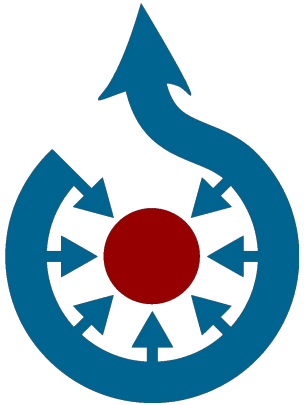
Structured Data on Wikimedia Commons

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)

Islahaddow, CC-BY-SA-4.0, based on Fisherman by the Volta River by Alimdaihli, CC-BY-SA 4.0

… so that they can be found in many more languages!

Depicts (P180)

Informação do arquivo | Dados estruturados

Itens retratados neste arquivo — Editar

retrata

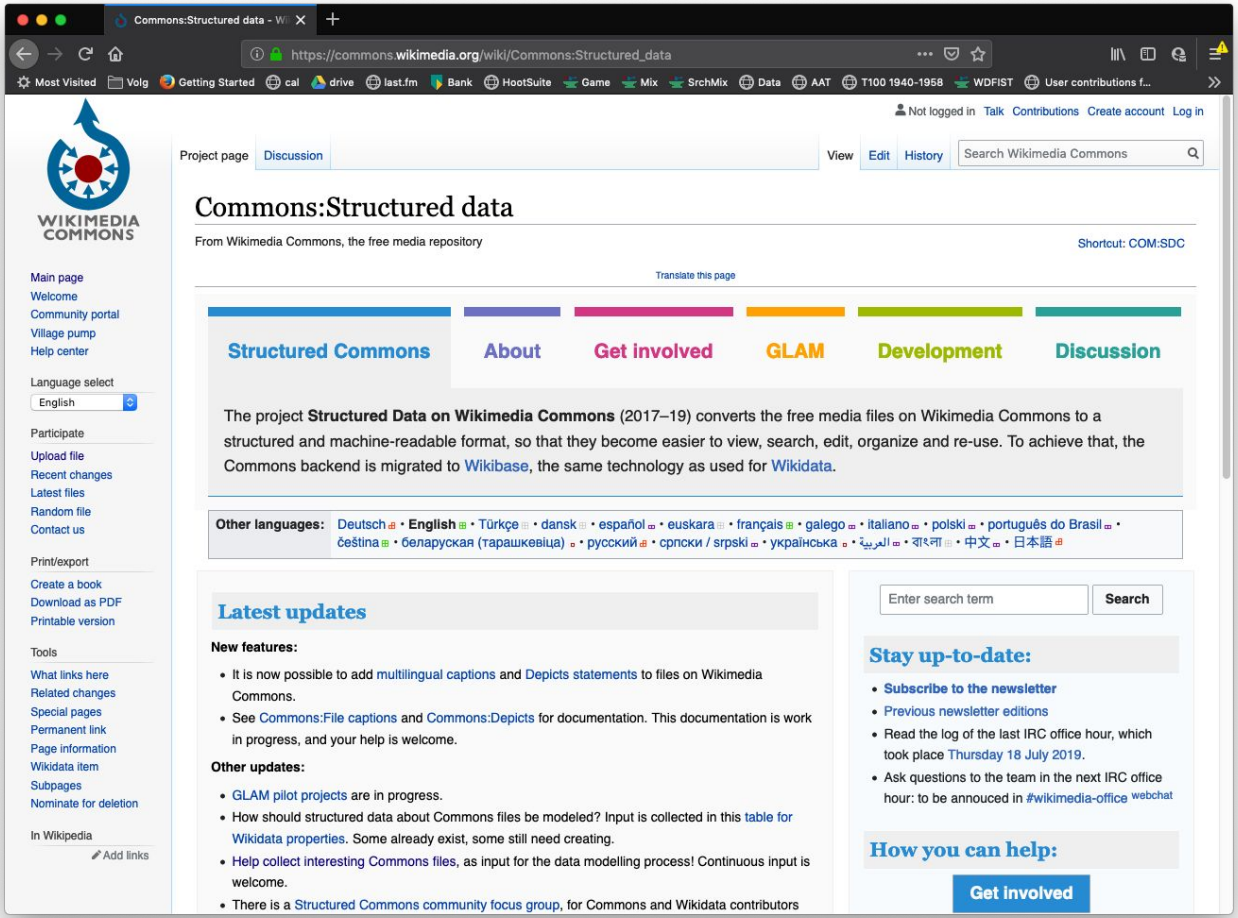
Pesquisar para adicionar elementos (gato doméstico, montanha, Taj Mahal, etc) de Wikidata

pescador	Marcar como proeminente
Rio Volta	Marcar como proeminente
onda capilar	Marcar como proeminente
água	Marcar como proeminente
canoa	Marcar como proeminente
rede de pesca	Marcar como proeminente

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.


Commons:Structured data - Wi
https://commons.wikimedia.org/wiki/Commons:Structured_data
Most Visited
Volg
Getting Started
cal
drive
last.fm
Bank
HootSuite
Game
Mix
SrchMix
Data
AAT
T100 1940-1958
WDFIST
User contributions f...
Not logged in
Talk
Contributions
Create account
Log in
Project page
Discussion
View
Edit
History
Search Wikimedia Commons
WIKIMEDIA COMMONS
Commons:Structured data
From Wikimedia Commons, the free media repository
Shortcut: COM:SDC
Translate this page
Main page
Welcome
Community portal
Village pump
Help center
Language select
English
Participate
Upload file
Recent changes
Latest files
Random file
Contact us
Print/export
Create a book
Download as PDF
Printable version
Tools
What links here
Related changes
Special pages
Permanent link
Page information
Wikidata item
Subpages
Nominate for deletion
In Wikipedia
Add links
Structured Commons
About
Get involved
GLAM
Development
Discussion
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.
Other languages: Deutsch • English • Türkçe • dansk • español • euskara • français • galego • italiano • polski • português do Brasil • čeština • беларуская (тарашкевіца) • русский • српски / srpski • українська • العربية • বাংলা • 中文 • 日本語
Latest updates
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
Enter search term
Search
Stay up-to-date:
Subscribe to the newsletter
Previous newsletter editions
Read the log of the last IRC office hour, which took place Thursday 18 July 2019.
Ask questions to the team in the next IRC office hour: to be annouced in #wikimedia-office webchat
How you can help:
Get involved

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

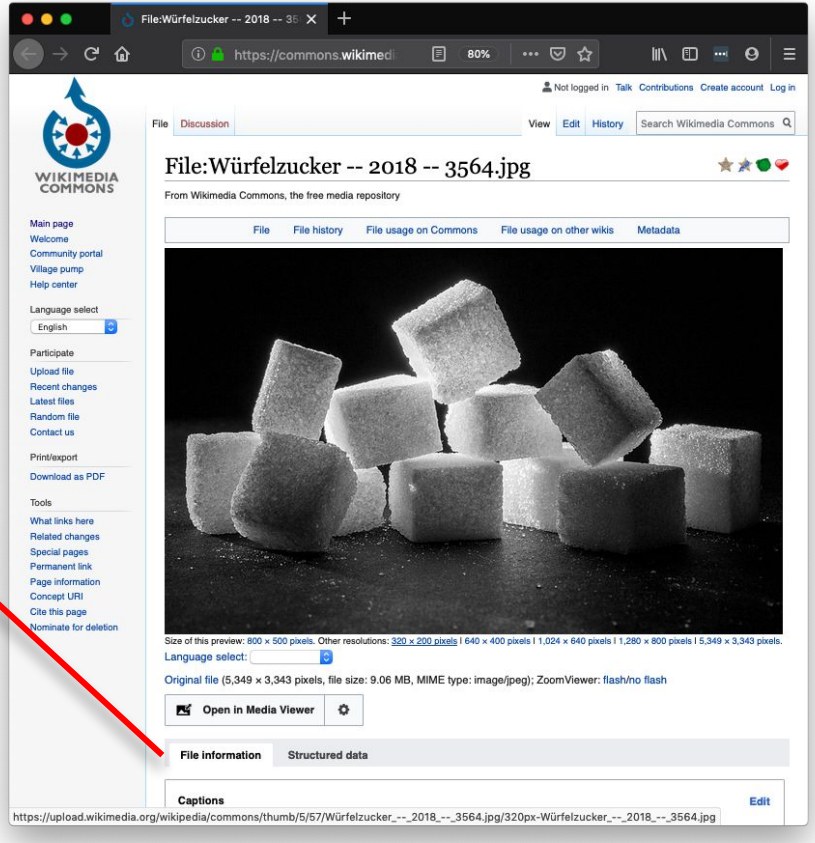
- Multilingual **file captions**
- **Depicts** statements with **Qualifiers**
- **Other statements** than Depicts

Add via file pages, UploadWizard, API

A simple example

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

Structured data tab:



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

Captions

File information | Structured data

Captions Edit

English	Sugar cubes (2018)
Nederlands	Suikerklontjes (2018)
Deutsch	Würfelzucker (2018)
русский	Кусковой сахар (2018)
français	Sucre en morceaux (2018)
polski	Kostki cukru (2018)

See fewer languages

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

Depicts

File information | Structured data

Items portrayed in this file Edit

depicts

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

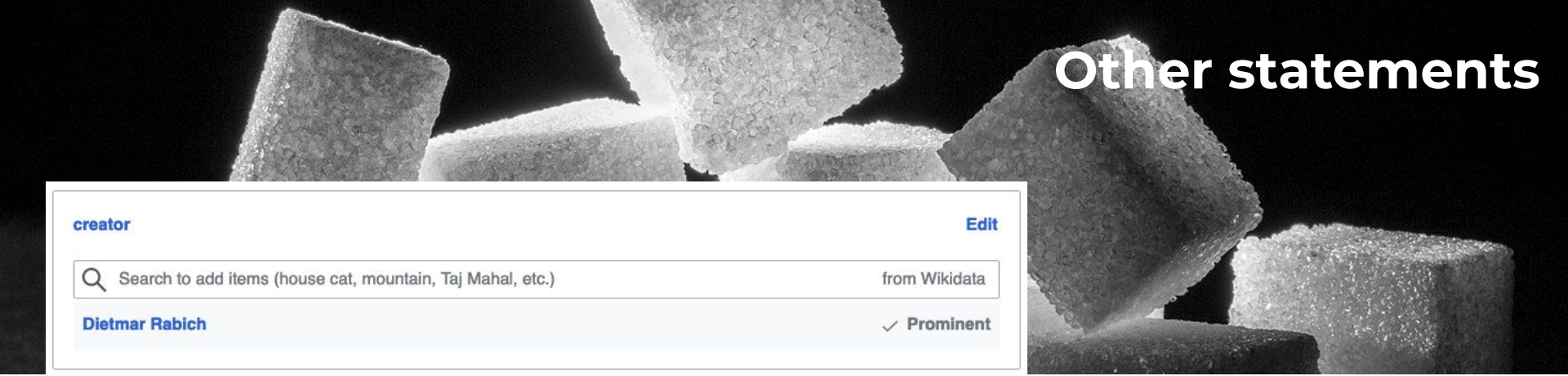
sugar cube ✓ Prominent

- quantity: 12
- color: white

FOUNDATION

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

Other statements



creator Edit

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

Dietmar Rabich ✓ Prominent

license Edit

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

Creative Commons Attribution-ShareAlike 4.0 International ✓ Prominent

Commons quality assessment Edit

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

Wikimedia Commons quality image Mark as prominent

Wikimedia Commons featured picture Mark as prominent

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

files are special

(creative work vs. digital representation)

Session on SDC hands-on training,
Sunday morning, room Strickland

GLAM pilot projects

- Content partnerships
- Community initiatives
- Tools / prototypes

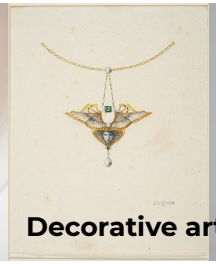
https://w.wiki/4QH

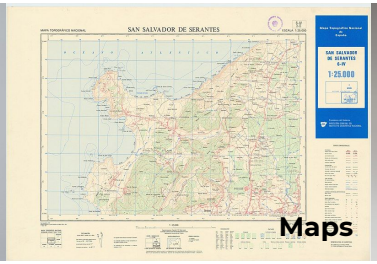
Alimdaihli, CC BY SA 4.0

GLAM pilot projects

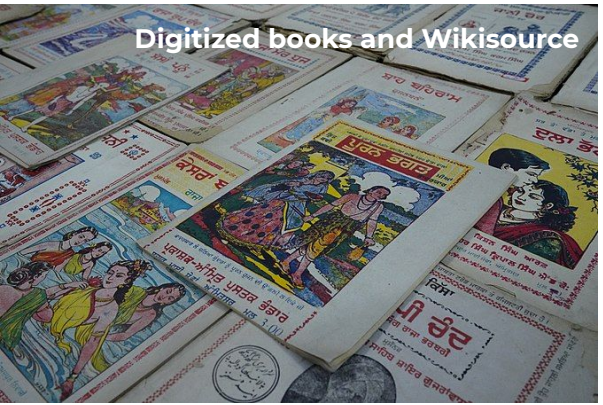
https://w.wiki/4QH

Content partnerships

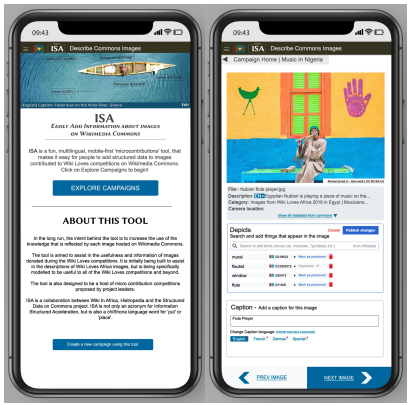




Community initiatives



Tools and prototypes



What's coming?

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

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

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 to 'put ' or 'place'.

campaign page

participation page

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

Wiki In Africa · histropedia · Wikimedia Commons · Wikimedia Foundation

A collaboration between Wiki In Africa, Histropedia and Structured Data on Commons.

Some first resources

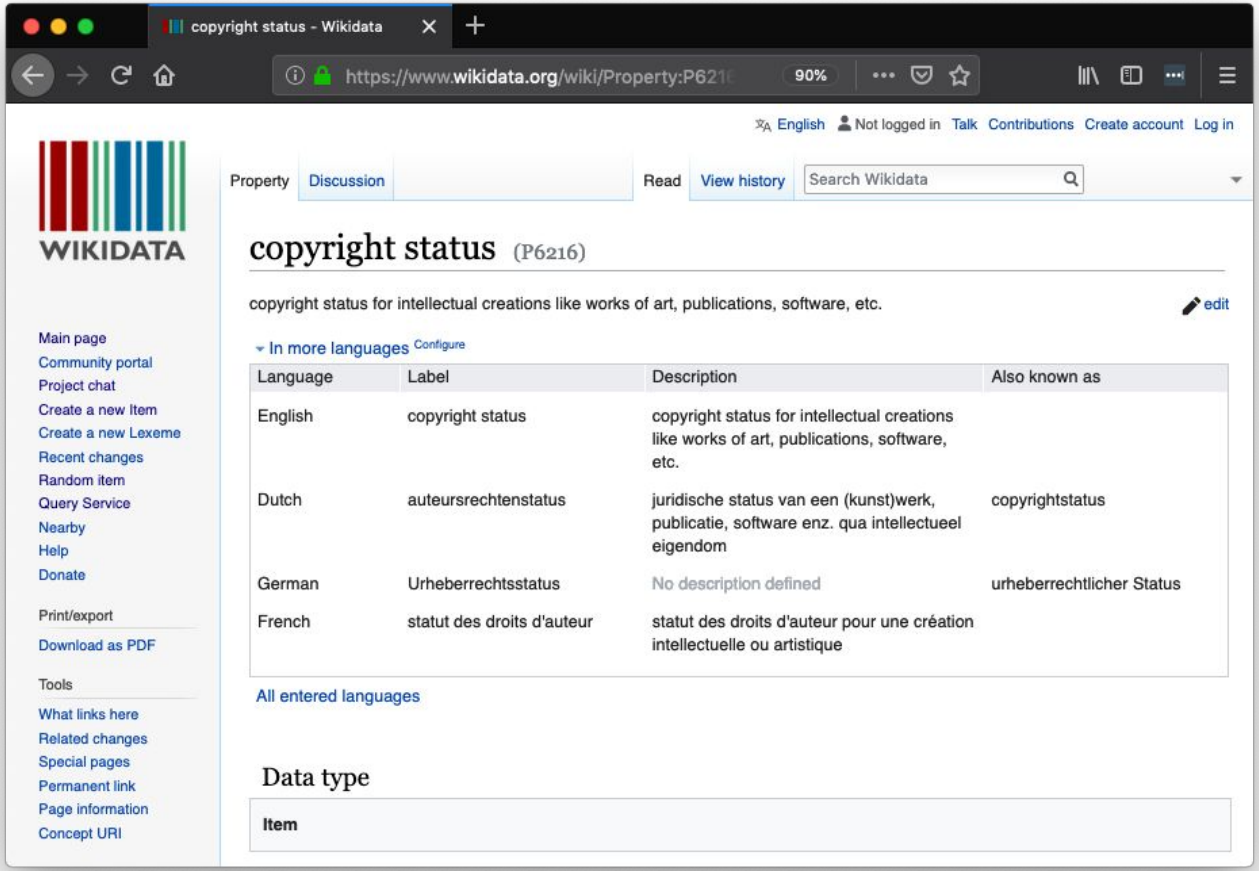
WIKIMANIA STOCKHOLM

Generic properties [edit source]

The properties below apply to all types of files.

Property / Aliases	Currently exists on Wikidata as	Data type	Qualifiers	Remarks
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)

WIKIMEDIA
FOUNDATION

Why start with SDC uploads in your GLAM-Wiki projects now?

- Multilinguality from the start
  (many more people will find the files)
- Uploads are much more re-usable
  (the new API is so much better!)
- Learning curve now, but will save you a lot of time with complex conversion work later

How can we help each other to get started?

- **GLAM & SDC pilot projects** will be documented
- Brand new **GLAMwiki documentation** on meta (September-December 2019)
- **What will help you learn the new things?**
  - Webinars?
  - In person courses?
  - Explainer videos?
  - Mentorship for your first 'new style' uploads?
  - How to best organize this?

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/

WIKIMEDIA
FOUNDATION

THANK YOU!

WIKIMANIA
STOCKHOLM