

Wikimedia Commons and **GLAM** needs around the world

Sandra Fauconnier - Jonathan Morgan



WIKIMANIA
CAPE TOWN

This presentation summarizes research that the Research team at the Wikimedia Foundation did in late 2017, asking GLAMs around the world how they use and experience Wikimedia Commons.



meta:Research:GLAM and Commons

Researchers: Jonathan Morgan and Niharika
Ved, Wikimedia Foundation



The research results are summarized at
https://meta.wikimedia.org/wiki/Research:GLAM_and_Commons

[https://etherpad.wikimedia.org/p/
Wikimania2018-Wikimedia_Common
s_and_GLAM](https://etherpad.wikimedia.org/p/Wikimania2018-Wikimedia_Commons_and_GLAM)



You can read notes taken for this Wikimania session at the etherpad:
https://etherpad.wikimedia.org/p/Wikimania2018-Wikimedia_Commons_and_GLAM

Regime visual transferable en tres temps (Juan Brozas), public artwork in Barcelona. Photo CC BY-SA 3.0 Spain, Dvdgmz



Why?

What were the reasons/incentives to do this specific research project?

Content diversity



credits/attributions
in final slides

Through collaborations with cultural institutions, since the mid 2000s, our communities have brought a huge diversity of interesting files to Wikimedia Commons.

Historical photos and audio; scientific specimens; digitized books and archival records; GLAM contribute lots of types of content to Wikimedia Commons (pushing the envelope)

[https://commons.wikimedia.org/wiki/File:BASA-RS-59-1-296-Nikola Obretenov, 1890.jpg](https://commons.wikimedia.org/wiki/File:BASA-RS-59-1-296-Nikola_Obretenov,_1890.jpg)

https://commons.wikimedia.org/wiki/File:Chromoblastomycosis_40x.jpg

<https://commons.wikimedia.org/wiki/File:Hebesfenomegacorona.jpg>

[https://commons.wikimedia.org/wiki/File:Annona_cherimola_Miller_\(AM_AK309529\).jpg](https://commons.wikimedia.org/wiki/File:Annona_cherimola_Miller_(AM_AK309529).jpg)

[https://commons.wikimedia.org/wiki/File:1056 Oslo. Utsikt fra Sjømannsskolen - no-nb digifoto 20151014 00002 blds_a_PK11792.jpg](https://commons.wikimedia.org/wiki/File:1056_Oslo._Utsikt_fra_Sj%C3%B8mannsskolen_-_no-nb_digifoto_20151014_00002_blds_a_PK11792.jpg)

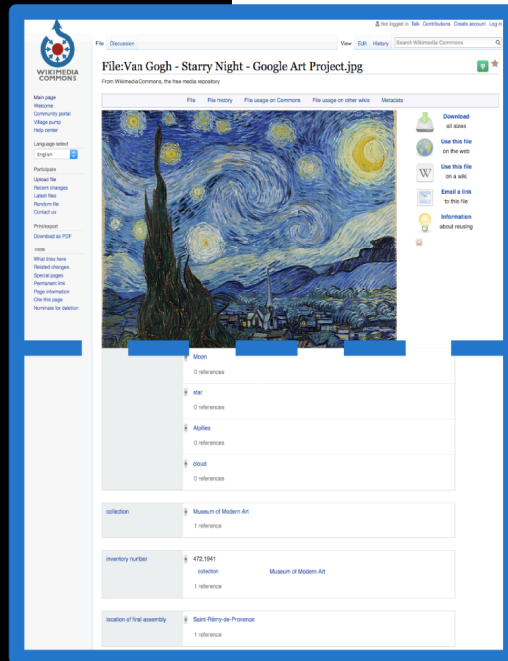
[https://commons.wikimedia.org/wiki/File:Mechelen, Belgium, Deventer, 1550.png](https://commons.wikimedia.org/wiki/File:Mechelen,_Belgium,_Deventer,_1550.png)

[https://commons.wikimedia.org/wiki/File:Abboekje voor lieve kinderen - met eene menigte prentjes en rijmpjes - PPN 852291590 - Image 11.jpeg](https://commons.wikimedia.org/wiki/File:Abboekje_voor_lieve_kinderen_-_met_eene_menigte_prentjes_en_rijmpjes_-_PPN_852291590_-_Image_11.jpeg)

https://commons.wikimedia.org/wiki/File:Avestruz_alta.jpg

Structured Data on Commons 2017-2019

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



File page



Data section



The Starry Night (1889) by Vincent van Gogh, Museum of Modern Art. From Google Art Project, Public Domain

It was also directly relevant to do research on GLAM use of Commons because in 2017-2019, Wikimedia Commons is enhanced with structured data from Wikidata. This is a good moment to check in with GLAMs to see what their needs are, and if/how they can be addressed with structured data.

How?



WIKIMANIA
CAPE TOWN

How was the research project performed?

Focus group (depth)

Interviews (diversity)

Survey (breadth)

Focus group (depth)

Get a bunch of people involved in GLAM projects in a room, ask them to describe their goals, process, pain points. This happened during a GLAM co-ordinators meetup in early 2017.

Interviews (diversity)

1-on-1 virtual sit-down with hand-selected people involved in GLAM projects, ask them questions about specific projects they worked on recently. Some questions are based on the issues that were raised during the focus groups. The interviews took place in Autumn 2017.

Survey (breadth)

Circulate a survey to a large audience with questions based on findings from the interviews and focus groups. A survey with more than 100 worldwide respondents was distributed late 2017.

What we learned



WIKIMANIA
CAPE TOWN

Research themes

1. Preserving important **metadata** about media items
2. Functionality and usability of **upload tools**
3. Monitoring activity and tracking **impact** after upload
- 4. Preparing media items** for upload
5. Working with **Wikimedia and Wikimedians**

Full report

meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions

When analyzing the responses in the focus groups, interviews and the survey, five main research themes emerged. They are described in detail on the report pages:

https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions

#	Gap or pain point for a GLAM contributor to Commons	Category	Product or Outreach?	Project Priority (5=high)	Point Person	In Structured Data on Commons scope?	SDC fix (if in scope)
1	Many GLAM metadata standards (vocabularies and ontologies) are not properly mapped on Wikidata yet	Social processes/Tech nology	Outreach	3	Sandra / Alex / CL (Developer relations)	In scope	Stimulate ontology mapping; inform tool builders about the usage of the mappings and how to do that
2	End-user documentation of monitoring tools is inadequate.	Documentation	Outreach, depending on Product	3	Sandra / Alex, dependance on Ramsey	In scope	Provide both better API / re-use data + we can probably build also some kind of reporting tools. Document those.
3	Some GLAM contributors to Commons experience unexpected file deletions, even when they believe they are following proper procedures	Social processes (copyright)	Both	2	(Ramsey)	Notifying users that OTRS exists = in scope. Facilitating the process of deletion and notification of deletion = out of scope.	Notify people that OTRS exists. Potentially in UploadWizard; deletion notifications
4	External reuse of Commons doesn't actually maintain connection between original files, attribution, and institutional metadata	Technology	Product	5	Ramsey	In scope	IIIF integration will address this.

Research findings breakdown, in large spreadsheet of 49 pain points around GLAM contributions to Commons (the 'List of Pain')



The Structured Data on Commons team read the research findings thoroughly and broke down the issues and pain points in a spreadsheet, lovingly called the 'List of Pain'. For each of the 49 pain points that GLAMs expressed, we investigated whether this pain point could be addressed with structured data on Commons, or not.

Personas

Developed by Niharika Ved and Jonathan Morgan.

Four 'archetypical' GLAM contributors to Commons:

1. **Alok**, University Professor from Delhi
2. **Sammy**, Wikimedian in Residence from Montréal
3. **Pat**, archivist from Budapest
4. **Nicholas**, Wikimedian from Monrovia



https://commons.wikimedia.org/wiki/File:GLAM_Personas_for_Structured_Data_on_Commons_v2.pdf

University Professor Alok

Age: 48 years old
 Education: B.Ed and M.A in History
 Languages: Hindi (fluent), English (proficient)
 Occupation: History Professor
 Location: Gurugram, Haryana, India

About
 Alok is a professor of Postcolonial History at the University of Delhi. Alok is also a Wikiversity editor and a Wikimedia Education Program partner. He believes in the power of free education to promote knowledge equality and empower Indian students to take an active role in shaping the story of their history and culture for posterity.

GLAM Involvement
 Organizing collections, digitizing content and uploading University owned historical content on Commons. Alok organizes field trips with students to teach them how to search and record meaningful historical content to be uploaded on Commons.

Goals	Activities	Challenges
<ul style="list-style-type: none"> Use of scalable upload carousels Documentation and guides to find and use upload tools Teach students to do research, help them gain digital literacy and digitization skills 	<ul style="list-style-type: none"> Carry out site visits with students to teach them how to record and upload content. Digitizing content from museum archives. Uploading historical photos and pictures of 16th century sculptures. 	<ul style="list-style-type: none"> Unclear of which licensing templates to use. Manual labor intensive task to search for the right upload tool to be used. Manual labor intensive task to search for the right upload tool to be used.

Editing Journey

The graph shows the relationship between **ATTITUDE** (Challenge and Confidence) and **STAGE** (PRE UPLOAD, DURING UPLOAD, POST UPLOAD, IMPACT).

- PRE UPLOAD:** Conducts field trips to collect content. Teachers receive how-to check and ready content for upload. (Challenge is high, Confidence is low)
- DURING UPLOAD:** Copyright empowers in consulting and their willingness to set up. (Challenge is high, Confidence is low)
- POST UPLOAD:** The uploaded content can be easily retrieved. It is easy to show where the images are stored. (Challenge is low, Confidence is high)
- IMPACT:** Future of uploads can be tracked to show how content is used to organizations. (Challenge is low, Confidence is high)

Reference: https://commons.wikimedia.org/wiki/File:GLAM_Personas_for_Structured_Data_on_Commons_v2.pdf

Niharika Ved, design intern at the Wikimedia Foundation, also created a set of four GLAM personas - archetypical GLAM contributors to Wikimedia Commons, based on the kinds of people we interviewed and surveyed.

You can find them at

https://commons.wikimedia.org/wiki/File:GLAM_Personas_for_Structured_Data_on_Commons_v2.pdf

Pain points addressed with structured data on Commons



WIKIMANIA
CAPE TOWN

The following three categories of pain points, experienced by GLAMs on Wikimedia Commons, can in some ways be addressed with structured data on Commons.



Demonstrate impact of contributing to Wikimedia Commons

[Specimen of ostrich skeleton](#), Museum of Veterinary Anatomy FMVZ USP /
Wagner Souza e Silva, CC BY-SA 4.0

After they contributed media files to Wikimedia Commons, GLAMs must be able to prove the value and impact of that work to their directors, boards, and funders. In many cases, GLAM representatives asked for better statistics and metrics tools and more detailed metrics.

Structured Data on Commons will bring more refined APIs that, among other things, will also allow more precise measurement and statistics of how (and how often) sets of media files are used.

"There are only a few tools to measure and monitor it but still on a very basic level."

Multilinguality, to make Commons uploads more findable and usable

[Abé-boekje voor lieve kinderen : met eene menigte prentjes en rijmpjes.](#) Published 1831 by Hendrik Frijlink, Amsterdam.
Koninklijke Bibliotheek / National Library of the Netherlands, Public Domain



Many GLAMs contribute media files for which there is no English description available at all. This makes it difficult to upload files with the right categories, and to make files findable.


Structured Data on Commons will integrate Wikidata into Wikimedia Commons file metadata, and the multilingual nature of Wikidata will make it much easier to contribute to (and search) Wikimedia Commons in a language that is not English.

Categories (+): Images from the Bulgarian Archives State Agency | Signatures of people from Bulgaria (-) (s) (i) (t) | Nikola Obretenov (-) (s) (i) (t) | BASA-Rousse-59-1-296 (-) (s) (i) (t) | (+)

Hidden categories: Artworks without Wikidata item | CC-PD-Mark | PD Old | Files with no machine-readable author

From Wikimedia Commons, the free media repository

Download Use this file Use this file Email a link Information



Size of this preview: 372 × 595 pixels. Other resolutions: 149 × 240 pixels | 298 × 480 pixels | 372 × 600 pixels | 800 × 1,289 pixels. Add a note

Original file (800 × 1,289 pixels, file size: 878 KB, MIME type: image/jpeg); (ⓘ request rotation)

Open in Media Viewer

Document creation place: Русе, България (Rousse, Bulgaria)

Description	Български: Снимка на Никола Обретенов с автограф. Русе, Полуфигура, Цивилно облекло. English: A photo of Nikola Obretenov with an autograph
Author	фотографско ателие на Франц Байер
Date	14 February 1890
Dimensions	13x21 cm
Current location	Archives State Agency ⓘ ⌵ State Archives - Rousse
Accession number	<ul style="list-style-type: none"> Fonds: 59 Inventory: 1 Archival unit: 296 Sheet no.: 1 Webpage: http://212.152.187.196:84/P/posna.aspx?type=documents&app=50&fpid=5977429&lgid=61107273&segid=6778715&dgid=2827

Data wrangling

'Mapping' data from GLAM databases to Commons is hard
Knowing about, and understanding Commons categories and templates
Lots of 'hacking'

[Photo of Nikola Obretenov with autograph](#), 1890, photographic studio of Franz Bauer. Bulgarian Archives State Agency, Public Domain

Many GLAMs report that they find it difficult, and very time-consuming, to translate their own metadata to the templates on Wikimedia Commons.

Structured Data on Commons will provide more refined APIs which should help with mapping metadata.

**"The categorization system
on Wikimedia Commons
requires insider knowledge."**



**WIKIMANIA
CAPE TOWN**

"We appreciate the many efforts of Magnus Manske and of many other developers to let us use tools like GLAMorous, Glamorgan, Baglama2, Mix'n'Match, and Locator tool. We would like to thank all them. We need more people involved in developing and refining GLAM tools, and also new tools. We would also thank the many WM Commons users who are already helping us to fix metadata of 17.000 images and locations of 8.000 of them."



Many GLAM contributors express that it's difficult to learn about (the very presence of) useful tools. But when they have found them, they are often very grateful that they exist.

Other pain points

(not related to structured data)

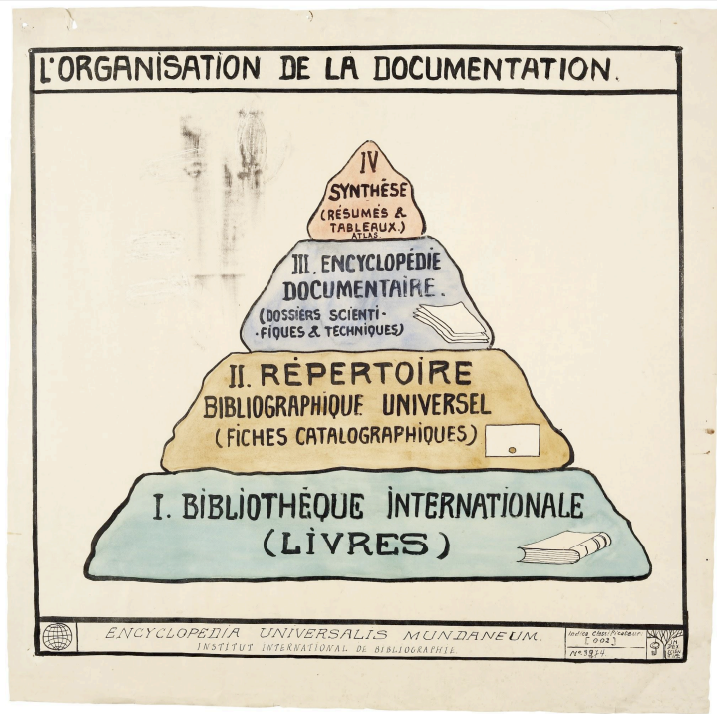


WIKIMANIA
CAPE TOWN

The research results also unearthed several pain points and issues that can't be addressed with structured data on Commons.

Better and more
easy to find
documentation
needed

[L'organisation de la documentation](#),
from the Encyclopedia Universalis Mundaneum, Paul Otlet,
1936, Mundaneum, Mons, Belgium; Public Domain



Firstly, GLAMs - especially newcomers - get lost in the beginning and ask for clearer, up to date, and easy to find documentation.

INDONESIA REPORT

From Eastern Indonesia on Commons

By Annayu Maharani & Rachmat04

Contribute — Share this [show]

Introducing open license & digitization at Dr. H.C. Oemboe Hina Kapita Museum, local museum in East Sumba

In Indonesia, The Law of Cultural Advancement No.5/2017 serves as a mandate to the government to improve public policy for the cultural sector. The law has become a tool to drive a healthy arts ecosystem, including the management of cultural institution like a museum. Therefore, this is a breakthrough to conduct file-sharing activity in a museum in Indonesia within the framework of GLAM-Wiki.

Wikimedia Indonesia in cooperation with [Indonesian Arts Coalition](#), a cultural policy and arts advocacy organization, organized a digitization activity at Dr. H.C. Oemboe Hina Kapita Museum, Waingapu, East Nusa Tenggara. On this occasion, Wikimedia Indonesia introduced how Wikimedia Commons and Creative Commons license could be used as the medium of knowledge dissemination in their collections.

The [two-day event](#) (30–31 May 2018) was divided into presentation and digitization sessions. On the first day, Indonesia Art Coalition shared about the change of the arts ecosystem in the digital era and the importance of cultural data, as stipulated in Law on Cultural Advancement. Further, Wikimedia Indonesia explained how to use Creative Commons and various types of license. Several cultural practitioners alongside with the local government from Office of Culture and Tourism Agency and Public Library and Archive were seen to come. The proposed audience at this event was the museum staffs, therefore they could conduct a better knowledge management through the museum collections, including weft ikat, pieces of jewelry and ritual types of equipment.

Sample collections from Dr. H.C. Oemboe Hina Kapita Museum



Copper bowl (top view)

Copper bowl (side view)

Ceramic bowl (top view)

Ceramic bowl (bottom view)

Copyright

Many GLAMs don't know much about it
And even less about our policies
Unaware of OTRS
Find OTRS very opaque

Indonesia Report: from This Month in GLAM, June 2018

https://outreach.wikimedia.org/wiki/GLAM/Newsletter/June_2018/Contents/Indonesia_report

Many GLAMs are not knowledgeable enough on the topic of copyright and get in trouble when attempting to upload files for which copyright is insufficiently cleared. When they start working with OTRS for that purpose, they find that process opaque and slow.

Metadata from GLAMs

sometimes very minimal
sometimes only available in analog
form
legacy software
in many places and many formats

[Mangkok Keramik](#), from Dr. H.C. Oemboe Hina Kapita
Museum, Kota Waingapu, Indonesia, CC BY-SA 4.0



Category (+): Images from Museum Dr. (H.C) Oemboe Hina Kapita (-) (1) (1) (+)
Hidden categories: CC-BY-SA-4.0 Self-published work
From Wikimedia Commons, the free media repository

[Download](#) [Use this file](#) [Use this file](#) [Email a link](#) [Information](#)

[Done](#) [Show/hide preview](#)
No galleries found.



Size of this preview: 506 × 600 pixels. Other resolutions: 202 × 240 pixels | 405 × 480 pixels | 645 × 768 pixels | 864 × 1,024 pixels | 1,792 × 2,124 pixels.
Add a note.
Original file (1,792 × 2,124 pixels, file size: 964 KB, MIME type: image/jpeg):
([p](#) request rotation); ZoomViewer: [flash](#)/[no flash](#)

[Open in Media Viewer](#)

Summary [[edit](#) | [edit source](#)]

Description	Bahasa Indonesia: Mangkok Keramik
Date	31 May 2018, 12:54:48
Source	Dr. H.C. Oemboe Hina Kapita Museum ↗ ↘
Author	Gusti/Museum Daerah Sumba

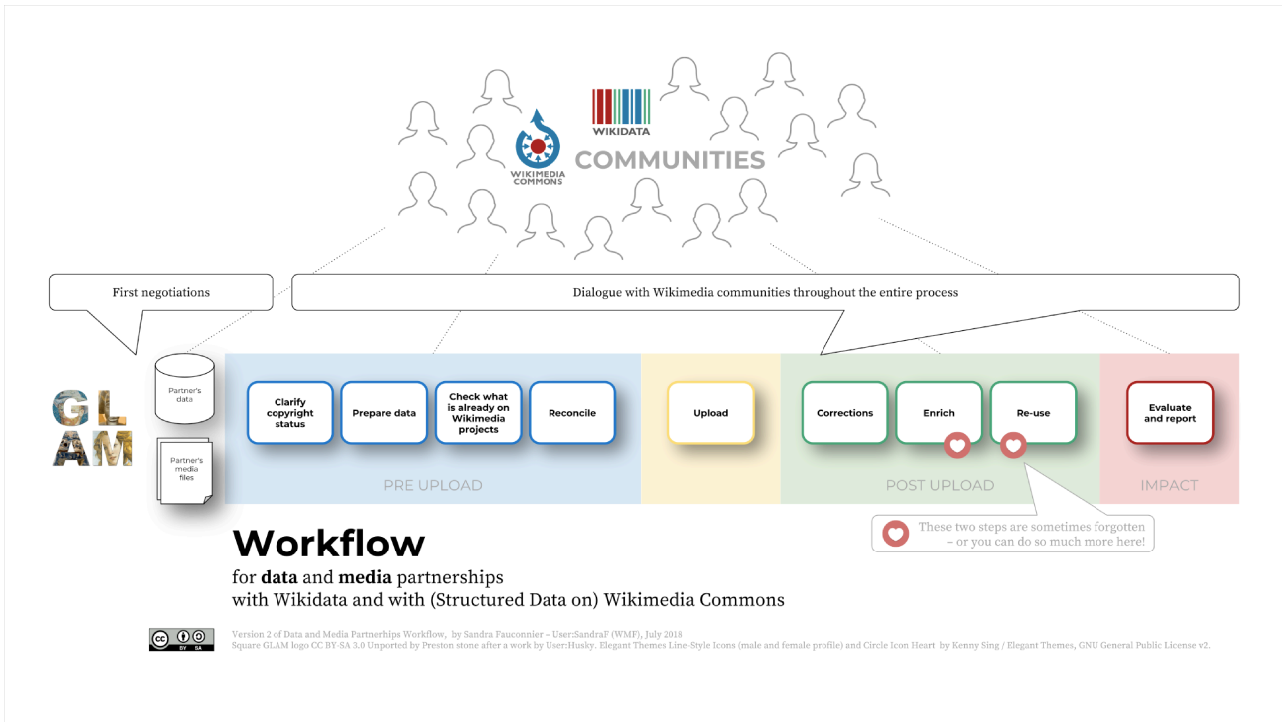
Especially (but not only!) in the developing world, many GLAM collections only have very minimal metadata, or no metadata at all. This is also an opportunity: our communities possess a lot of knowledge and can help with enriching files that are only minimally described upon upload.

Next steps



WIKIMANIA
CAPE TOWN

What will we do to follow up on these research results and insights?



Firstly, we have started thinking about how to organize (new) documentation in a better way. Structured data as a new technology on Commons will need new documentation, which is a good opportunity to re-think how we document things.

This workflow diagram is a first step:

https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow



GLAM pilot projects

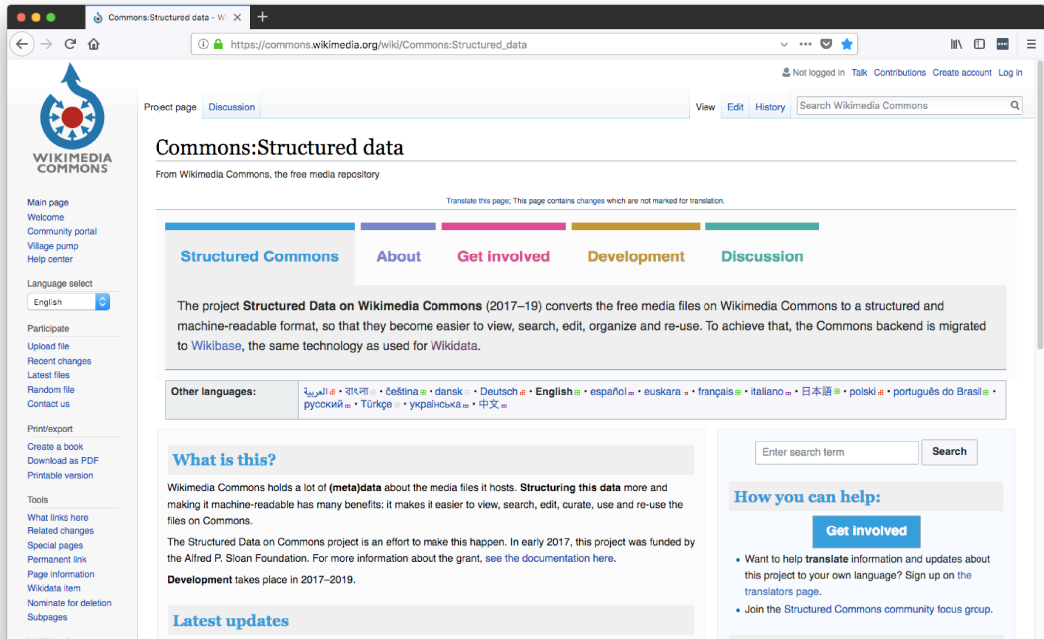
- Diverse and representative
- Support and documentation

In Summer-Autumn, we are starting to define the first GLAM pilot projects that will use structured data on Wikimedia Commons. We want to make sure that they are diverse, and representative for the different kinds of GLAM projects that the Wikimedia communities perform.

These projects will be executed in Spring-Summer 2019, when Structured Data on Commons is deployed, and will be well documented so that everyone can replicate and be inspired by them.

File credits (if not mentioned in previous slides)

- [Micrograph of chromoblastomycosis showing sclerotic bodies in dermis](#). Department of Pathology, Calicut Medical College, CC BY-SA 4.0
- [Different views of the hebesphenomegacorona](#). Matematica (IME/USP)/Rodrigo Tetsuo Argenton, CC BY-SA 4.0
- [Annona cherimola Miller](#), by Ewen Cameron, 2010. Auckland War Memorial Museum, CC BY 4.0
- [Mechelen, België, uit Atlas des villes des Pays-Bas](#), ca. 1558, Jacob van Deventer. Koninklijke Bibliotheek, Public Domain
- [Bildet er hentet fra Nasjonalbibliotekets bildesamling Ekeberg,Sjømannsskolen, Oslo, Oslo](#), Unknown date, National Library of Norway, Public Domain
- [Sandwich terns brooding](#), 2013, by Marc Plomp / Stichting Natuurbeelden, CC BY-SA 3.0 Netherlands



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

This is the information portal for Structured Data on Commons. Here you can find information about the project in general, and all the community consultations. You can subscribe to the newsletter here. Input on community consultations is very welcome!

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

Enkosi!

sandraf@wikimedia.org

