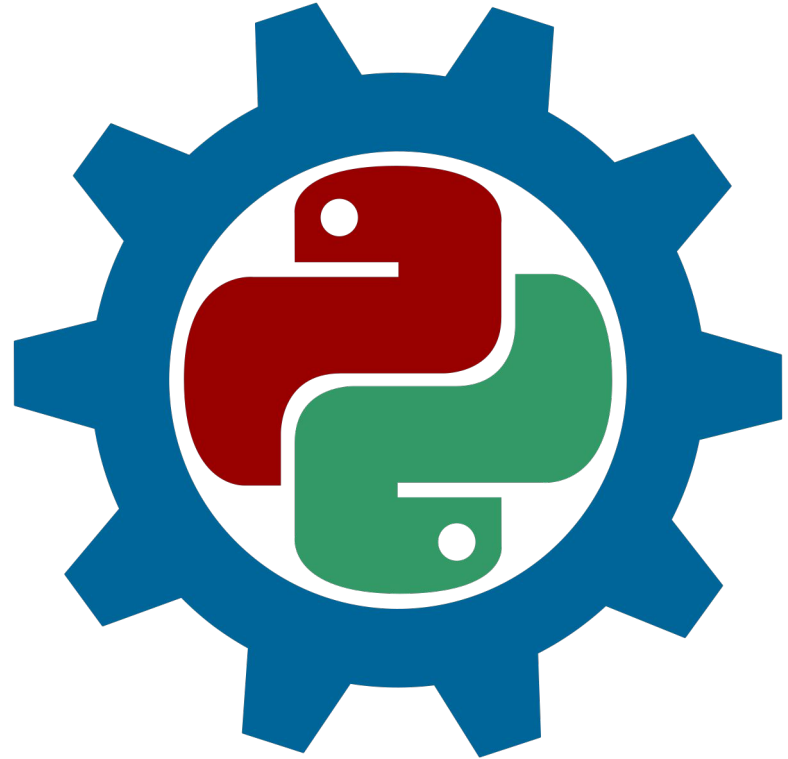
- Framework that offers layers for large-scale batch operations
- Bots run by **users** to do batch uploads
- No SDC support
 - Ticket ([T223820](#))
 - Properly implement structured data access on Commons in Pywikibot
 - "Writing Structured Data on Commons with Python" ([bit.ly/3Ba1MW2](#))



WIKIMEDIA
FOUNDATION

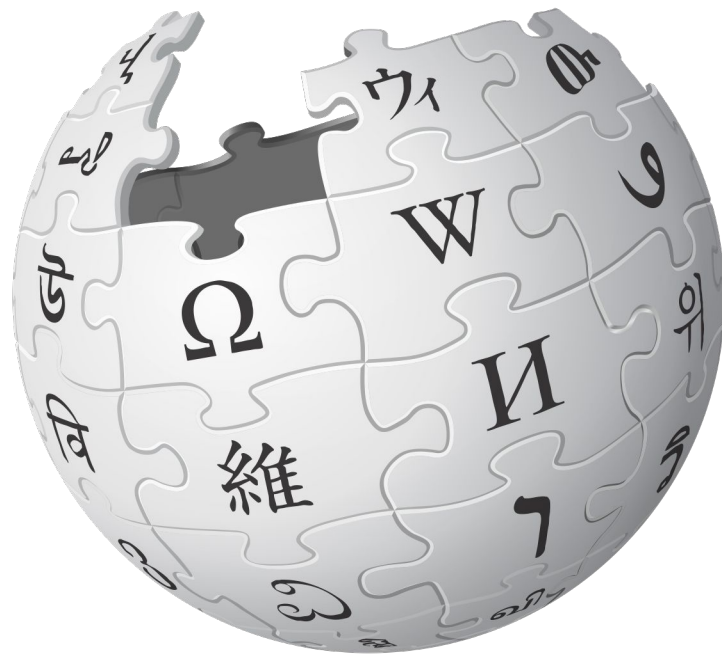
Pywikibot

- No SDC support
 - SDC bot operations are done by hand
 - BotMultichillT, SchlurcherBot, JarektBot
 - GLAM bots
 - AliciaFagervingWMSE-bot, METbot, GeographBot, DPLA bot



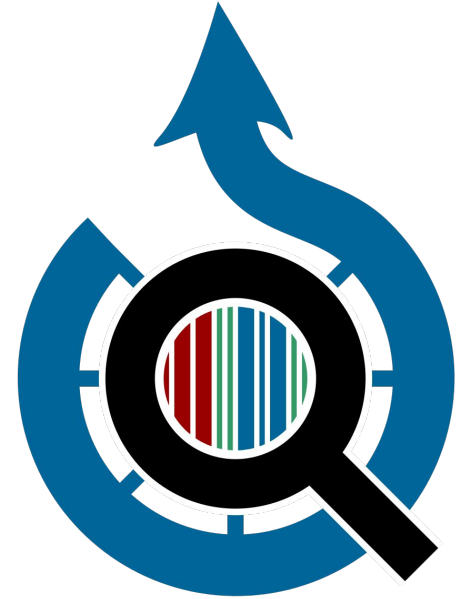
Reuse of content

- Forgotten phase
- Useful for GLAMs
- Structured Data Across Wikimedia
 - Image Suggestions
 - Uses certain metadata
 - Depicts (P180)
 - Digital representation of (P6243)
 - Descriptive title, file captions in multiple languages, detailed descriptions, relevant categories...



Measuring and visualizing impact

- Wikimedia Commons Query Service (WCQS)
- Commons Searches
- Outreach Dashboard



Commons Searches

Media search or Search

- Use the code "*haswbstatement:*", followed by the properties and QIDS you want to search.
- Based on the structure available on Wikidata for Commons files.
- Example:
 - *digital representation of* (P6243) > *The Kiss* (Q698487)
 - "*haswbstatement:P6243=Q698487*"


Search media

Search media

haswbstatement:P6243=Q698487

[Images](#) [Audio](#) [Video](#) [Other Media](#) [Categories and Pages](#)

All licenses ▾ All file types ▾ All image sizes ▾ Sort by: Relevance ▾



[haswbstatement:P6243=Q698487](#)

Commons Searches

Media search or Search

- Searching Structured Data on Commons statements in different Commons categories

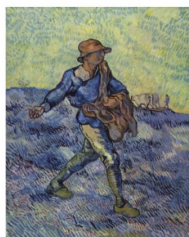
1. Add "incategory:"
2. Followed by the name of the category with underlines (_) between each space:
"Images_from_Metropolitan_Museum_of_Art"
3. Add "haswbstatement:" and the property you wish to search, like depicts, "P180".

[incategory:Images_from_Metropolitan_Museum_of_Art haswbstatement:P180](#)

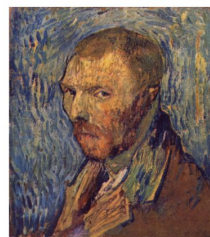
54,400 results

Commons Query Service

- Based on the Wikidata Query Service
- It's a beta version and not always up to date.
- Example:
 - Images depicting Van Gogh works
 - Painting files with creator (P170) set to Vincent van Gogh (Q5582).



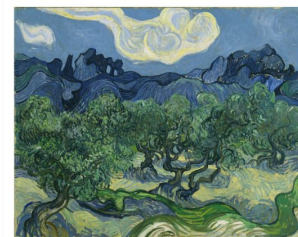
[commons:Vincent_Willem_van_Gogh_...](#)
Q sdc:M151853
Q Sower, The (after Millet)
painting by Vincent van Gogh



[commons:Vincent_Willem_van_Gogh_10...](#)
Q sdc:M151945
Q Self-portrait
painting by Vincent van Gogh (F528)



[commons:Two_sunflowers.jpg](#)
Q sdc:M2428313
Q Sunflowers
painting by Vincent van Gogh in the Metropolitan Museum of Art



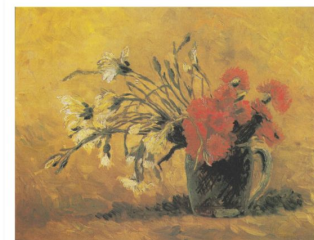
[commons:Van_Gogh_The_Olive_Trees_.jpg](#)
Q sdc:M4104882
Q The Olive Trees
painting by Vincent van Gogh



[commons:Van_Gogh_-_Neugeborenes_Kalb_auf_Stroh.jpeg](#)
Q sdc:M9108161
Q Newborn calf on straw



[commons:Van_Gogh_-_Landschaft_mit_Kirche_und_Häusern.jpeg](#)
Q sdc:M9363025
Q Landscape with Church and Houses

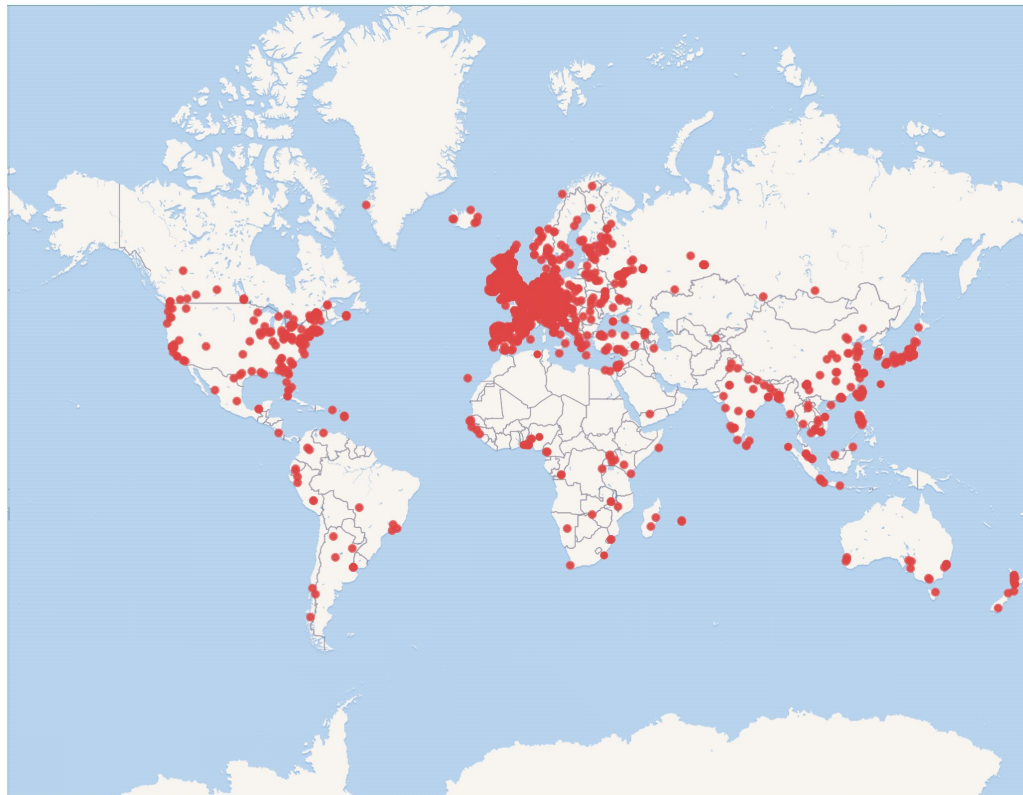


[commons:Van_Gogh_-_Vase_mit_roten_und_weißen_Nelken_a...](#)
Q sdc:M9403086
Q Vase with red and white carnation on a yellow background



Commons Query Service

- Example:
 - Depictions of bridges
 - Map of some images tagged with "depicts" (P180) = "bridge" (Q12280)



Metropolitan Museum of Art

- Uses PetScan and QuickStatements:
 - depicts (P180), creator (P170), and inventory number (P217)
 - 43 thousand files
- Uses Python through the METbot:
 - The Met object ID (P3634) and collection (P195)
 - 140 thousand files
- Enhance the number of files being used across Wikimedia.

THE
MET

Metropolitan Museum of Art

Items portrayed in this file	Edit
<i>depicts</i>	
<input type="text" value="Search to add items"/>	from Wikidata ...
John Surratt	Mark as prominent
<input type="text" value="determination method: Metropolitan Museum of Art Tagging Initiative"/>	
portrait	Mark as prominent
<input type="text" value="determination method: Metropolitan Museum of Art Tagging Initiative"/>	
American Civil War	Mark as prominent
<input type="text" value="determination method: Metropolitan Museum of Art Tagging Initiative"/>	
David Herold	Mark as prominent
<input type="text" value="determination method: Metropolitan Museum of Art Tagging Initiative"/>	
John Wilkes Booth	Mark as prominent
<input type="text" value="determination method: Metropolitan Museum of Art Tagging Initiative"/>	
man	Mark as prominent
<input type="text" value="determination method: Metropolitan Museum of Art Tagging Initiative"/>	

WAR DEPARTMENT, WASHINGTON, APRIL 20, 1865.

\$100,000 REWARD!

THE MURDERER

Of our late beloved President, Abraham Lincoln,
IS STILL AT LARGE.

\$50,000 REWARD

Will be paid by this Department for his apprehension, in addition to any reward offered by the President or the Executive.

\$25,000 REWARD

Will be paid for the apprehension of JOHN H. SURRATT, one of Booth's Accomplices.

\$25,000 REWARD

Will be paid for the apprehension of David C. Herold, another of Booth's accomplices.

LIBERAL REWARDS will be paid for any information that shall render to the arrest of either of the above-named criminals, or their accomplices.

All persons harboring or concealing the said persons, or either of them, or aiding or assisting their concealment or escape, will be treated as accessories in the murder of the President and the attempted assassination of the Secretary of State, and shall be subject to trial before a Military Commission and the punishment of DEATH.

Let the stake of interest bleed be removed from the back by the arrest and apprehension of the murderers.

All good citizens are exhorted to aid public justice in this connection. Every man should consider his own convenience aligned with this solemn duty, and that neither night nor day shall it be unaccomplished.

EDWIN M. STANTON, Secretary of War.

DESCRIPTIONS.—**BOOTH** is five feet 7 or 8 inches high, slender build, light forehead, black hair, blue eyes, and wears a heavy black mustache.

JOHN H. HEROLD is about five feet 10 inches high. Hair rather thin and dark; eyes rather light; an honest, broad, well-set, and stout man. Complexion rather pale and clear, with color in his cheeks. Wears light-colored, fine quality, blue-denim apparel; always wears red suspenders; also wears a cap projecting at the top. Forehead rather low and square, but broad. Parts his hair on the right side; neck rather long. His lips are firmly set. A thin nose.

DAVID C. HEROLD is a five feet six inches high, fair skin, dark eyes, dark, wavy hair, full nose, thin cheeks, head short and fleshy, feet small, heavy high, round bottomed, naturally quick and active, slightly closes his eyes when looking at a person.

NOTE.—In addition to the above, there and other individuals have offered rewards amounting to about one hundred thousand dollars, making an aggregate of about **TWO HUNDRED THOUSAND DOLLARS.**

Digital Public Library of America

- Largest single contributor to Commons + biggest GLAM contributor
 - Not on Wikidata
- Approved proposal ([w.wiki/3ejE](https://www.wikidata.org/wiki/3ejE))
 - *"... improve the DPLA digital asset pipeline by implementing Structured Data on Commons, continuous data synchronization, improved analytics for partners, and experimentation with SDC-based image citation solutions."*
 - 2 million edits to add structured data and make metadata updates



Digital Public Library of America

- Discussion about the proper Structured data on Commons modeling to be used on DPLA's files.
 - [w.wiki/3ejC](https://www.wikidata.org/wiki/3ejC)
- DPLA bot
 - add/update SDC statements and image metadata to sync with data sources



Finna

- A project maintained by the National Library of Finland.
- Uploads and edits by the users Zache and FinnaUploadBot.
- Adding statements such as file captions in different languages, inception (P571), source of file (P7482), depicts (P180), creator (P170), location (P276), and coordinates of the point of view (P1259).
- More than 16 thousand files with SDC



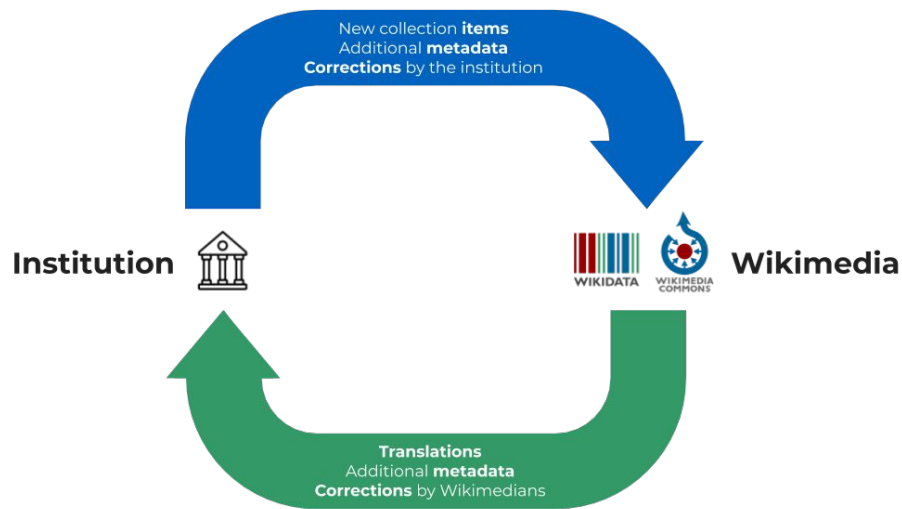
Punjabi Qisse

- Official pilot project for Structured data on Commons.
- Structured data was added to a set of old, digitized Punjabi books in the Qisse genre.
- Re-use of data between various Wikimedia projects: Wikidata, Wikimedia Commons, and Wikisource, improving the workflow of contribution to these projects.
 - w.wiki/3eja



Wikimedia Commons Data Roundtripping

- Official pilot project for Structured data on Commons.
- Research and experiments around a area unlocked by structured data: roundtripping data between GLAM databases and Wikimedia Commons / Wikidata.
- A project by the Wikimedia Foundation and the Swedish National Heritage Board (Riksantikvarieämbetet).
- w.wiki/3ejV



Activities

- [ISA \(isa.toolforge.org\)](https://isa.toolforge.org)
 - Create a campaign for your library or contribute to an existing campaign!
- AC/DC (w.wiki/3ekA)
- SDC (w.wiki/3ekB)
- Image Annotator (image-annotator.toolforge.org)
- Help with the Structured data on Commons Modeling:
 - author, copyright, date, depicts, image captured with, location, maps, participant, quality, significant event, source...
 - Modeling page (w.wiki/9Z8) + Talk page (w.wiki/3ekC)
 - Properties table (w.wiki/YRm) + Talk page (w.wiki/3ekD)
- Join the Digital Public Library of America with their modeling discussion:
 - w.wiki/3ejC

Discussion



WIKIMEDIA
FOUNDATION



Thank you!

Giovanna Fontenelle | GFontenelle (WMF)
gfontenelle@wikimedia.org



WIKIMEDIA
FOUNDATION