Structured Data on Wikimedia Commons and knowledge equity

Wikimania 2018, Cape Town



Etherpad

https://is.gd/sdequity



What is happening? When? Upcoming work Benefits

How can this advance knowledge equity?

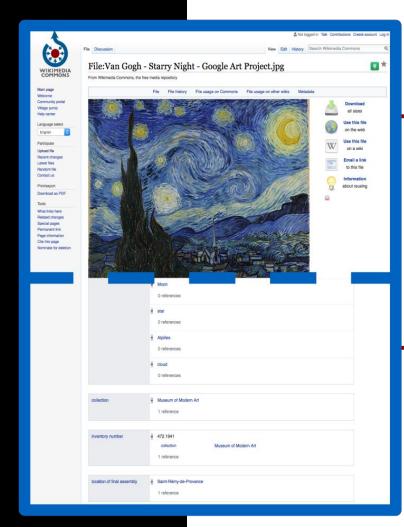


Structured Data on Commons 2017-2019

adding metadata on Commons in a **structured** & **machine-readable** format making Commons files easier to **view**, **search**, **edit**, **organize** and **re-use**, in many **languages**



What are we doing?



··File page



Data section



Statements Properties Items

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

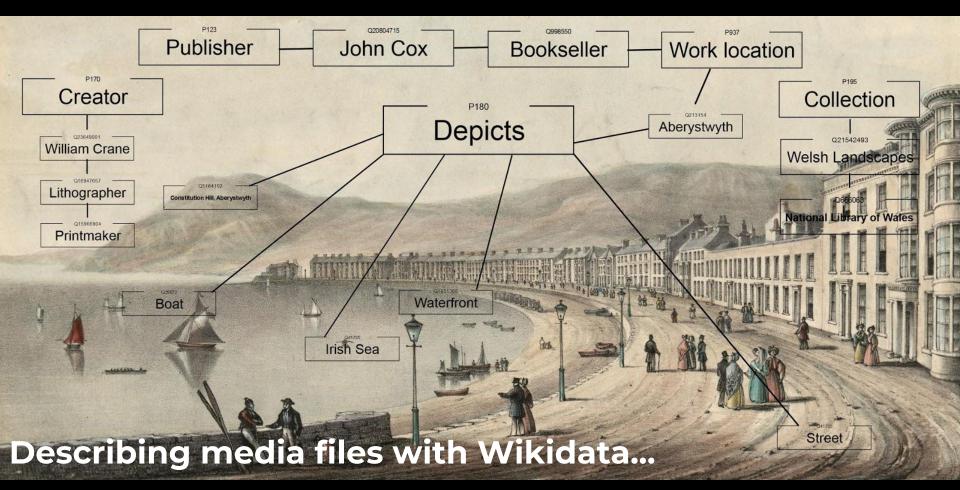




Photo by Jason Krüger for Wikimedia Deutschland, 20 Ap 2018, CC BY-SA 4.0

What we're working on

We worked / are working on

- Multi-content revisions
- Designs and prototypes for
 - file pages
 - UploadWizard
 - search
- Prepare for GLAM pilot projects

~ October 2018:

multilingual captions for files

~ Early 2019:

'depicts' and other structured data



Main page Community portal Current events Recent changes Random page

Donate Tools

What links here Related changes Special pages Printable version Permanent link Page information

In other languages Add links

File:Fullsizeoutput 235.jpg

File File history File usage Metadata



Size of this preview: 600×600 pixels. Other resolutions: 240×240 pixels I $2,000 \times 2,000$ pixels. Original file ($2,000 \times 2,000$ pixels, file size: 536 KB, MIME type: image/jpeg)

Summary [edit]

 Description
 English: A wikitext description

 Date
 2018-07-19

 Source
 Own work

 Author
 Wubwub

Licencing [edit]

I, the copyright holder of this work, hereby publish it under the following licence:

This file is licenced under the Creative Commons Attribution-Share Alike 4.0 Internationat@ licence.

Licencing [edit]

I, the copyright holder of this work, hereby publish it under the following licence:

This file is licenced under the Creative Commons Attribution-Share Alike 4.0 International legislation.

You are free:

. to share - to copy, distribute and transmit the work

w:en:Creative Commons

attribution share alike to remix – to adapt the work
 Under the following conditions:

- attribution You must attribute the work in the manner specified by the author or licencer (but not in any way that suggests that they endorse you or your use of the work).
- share alike If you alter, transform, or build upon this work, you may
 distribute the resulting work only under the same or similar licence to this
 one.

File history

Click on a date/time to view the file as it appeared at that time.

	Date/Time	Thumbnail	Dimensions	User	Comment
current	14:26, 19 July 2018	6	2,000 × 2,000 (536 KB)	Wubwub (talk I contribs)	User created page with UploadWizard

Description

You cannot overwrite this file.

File usage

There are no pages that link to this file.

MediaInfo:M10306

File:Fullsizeoutput 235.jpg

No description defined

In more languages

In more languages
 Language Label

British English No label defined No description defined
English Rita the structured data on No description defined

Commons bee, an amigrumi

Statements

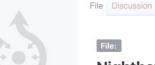
+ add statement







View Edit History ☆ More ▼ Search Wikimedia Commons

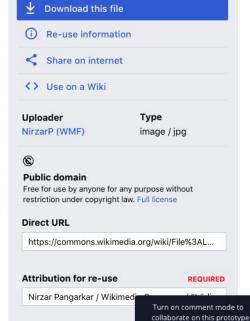


COMMONS

Nighthawks by Edward Hopper 1942.png



Nighthawks is a 1942 oil on canvas painting by Edward Hopper that portrays people in a downtown diner late at night. It is widely considered Hopper's most famous work, and one of the most recognizable paintings in American art.



File information

Metadata







Existing file page as is

Turn on comment mode to collaborate on this prototype



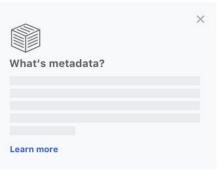






Metadata





material used	■■ P186 • Edi
Q. Search to add value	from wikidata
oil paint	Q6955 • ✓ Primary
canvas	Q4259 • Make primary

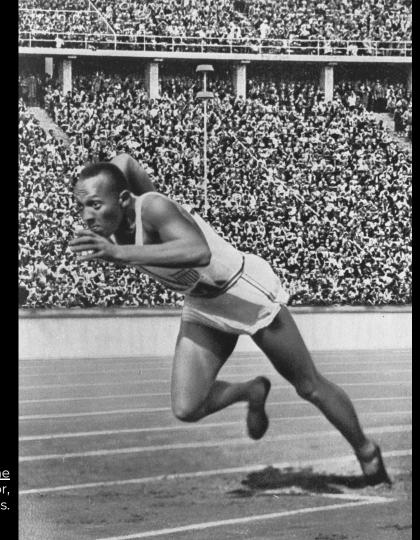
Turn on comment mode to collaborate on this prototype







Upcoming, now and later in 2018-19



Jesse Owens at start of record breaking 200 meter race during the Olympic games 1936 in Berlin (photographic montage). Unknown author, 1936, Public Domain, from the United States Library of Congress.



Properties for structured data on Commons

- → Community consultation runs NOW. Participate!
- → The basic properties that will be available for media files



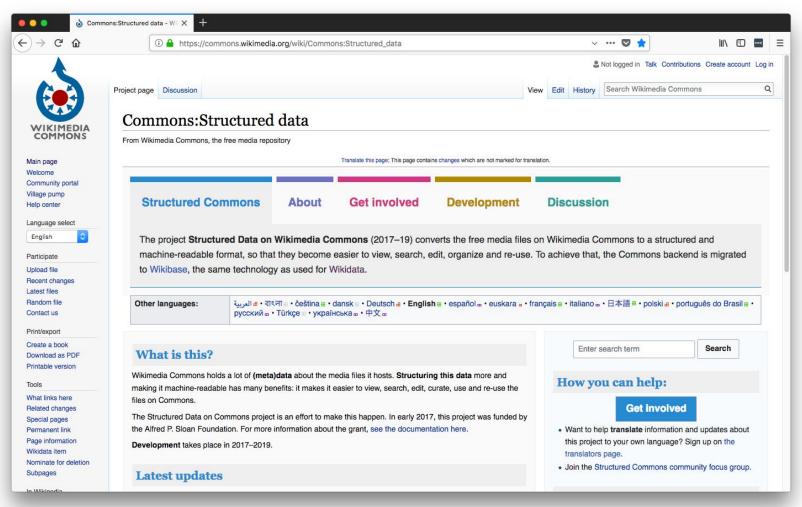


GLAM pilot projects

- → Diverse and representative
- → Support and documentation



Support for tool developers

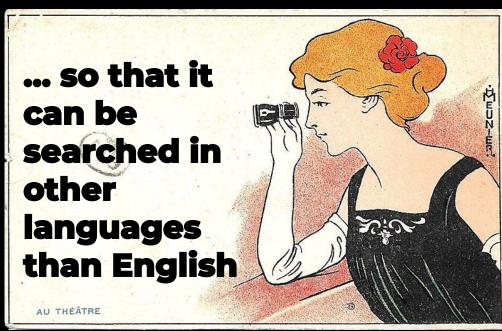


Benefits for Wikimedia Commons?

(With some examples of what people are currently already doing with Wikidata...)

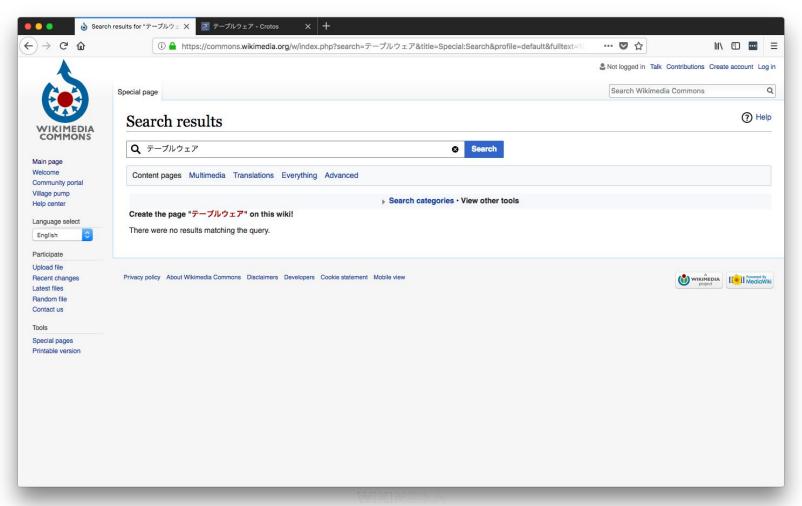




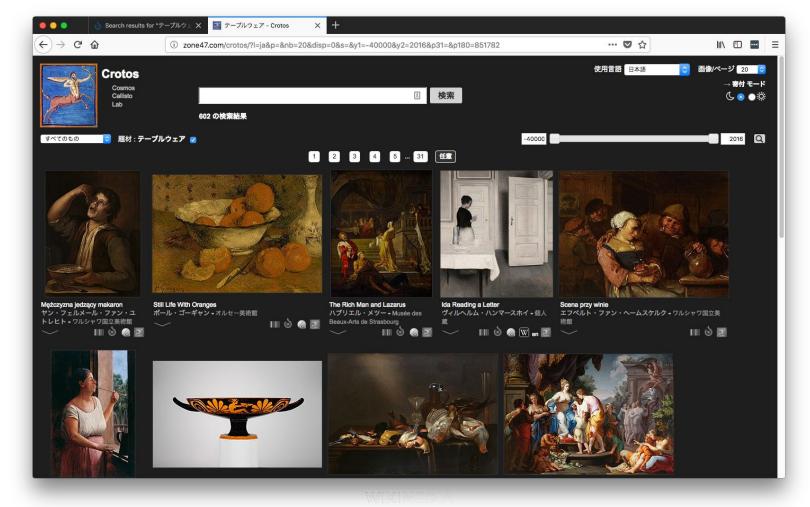


<u>Marten van Valckenborgh: The Tower of Babel</u> (1595), Gemäldegalerie Alte Meister, Public Domain

Henri Meunier: <u>Au theâtre</u>, before 1922, Public Domain

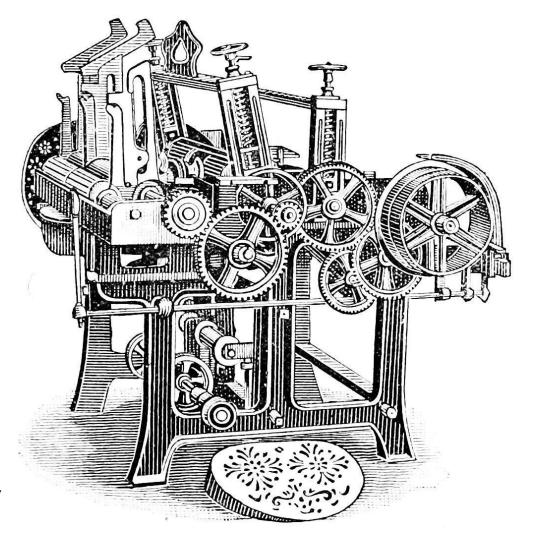


'tableware' in Japanese - テーブルウェア

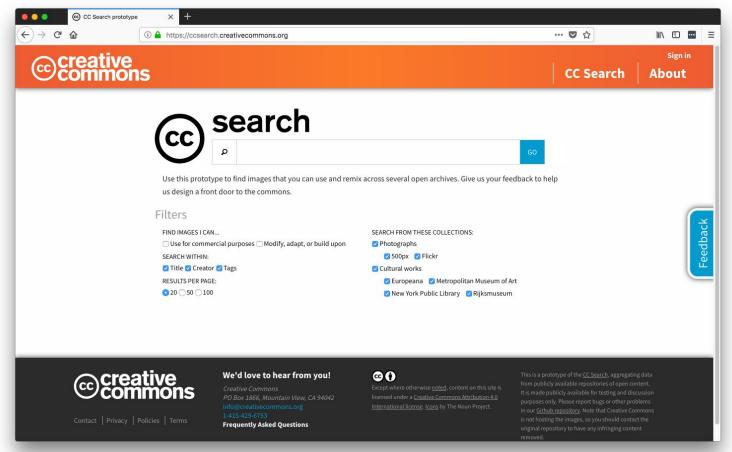


http://zone47.com/crotos/?l=ja&p=&nb=20&disp=0&s=&y1=-40000&y2=2016&p31=&p180=851782

Machinereadable

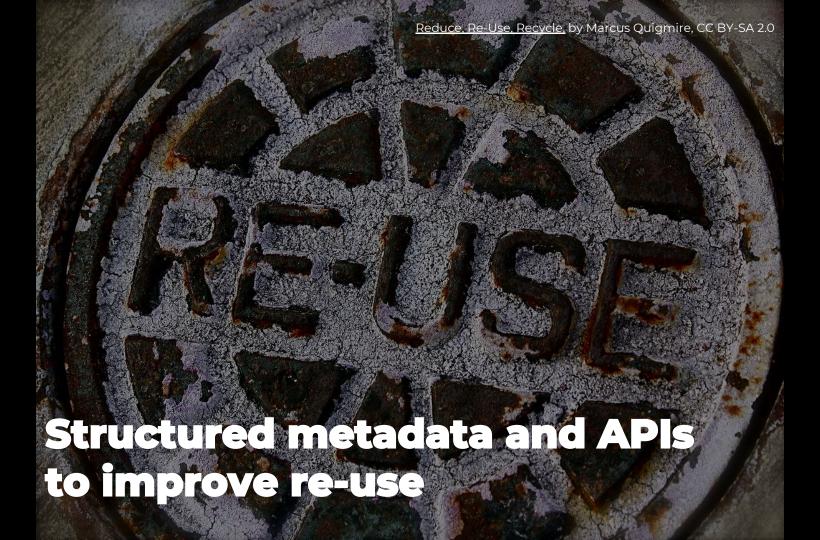


<u>A gilling machine</u>, from Popular Science Monthly Volume 39, 1891. Unknown author, Public Domain

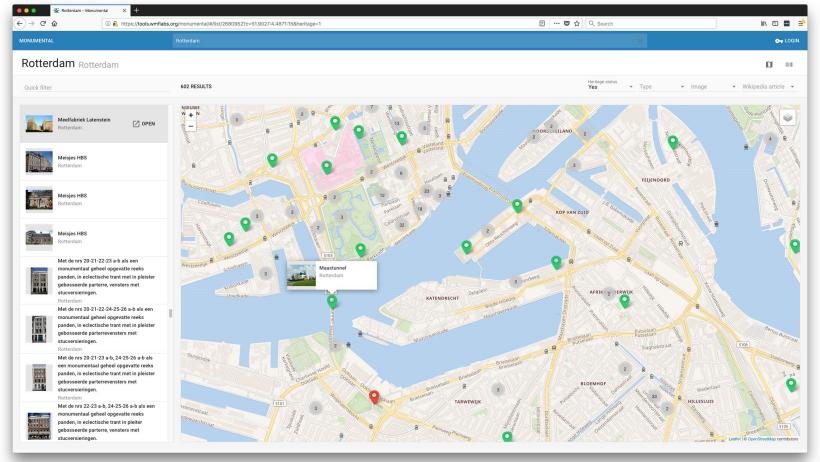


http://ccsearch.creativecommons.org/

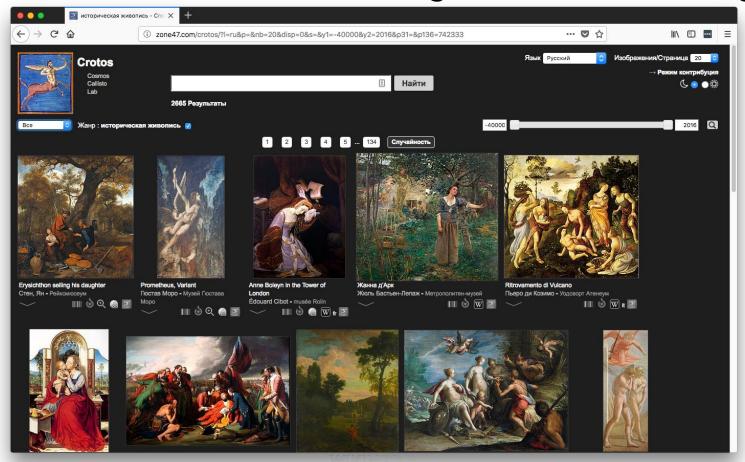
By becoming compatible with structured data used by other websites, we can be part of more discovery platforms

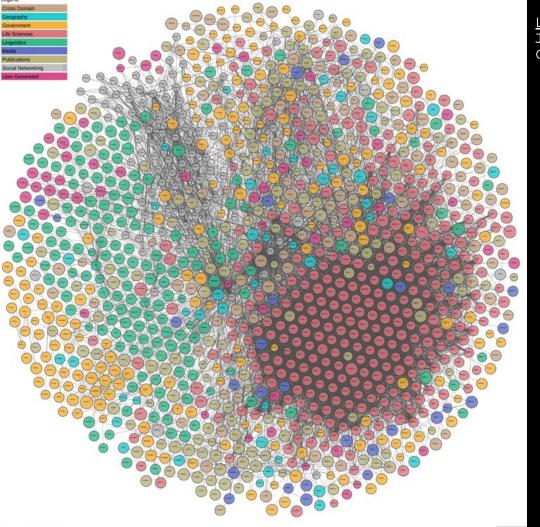


More value and potential of Commons' media files beyond Wikipedia



... where Wikidata enables multilingual search and browsing





<u>Linked Open Cloud diagram as of 30 May 2018</u>, Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak / contributors to lod-cloud.net, CC BY-SA 3.0

Linked data -Wikimedia's data connected with other datasets around the world

(c) (i)



tableware (Q851782)

Reasonator

食器 | 食具 | أواني | الاواني | آنية المائدة | Eetgerei | eet- en drinkgerei | Посудина | Vaisselles | stoviglie | Speiseservice | Kaffeeservice | Tafelservice | Kaffeegeschirr | Tafelgeschirr | Teeservice | dishes | dish | dishware | jídelní nádobí

plates used for setting a table and serving food

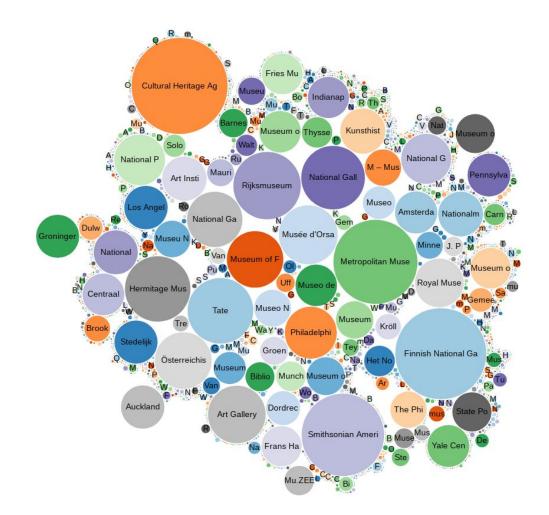
Other pro	perties	~
described by source		
subclass of	manufactured object in the strict sense, object made by humans with their hands	
	product anything that can be offered to a market that might satisfy a want or need	
	kitchenware object used in the kitchen	
part of	eating ingestion of food to provide for all organisms their nutritional or medicinal needs	
	everyday culture customs, habits, and everyday objects that are not perceived as a culture in terms of visual arts, music and literature	
has part	plate type of dish	
	cutlery eating utensils	
	teacup tableware: cup, with or without a handle	
	cup vessel intended for an individual to use for drinking wine, water, or other beverage	
topic's main category	Category:Tableware Wikimedia category	



External sources	~
AAT	300043196
BabelNet	00075834n
BNCF Thesaurus	566
EAGLE	objtyp/lod/1780
GND	4124863-6
Iconclass notation	41C31
JSTOR topic	tableware
PACTOLS thesaurus	pcrtM4tR8ZGTvw
PSH	10251
Quora topic	Dishes-and- Tableware
US National Archives	10642939

Wikidata's WikiProject Sum of all Paintings...

All the museum collections which have paintings in them with the count per location http://tinyurl.com/zqjz6q2



Querying & bringing together (thematic) information that may not have been combined before (across & also outside heritage collections!)



How can this advance knowledge equity?

(With some examples of what people are currently already doing with Wikidata...)



More potential for community improvements to media files

... even if they're minimally described when uploaded

File:UvA-BC 300.413 - Siboga - inheemse bewoners van Timor op een rif aan de oostkant van het eiland.jpg

Categories (++): Photographs of the Siboga Expedition (-) (±) (1) | Portuguese Timor (-) (±) (1) (↑) | Valu Beach (-) (±) (1) (↑) | People of East Timor (-) (±) (1) (↑) | Fataluku (-) (±) (1) (↑) | (+) | Hidden categories: Images from the Special Collections of the University of Amsterdam (-) (±) (1) (↑) | Files from the Expedition Wikipedia project | Template Unknown (author) | Anonymous-EU | Files with no machine-readable author | Files with no machine-readable source

From Wikimedia Commons, the free media repository

om wikimedia Commons, the free media repositor



<u>Fatalulu on the East coast of Tutala (Timor). Siboga Expedition</u> <u>1899-1900.</u> Public Domain, Special Collections of the University of

Better crowdsourcing

with data that can be given back to an institution in a format that works for them



On Sunday October 4th, World Animal Day, the Rijksmuseum will host a birdwatching day in collaboration with Naturalis Biodiversity Center, Wikimedia Netherlands and the COMMIT SealincMedia project.

Rijksmuseum – Naturalis Biodiversity Center – Wikimedia Nederland & COMMIT/ SEALINCMedia present a unique birdwatching event

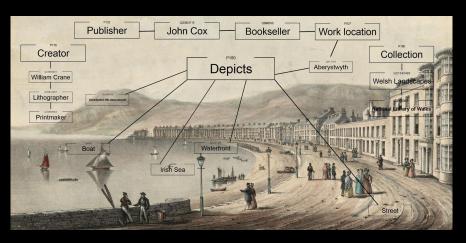
Birds are everywhere. In your own garden, in nature, and also in art. Among the Rijksmuseum's 1,2 million collection objects are many prints, paintings and artefacts that have bird species depicted on them. Among the 37 million objects in the Naturalis collection are many birds from all over the world that have been collected in the last 200 years, as well as historical drawings of plants and animals in which many birds are depicted.

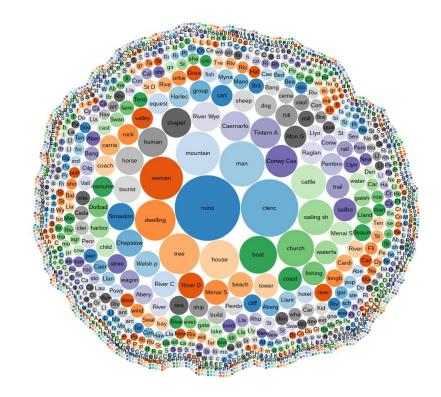
Wikimedia Commons

Some of the depicted birds are easily identified. Others require a trained eye to determine which species the artist has pictured. The Rijksmuseum and Naturalis are looking for experienced bird watchers and other avian enthusiasts to join an expedition through their digital collections and help the museums identify bird species in works of art.

The Rijksmuseum and Naturalis are currently in the process of donating large parts of their digitized collections of bird images to Wikimedia Commons, Wikipedia's open multimedia library. Participants of the birdwatching day are challenged to collaboratively identify as

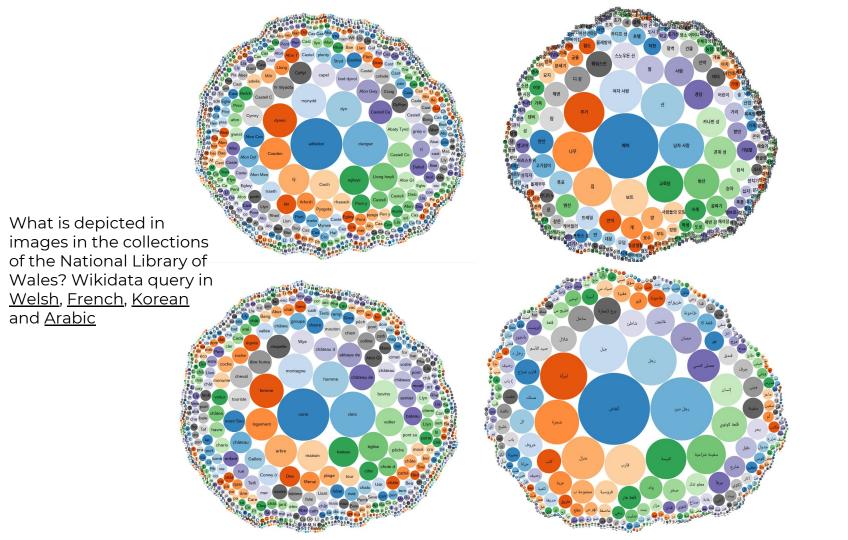
Multilinguality ... also for small languages!

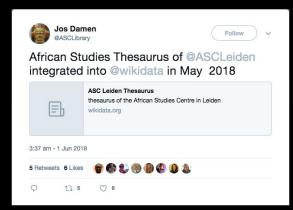




What is depicted in images in the collections of the National Library of Wales? Wikidata query in English...

Jason.nlw, CCO, based on a lithograph (ca. 1840) by W. Crane, Public Domain





Equity in vocabularies

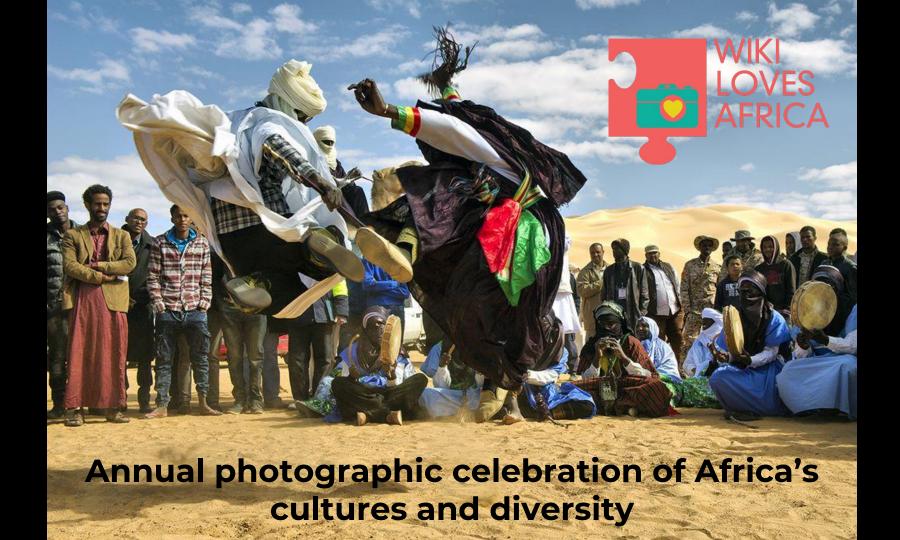
Actively add more diverse terms, people, places... to Wikimedia projects

ethnic group [edit | edit | source]

image +	number +	Article +	item +	description +	ASC Leiden Thesaurus ID	instance of	country +
	1	Abe	Q55096070	ethnic group	294896163	ethnic group	
	2	Aboure	Q2821774	Ivoirian people	294896309	ethnic group	Ghana Ivory Coast
	3	Abron	Q927326	ethnic group	294896252	ethnic group	
	4	Acholi people	Q340581	ethnic group	294896473	ethnic group	
	5	Ada	Q55342471	Ethnic group in Ghana and Togo	294896538	ethnic group	Togo Ghana
	6	Adangbe	Q352275	ethnic group	294896570	ethnic group	
	7	Adangme	Q1159639	ethnic group in Togo and Ghana	294896570	ethnic group	Ghana Togo
	8	Adyukru	Q55317652	Ivoirian people	294896716	ethnic group	Ivory Coast
	9	Afar people	Q211720	ethnic group	294896759	ethnic group	
PA.	10	African Americans	Q49085	racial or ethnic group in the United States with African ancestry	294896813	ethnic group ethnic community	United States of America
	11	Afrikaner people	Q841351	Germanic ethnic group in Southern Africa	294897089	ethnic group	
	12	Afro-Brazilian	Q388315	ethnic group	294897119	ethnic group	

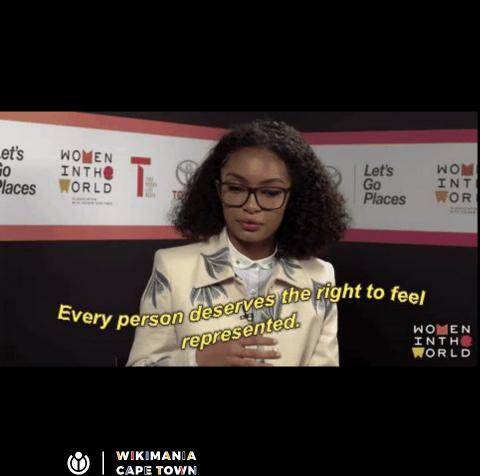
https://www.wikidata.org/wiki/User:Spinster/ASC_Leiden_Thesaurus





65% of people are visual learners

Africa is the least covered region on Wikimedia projects



Wiki Loves Africa

- Annual photographic and media competition.
- Celebrates all that is quintessential about the African continent.
- Each year, submitted images illustrate a specific theme that highlights individual passions and daily encounters, collective idiosyncrasies and universal humanity.
- •
- Run on a continental level with local events organised by participating usergroups and volunteers.



2014: Cuisine

2015: Cultural Fashion and Adornment

2016: Music and Dance

2017: People at Work

2019: Play!

2014: Cuisine 1st prize winner



WIKIMANIA CAPE TOWN

2015: Cultural Fashion & Adornment 1st prize winner

2014: Cuisine

- 873 people contributed
- · 6,116 photographs entered
- · 2,420,791 page views
- 969 images used (15.44%) March '18

2015: Cultural Fashion and Adornment

- 722 people contributed
- · 7,500 photographs entered
- · 10,482,196 page views
- · 609 images used (8.25%) March '18



2014: Music and Dance

- · 836 people contributed
- · 7844 photographs entered
- · 1,206,727 page views
- · 319 images used (4.04%) March '18

2017: People at Work

- · 2,473 people contributed
- · 18,294 photographs entered
- · 2,929,945 page views
- 1,124 images used (6.14%). March '18



2016: Music & Dance 1st prize winner





2017: People at Work

1st prize winner

Challenges

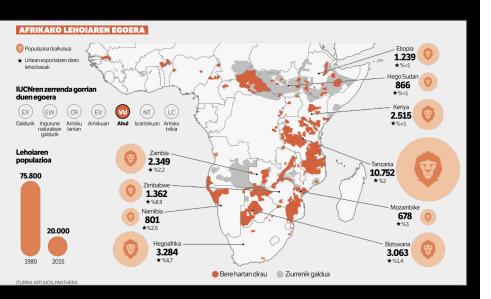
Majority of users are first time:

- **File names** are tricky
- **Descriptions** are difficult
- Categorisation is completely foreign and therefore usage is relatively low unless actively tasked by a member of the team (some kind of AI would definitely help here ;-))
- Images are "dumped" with the country tying images together, but nothing else.
- Cleaning up what suits criteria is manual and normally done by one person.
- **Small teams** very little local team input (confidence)
- **Photo Essays** keeping them together and automatically representing
- Need to **curate** the process more ...

Berria infographics

How we can reuse newspapers' content

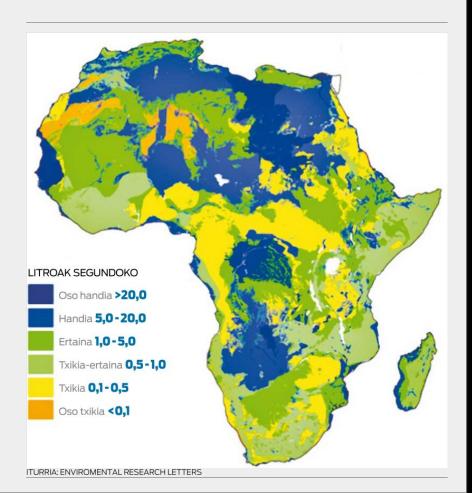






- Berria newspaper now is publishing the web version under cc-by-sa.
- The have published more than 16.000 infographics in their history.
- Now they are uploading them to Commons, with the tags they use in their newspaper (i.e. Africa).

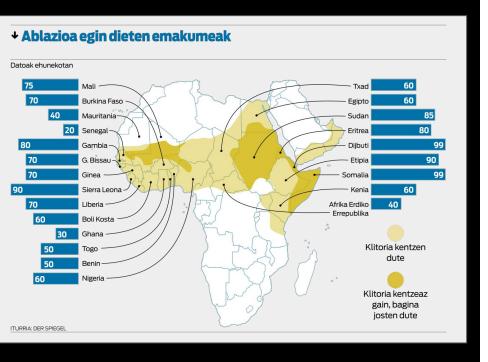
◆ Akuiferoen emankortasuna



- Everything is uploaded in svg format, so i18n is easier.
- Lots of infographics are about very specific subjects.

(i.e. unemployment in the Basque Country in August 2015)

 Some others are highly valuable for everyone.





- Tags and topics can benefit broadly using P180.
- Language translations can be easier once they start to be better organized.
- Infographics have text, let's dream a future with automatic svg translation from Wikidata entities.

Digitized publications on Commons from the Indian Wikimedia communities



Books/Validated Books updated on 26 Feb 2018

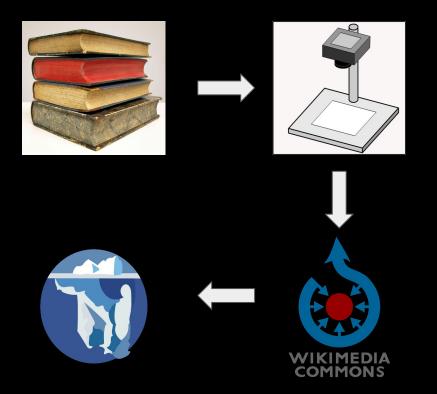
Languge +	Index File/ Book	Readable Book +
Assamese ^[1]	14	00 ^[2]
Bengali ^[3]	2858	37 ^[4]
Gujarati ^[5]	61	00 ^[6]
Kannada ^[7]	162	06 [8]
Malayalam ^[9]	242	00 ^[10]
Marathi ^[11]	36	00 ^[12]
Odia ^[13]	54	02 [14]
Sanskrit ^[15]	166	00 ^[16]
Tamil ^[17]	2169	30 [18]
Telugu ^[19]	349	48 [20]
Punjabi ^[21]	14	00 ^[22]

Statistics of Indic Wikisource

• 11 languages

123 readable books

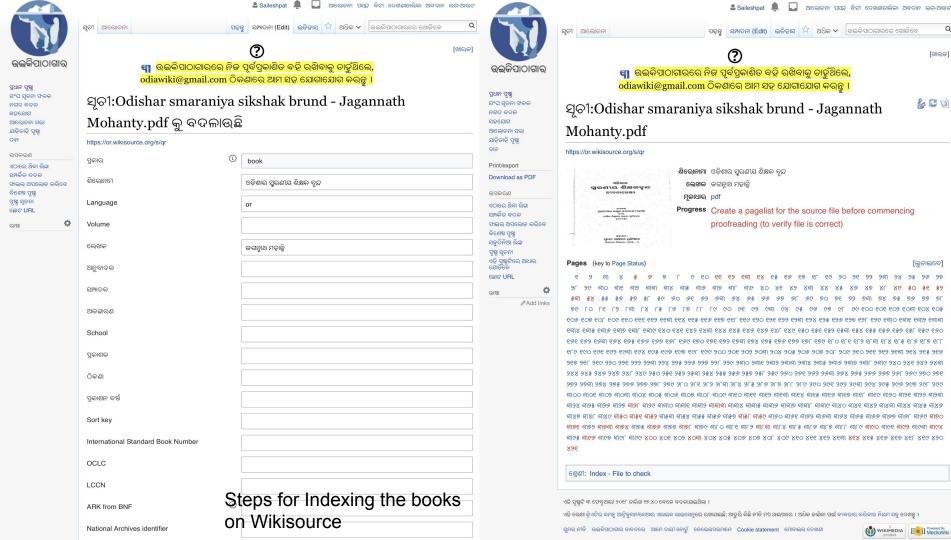
(As of 26 Feb 2018)



Workflow

- Copyright free/PD books
- Digitization
- Wikimedia Commons
- Wikisource
 - OCR
 - Proofreading
 - Validating





[ଖାରଜ]

(S 3)

ଲ୍ଟଚାଇବେ

Structured data and Wikisource



- It will make the Indexing of the books on Wikisource more easier. Will save more time.
- Extract the data related to the work of the author/publisher
- Activate Listeria bot when a new file will be added to Commons it will auto-update the list.



ф	Name +	image +	Sports ¢	country for sport	Stmnts +	SL +	IAAF ◆	Wikidata ◆
1	Alia Saeed Mohammed		5000 metres 1500 metres	United Arab Emirates	19	6	235683	Q16211257
2	Annu Rani		javelin throw	India	16	6	256879	Q17403394
3	Anu Raghavan		400 metres hurdles	India	11	1	261260	Q33104653
4	Aya Ota		shot put	Japan	7	0	292528	Q33111803
5	Ayako Jinnouchi	MAAA	1500 metres	Japan	16	4	204522	Q11658164
6	Ayako Kimura		100 metres hurdles	Japan	29	8	259312	Q4830847
7	Bùi Thị Thu Thảo		long jump	Vietnam	11	4	294015	Q32177640
8	Cecilia Yeung		high jump	Hong Kong	16	4	308354	Q29698185
9	Chen Qiaoling		pole vault	People's Republic of China	9	2	298726	Q33104570
10	Chen Yang		discus throw	People's Republic of China	19	7	256859	Q26404027
11	Darya Maslova		5000 metres 10,000 metres	Kyrgyzstan	17	7	291067	Q20745026
12	Debashree Mazumdar	IND	4 × 400 metres relay	India	16	6	271416	Q26265953
13	Dilhani Lekamage		javelin throw	Sri Lanka	6	0		Q33112318
14	Dutee Chand		100 metres 200 metres 4 × 100 metres relay	India	24	17	275950	Q15020846

- Extract the data as you want to show
- Customise it according to the interest of local readers.
- Digital catalogue.

Images from State Government of Odisha List by: https://en.wikipedia.org/wiki/User:MKar/List/Participants_of_2017_A sian Athletics Championships

