Structured Data for Multimedia

Project Overview

Prepared by
the Multimedia and Wikidata teams,
with members of the Wikimedia Community

Updated 11/06/2014

Issues

Current problems with metadata for multimedia:

- It's hard to **search for images**.

- Many file labels are **only in English**.

- Some metadata is **duplicated or missing**.

- It's **hard to develop tools** with unstructured data.

[Learn more](#).

Mission

We aim to **use structured data** for all **media files**, to:

- make it easier for **users** to **read & edit file info**

- enable **developers** to **build better multimedia tools**

We are starting a **collaborative investigation**, as a community.

This **major infrastructure project** could take several years.

Users

This project impacts **a range of users**:

- **readers**
- **contributors**
- **editors/curators**
- **developers**
- **3rd-party site users**
- **cultural institutions**

Notes:
- initial focus on **Commons**, then **Wikipedia** and other sites.
- initial focus on **advanced users** and **developers**.

Many different uses

Upload

Edit/Curate

Search

View

Re-use

User stories

Here are some user needs that structured data could improve:

- **Search media files** more effectively.
- View information **in my language**.
- **Learn more** with **related information**.
- **Edit file information** more easily.
- **Contribute data** from other pages.
- **Get smart suggestions** during upload

See more user stories.

WIKIMEDIA COMMONS

Search media

★ ⬚ 0 👤 Cronopio ▾

File:Vincent Willem van Gogh 127.jpg

Vincent van Gogh

Dutch painter, draughtsman and printmaker

🖼 W ❞ •••

Also known as **Vincent Willem van Gogh**
Live period **1853 - 1890**
Work locations **Netherlands** and **2 more**

More details from Wikidata

Details	Data

Artist   Vincent van Gogh ✎ •••
Title   Vase with Fifteen Sunflowers
Date   18888
Medium   Oil on canvas

Learn more with related information

When selecting a structured data item on a file page, related content from Wikidata appears in an info card.

(first concept mockup)

Vase with Fifteen Sunflowers by Vincent Van Gogh, from National Gallery, public domain. Vincent van Gogh - Self-Portrait by Vincent Van Gogh, from Google Art Project, public domain. Both files via Wikimedia Commons.

WIKIMEDIA
COMMONS

File:Vincent Willem van Gogh 127.jpg

Details	Data

Artist | Vincent Van Gogh |

By saving changes, you agree to our Terms of Use and agree to irrevocably release your text under the CC BY-SA 3.0 License and GFDL

Cancel   **Change author**

Title   Vase with Fifteen Sunflowers
Date   18888
Medium   Oil on canvas

Edit file information more easily

I can easily edit any structured data item on a file page, by opening a simple editing tool.

(first concept mockup)

Vase with Fifteen Sunflowers by Vincent Van Gogh, from National Gallery, public domain.

Contribute data from other pages

round black spots measuring from 2 to 3 cm (0.79 to 1.18 in) across, affording it some camouflage while hunting. There are no spots on its white underside, but the tail has spots, which merge to form four to six dark rings at the end. The tail usually ends in a bushy white tuft. The cheetah has a small head with high-set eyes. Black "tear marks" running from the corner of its eyes down the sides of the nose to its mouth keep sunlight out of its eyes and aid in hunting and seeing long distances. Its thin and fragile body make it well-suited to short bursts of high speed, but not to long-distance running.

Agility, rather than raw speed, accounts for much of the cheetah's ability to catch prey. Cheetahs can accelerate four times as fast as a human (thanks to greater muscle power) and can slow down by 14 kilometers per hour in one stride. They can hunt successfully in dense forests[16]

The adult cheetah weighs from 21 to 72 kg (46 to 159 lb). Its total head-an measure 60 to 84 cm (24 to 33 in) in length.[17][18][19][20] Cheetahs are 66 larger than females and have slightly bigger heads, but there is not a great apart by appearance alone. Compared to a similarly sized leopard, the che averages about 90 cm (35 in) tall) and so it appears more streamlined.

Some cheetahs have a rare fur pattern mutation of larger, blotchy, merged constitute a separate subspecies but are in fact African cheetahs; their unu "king cheetah" has only been seen in the wild a handful of times, but it has

The cheetah's paws have semi-retr the flat-headed cat and the Iriomote structure of the cheetah's claws is t and fur present in other varieties, a always visible. The dewclaw is muc

Comparative illustration of a leopard (left) and cheetah (right)

Adaptations that enable the cheeta increased oxygen intake, and an enlarged heart and lungs that work together to circulate oxygen efficiently. During a typical chase, its respiratory rate increases from 60 to 150 breaths per minute.[11] While running, in addition to having good traction due to its semi-retractable claws, the cheetah uses its tail as a rudder-like means of steering[citation needed] to allow it to make sharp turns,

with

Image details missing
The image will not be visible for readers until the details are provided

Cheetah running | Delete image

Cronopio

Source | This is my own work

License | Creative Commons Attribution ShareAlike 3.0

Add a description

English | + | Created today | More

Add categories

Cancel | Save page

Get smart suggestions

Product

How can we support these uses?
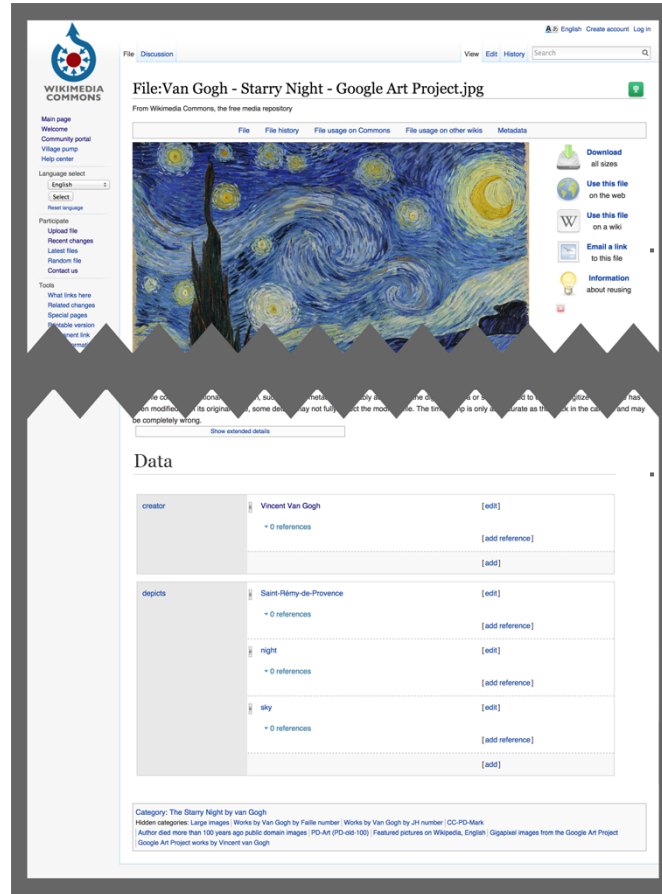
Touchpoints

Where will people find structured data?

Wikipedia

Commons

Wikidata

user

article

file page, with:

data section

item page
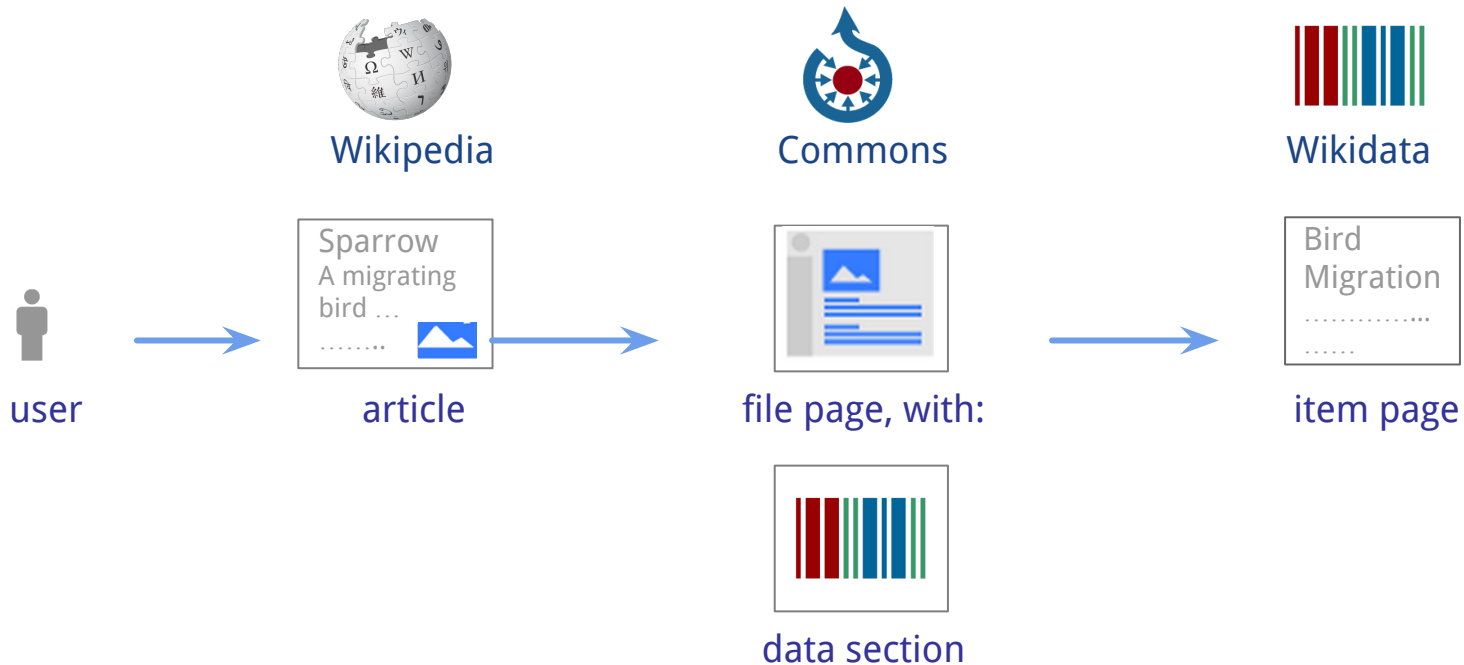
Data on file pages

First concept

File page

Data section

File page vs. Data section

Who are they for? When should you use them?

File page sections

For:
• **all users**

Used to:
• **view** all metadata
• **edit** non-structured data
• **access** data section

Data section

For:
• **advanced users**

Used to:
• **view** structured data
• **edit** structured data
• **migrate** structured data

Unstructured file information would be gradually migrated into structured format in the data section.

Data section

Hamster Berta (Q6573)   [edit]

My hamster Berta in 2007.   [edit]

Statements

contributor	http://commons.wikimedia.org/wiki/User:Hugo85	[edit]
	role — photographer	
	[add]	

| time of creation | March 2007 | [edit] |
| | [add] | |

topic	hamster	[edit]
	photograph	
	[add]	

| license | CC-BY-SA-3.0-DE | [edit] |
| | | [add reference] |

This is a rough example of what a data section *might* look like, with the current Wikidata user interface.

Data view

Title	English ⌄	Lake Laanemaa	文A 2
Description		Lake Laanemaa at Orkjärve Nature Reserve, Estonia. Sunrise. Intense colors were present for just couple of minutes.	文A 2
About		lake	
Type of work		Photograph	More about the same work?
Author		*Heino Ruiso?*	
Rights		Creative Commons Attribution-ShareAlike 3.0 (unported) license own work verified	
Campaign		Eesti Looduse fotovõistlus 2013	
Quality assessment		Featured picture on Commons and Persian Wikipedia	

+ Add more information

Title in other languages

Eesti
Laanemaa järv ✏ 🗑

文A Add in another language

This is a first example of what a data view *might* look like, with a new user interface.

Data editor

Description in [English ▾]

Painting about a big lake.

文A 2   +

Cancel   **Create description**

Description in two languages

English
Painting about a big...   ✎ 🗑

文A |Language

French
German
Spanish
Add

Description in

Description in [English ▾]

Painting about a big lake.

文A 2   +

Cancel   **Create description**

Description in two languages

English
Painting about a big...   ✎ 🗑

Spanish
Pintura sobre un gran...   ✎ 🗑

文A Add another language   🔍

This is a first example of how a structured data editor *might* help edit descriptions in different languages.

Data editor

Rights   + Add rights information

Rights for   the whole work ✕
License   CC-BY-SA   +
Contributor   Cronopio
Justification   own work ✓
                          Cancel   Create rights info

Rights for |Indicate the part of the work
License   The whole work   +
                 Painting
Contributor   Scan
Justification   own work ▾
                          Cancel   Create rights info

Rights for   Painting ✕
License   CC-BY-SA   +
Contributor   Cronopio
Justification   own work ✓
        + Add rigths for another part of the work
                          Cancel   Create rights info
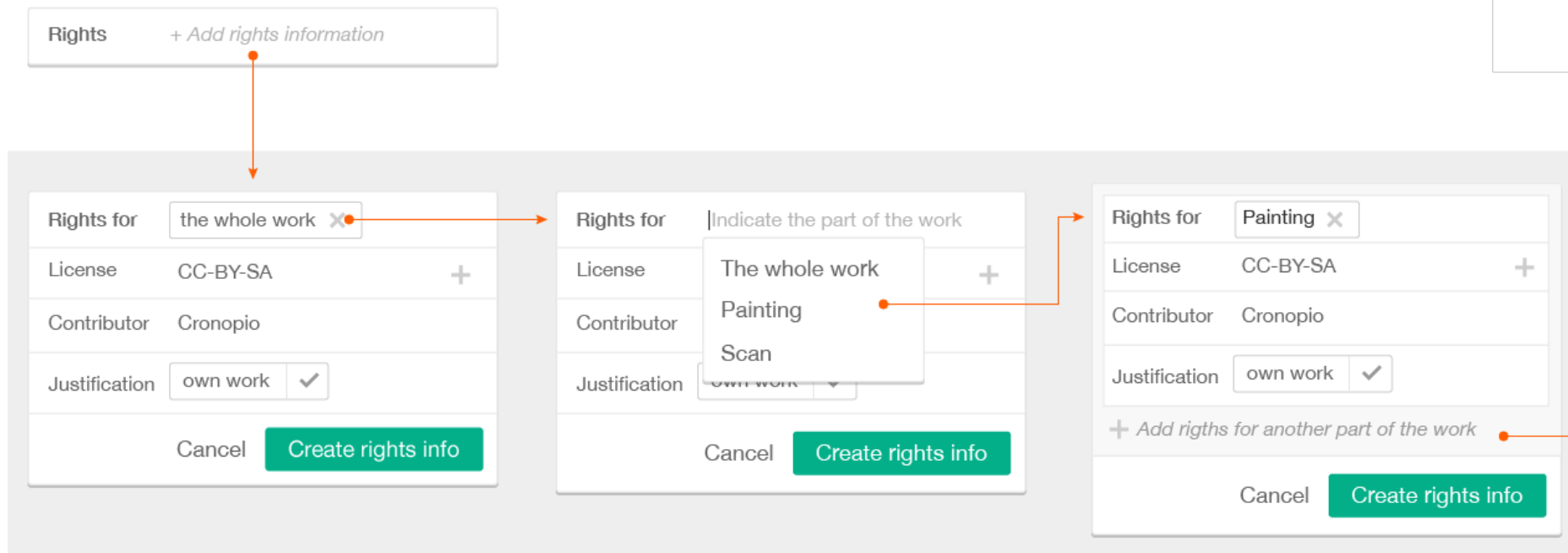
This is a first example of how a structured data editor *might* help edit rights or licenses for different works.

Key data

In a **first phase**, we would focus on these key data for each file:

- **title**
- **description**
- **work**
- **contributor**
- **role**
- **right/license**
- **restriction**
- **category**
- **topics**

See Data Model in progress

Key concepts

Here's one way to visualize some of the structured data clusters, at a high level:

File
- title of file
- description
- metadata

Links to:
- contributor
- rights/license
- work
- etc.

1.n

Work *
- title of work
- role
- etc.

1.n

Contributor
- title of contributor
- description
- etc.

1.n

Rights
- title of right/license
- description
- etc.
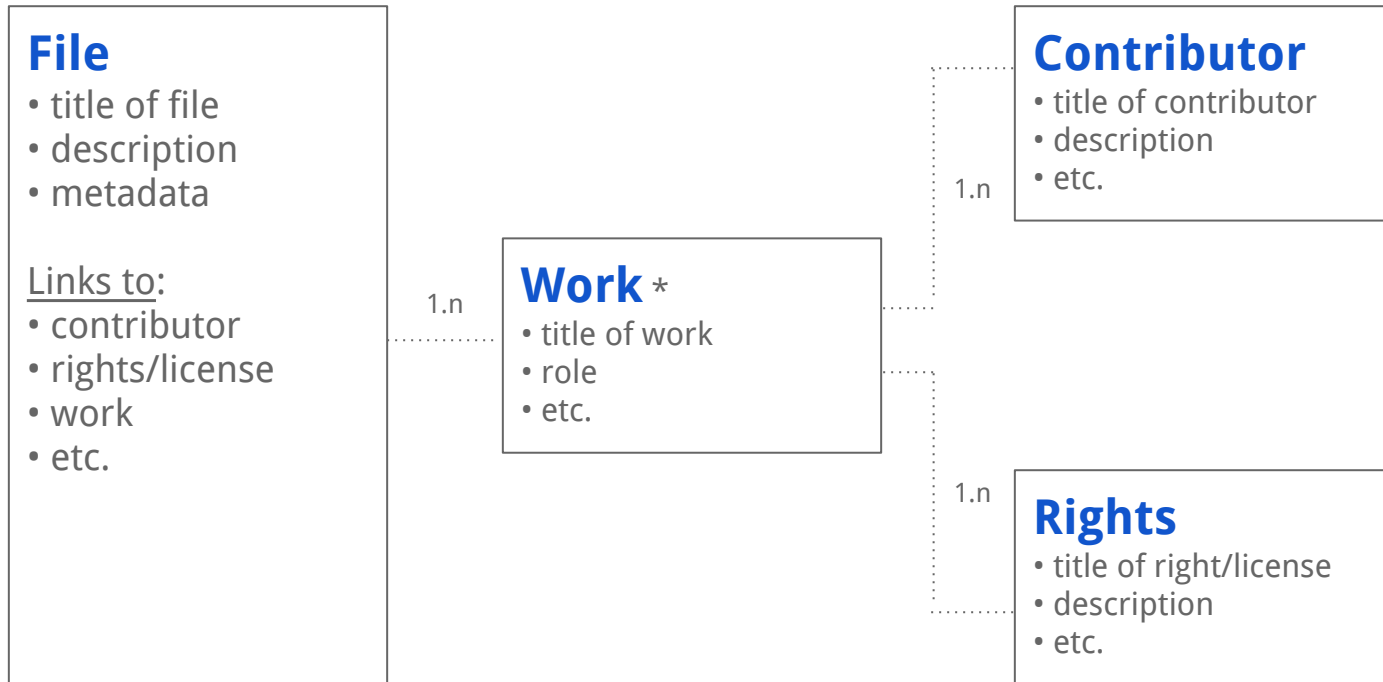
Storage

Where might data be stored?

Wikibase on Commons	Wikidata.org	Commons MediaWiki
title	contributor *	file name
description	rights/license	format
contributor link	work *	resolution
rights/license link	topic *	size
work link	category item *	date
topic link	media properties	geo-tags
file link		category
	* = optional	EXIF metadata

API Class Diagram
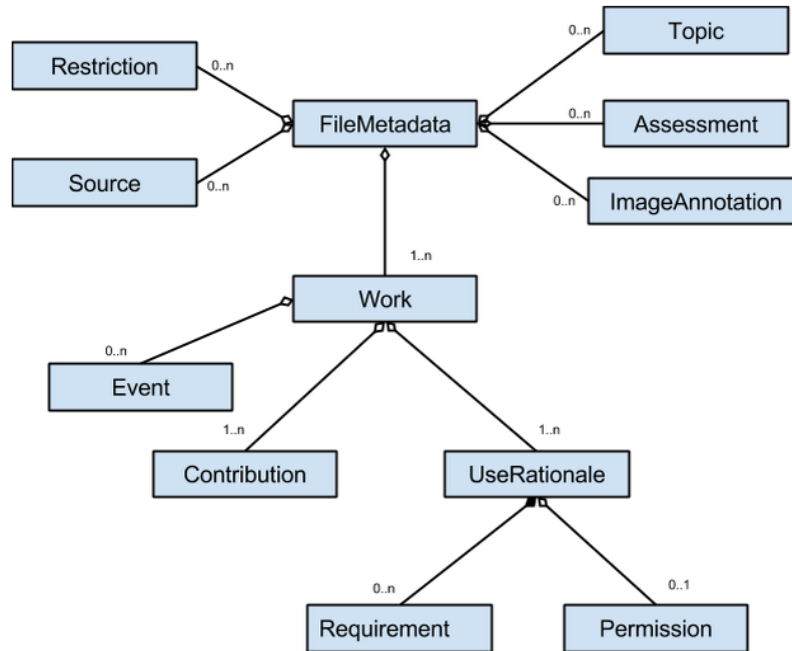
- **Work** - component with separate authorship (e.g. statue on a photograph)
- **Restriction** - usage limitations (trademarks, personality rights etc.)
- **Source** - origin of a work (museum or WLM campaign)
- **Event** - point in time relevant for copyright assessment (publication date, author death)
- **Contribution** - an item from the author list
- **UseRationale/Rights/License** - CC-BY, PD, etc.
- **Requirement** - copyright limitations on the reuser (e.g. share-alike)
- **Permission** - OTRS ID, etc.
- **Topic/Category** - properties describing what's visible in the image
- **Assessment** - featured etc
- **ImageAnnotation** - highlight areas of the image

See Data Model in progress

Categories and topics

Files could be associated with a range of **topics** and **categories** (which can 'bundle' multiple topics).

Next steps

Goals

- **discuss & define** this project, **as a community**

- **allocate resources** to support development

- **design** the data structure & user interface

- **develop & test** prototypes

- **iterate** on key features and infrastructure

- **migrate** unstructured data collectively, over time

Learn more.

Workflows

FIRST WORKFLOWS

• **view** structured data

• **edit** structured data

• **migrate** data in structured form

NEXT WORKFLOWS

• **search** by multiple criteria

• **upload** with structured data

Components

Here are some of the first modules we might work on:

- **high-level API**

- **data back-end** storage

- **data section** on file page

- **edit tool** on file page

- **review** tools + processes

- **basic search**

Questions

- Which **components** to build first?

- What **data model** should be used?

- How to **test** first models / functions?

- How to gradually **roll out** first functions?

- How to coordinate the **data migration**?

- How to **measure** and **validate** each function or data transfer?

Measures of success

Here are some of the metrics we are **considering** to track success:

- **# files used** on WMF-hosted projects

- **# uncategorized files** on Commons

- **# search queries** for media files

- **% successful searches** for files

- **# views per file**

- **# files with data**

See more metrics under consideration.

Measures of success (ctd.)

Here are more metrics we are **discussing** to track success for different use cases:

All	Upload	Edit	Search	View	Re-use	Progress
# files used	# uncategorized files	# editors per file	# file search queries	# views per file	# API use (volume)	# files with data properties
task performance:		# active file editors	# file search volume		# links to files	% new uploads with data
• add/edit info • find files			% successful searches			% template calls without params.

We plan to host **community discussions** to evaluate and select key metrics for this project.

Roadmap 2014-15

	Stage 1	Stage 2	Stage 3
product	• planning • design	• experiment • specs	• learn / metrics • update plans/specs
developers	• code review • code cleanup	• prototypes • small features	• info / file pages • editing tools
community	• discuss • guide	• discuss • test	• migrate data • edit data

We are now in stage 1. We will update this roadmap regularly, as a community.

The future

How structured data can help our movement:

- improve **search** for media files
- show information in **many languages**
- find information on **related topics**
- make it easier to **edit** and **re-use files**
- drive **curation** and **contributions**
- offer a **better multimedia experience**

Learn more.

Collaboration

This project is a **partnership** between:

WIKIMEDIA COMMONS

WIKIMEDIA DEUTSCHLAND

WIKIMEDIA FOUNDATION

This is a promising **collaboration model** for software development, with both challenges and opportunities -- and a chance to **learn together**.

Get involved

Thanks to all our collaborators on this project!

Join us.

Learn more:

commons.wikimedia.org/Commons:Structured_data

A collaboration between:

Wikimedia Community

Wikidata Team (Wikimedia Deutschland)

Multimedia Team (Wikimedia Foundation)