Structured Data for Commons

Project Overview

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

Updated 10/14/2014

Purpose

The Stuctured Data project aims to:

- **use structured data** for all media files on our sites

- make it easier for **users** to **read / write file info**

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

This is a major infrastructure project that is likely to 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**.

Benefits

How structured data can help our movement:

- offer a **better user experience**
- **find relevant content**
- offer metadata in **more languages**
- make it easier to **re-use files** in a license-compliant way
- drive **curation** and **contributions**
- support & create feedback loop for **GLAMs**

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 file information **in my language**
- **Learn more** with **related information**
- **Edit file information** more easily
- **Contribute data** from other pages
- **Get smart suggestions**

WIKIMEDIA COMMONS

Search media

⭐ 📑 0 👤 Cronopio ▾

File:Vincent Willem van Gogh 127.jpg

| Details | Data |

Artist Vincent van Gogh ✏ ⋯

Title Vase with Fifteen Sunflowers

Date 18888

Medium Oil on canvas
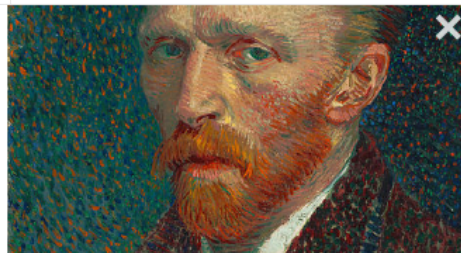
Vincent van Gogh

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

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.

Search media

⭐ 📑 0 👤 Cronopio

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-and 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

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

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

Adaptations that enable the cheetah 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,
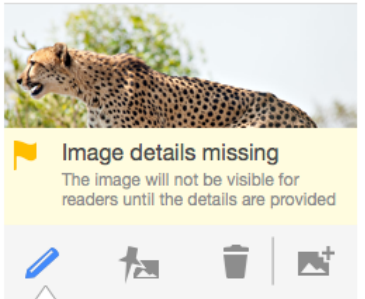
Get smart suggestions

Product

How can we support these uses?

Touchpoints

Where will people find structured data?

Wikipedia

Commons

Wikidata

user

Sparrow
A migrating
bird …
……..

article

file page, with:

Bird
Migration
…………...
……

item page
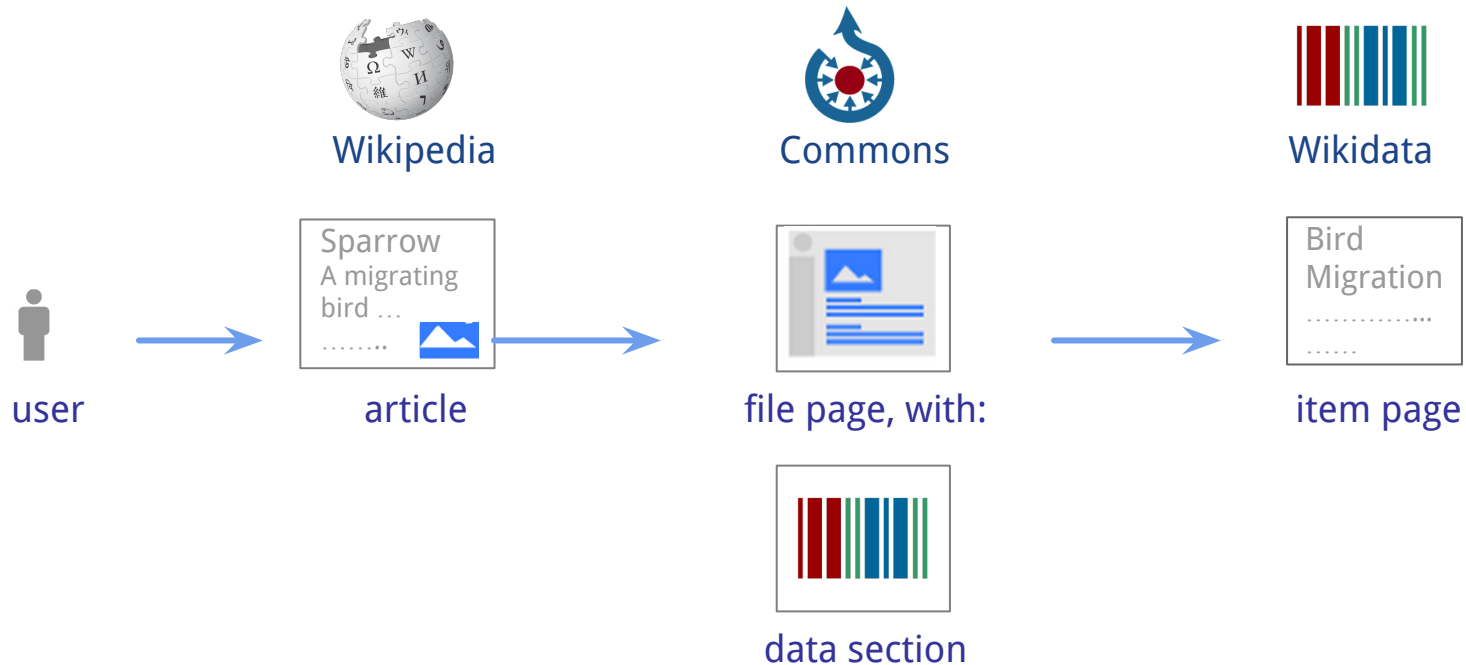
data section

New Data section in File page
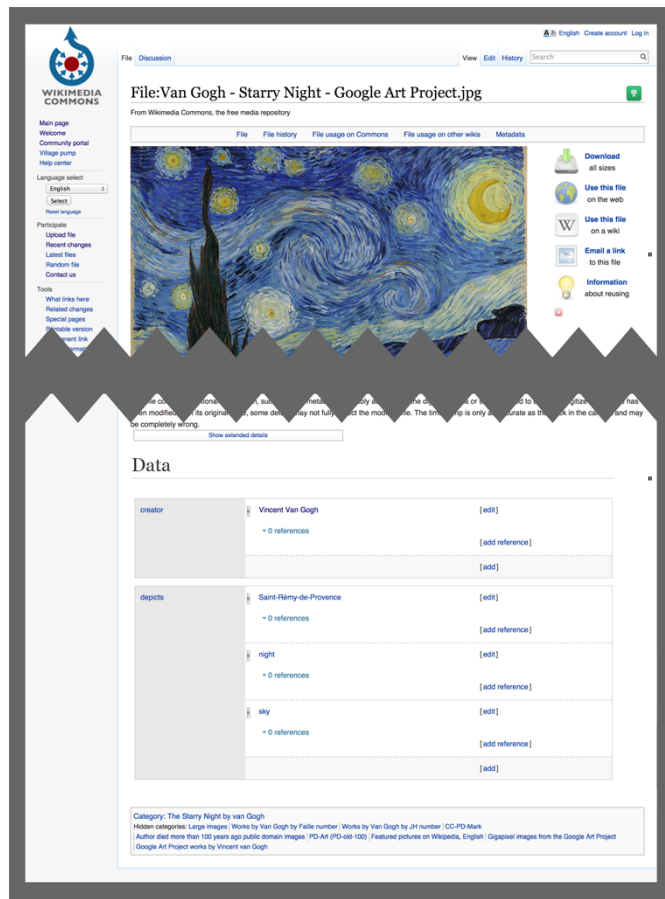
(first concept)

File page

Data section

File page vs. Data section

Who are they for? When should you use them?

File page top sections

For:
- **all users**

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

Data section

For:
- **advanced users**

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

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

Data example

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 structured data might look with the current Wikidata user interface.

Data editor

Creator	Vincent van Gogh	+Add another
Title	*Is it The starry night?*	
About	Saint-Rémy-de-Provence night	
	sky	
Located in	Museum of Modern Art	
+ Add more information		

Creator

Vincent van Gogh

🖉 🗑

🔖 No references

This is a first example of how a structured data editor might look with a new user interface.

More criteria

Churc|

🏷 **Church**
Religious building for Christian worship

🏷 **Churchill**
Prime Minister of the United Kingdom during World War II

📁 **Churches in Berlin**
Churches located in Berlin

Files could be associated with a range of topics and categories (bundling multiple topics).

Key data

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

- **title**
- **description**
- **work**
- **contributor**
- **role**
- **license**
- **restrictions**
- **category**
- **topics**

Key concepts

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

File
- title
- description
- metadata

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

1..n

Work *
- name
- role
- etc.

1..n

Contributor
- title
- description
- etc.

1..n

License
- title
- description
- etc.
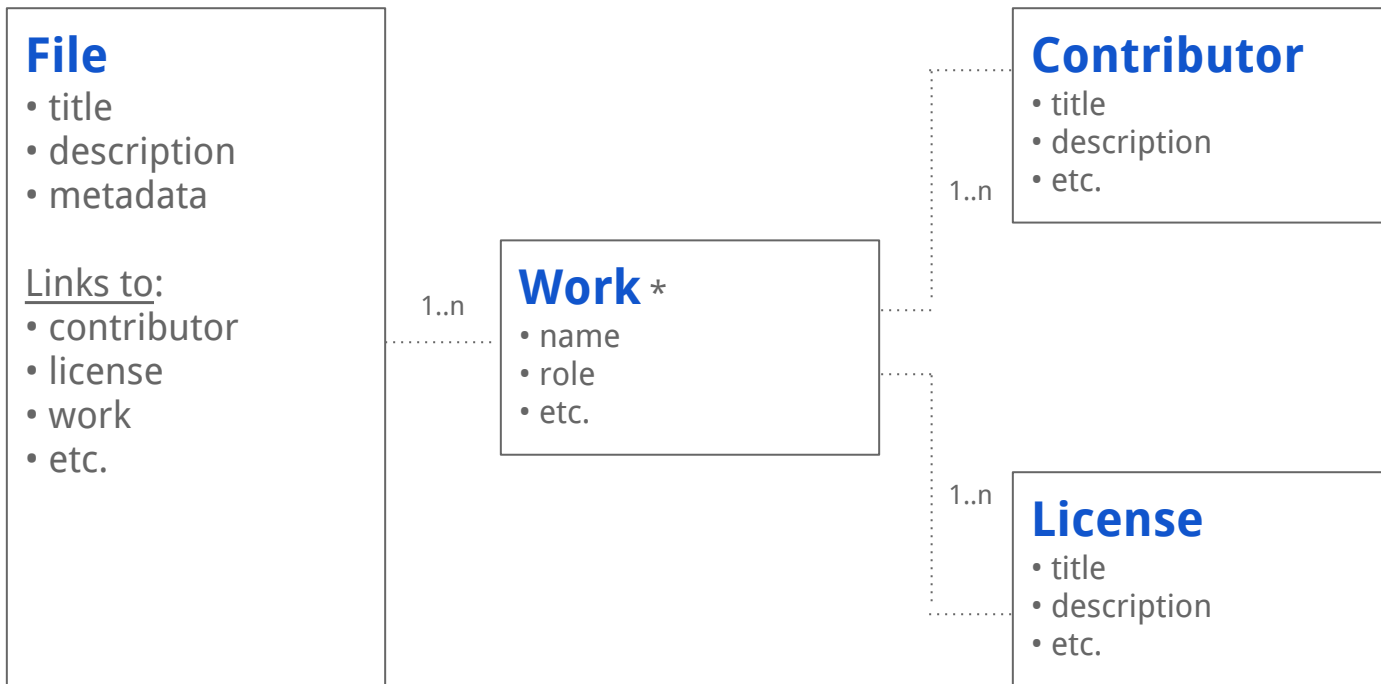
Storage

Where might data be stored?

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

API Class Diagram
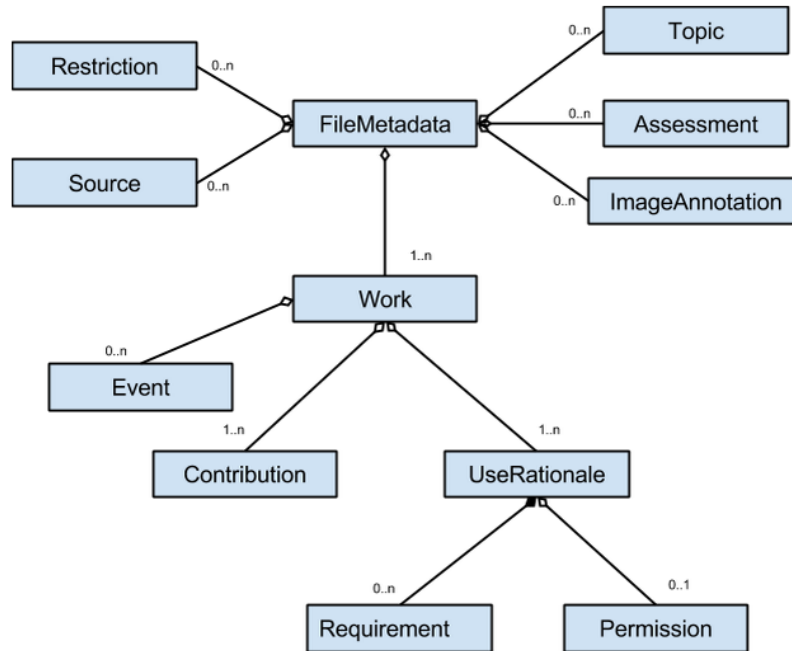
- **Work** - component with separate authorship (e.g. statue on a photograph)
- **Restriction** - usage limitations (trademarks, personality rights etc.)
- **Source** - origin of a work (museum or WLM campaign)
- **Event** - point in time relevant for copyright assessment (publication date, author death)
- **Contribution** - an item from the author list
- **UseRationale/License** - CC-BY, PD claim, 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

Next steps

Goals

- **Plan & discuss** this project with our communities

- **Design & test** the data structure and user interface

- **Develop** the code and functions for this project

- **Migrate** some unstructured data to the new format

- **Test, observe** and adjust plan as needed

Components

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

- **Data Model** - map key data
- **High-level API** - read/write data
- **File Page Update** - access to data section
- **Data Section** - view, edit structured data

Workflows

FIRST WORKFLOWS

- **migrate** data in structured form

- **view** structured data

- **edit** structured data

NEXT WORKFLOWS

- **search** by multiple criteria

- **upload** with structured data

Functions

- **high-level API**

- **data back-end** storage

- **data section** on file page

- **edit tool** on file page

- **review** tools + processes

- **basic search**

Questions

- Which **functions** 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 feature/data?

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

Stage 2 is now starting: these goals will be discussed with the community.

What now?

Here's what will happen in coming months:

- **more discussions**
- **planning, design, specs**
- **first prototypes & tests**
- **developer bootcamps**
- **learn, adjust, try again**

- **be patient: this will take time ...**

Learn more:

commons.wikimedia.org/Commons:Structured_data

A collaboration between:
- **Wikimedia Community**
- **Wikidata Team** (Wikimedia Deutschland)
- **Multimedia Team** (Wikimedia Foundation)