

The Path to Structured Metadata for Multimedia

Presentation for
Wikimedia Foundation's Metrics Meeting

Prepared by
the [Multimedia](#) and [Wikidata](#) teams,
with members of the [Wikimedia Community](#)

Presented on November 6, 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.](#)

Goals

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

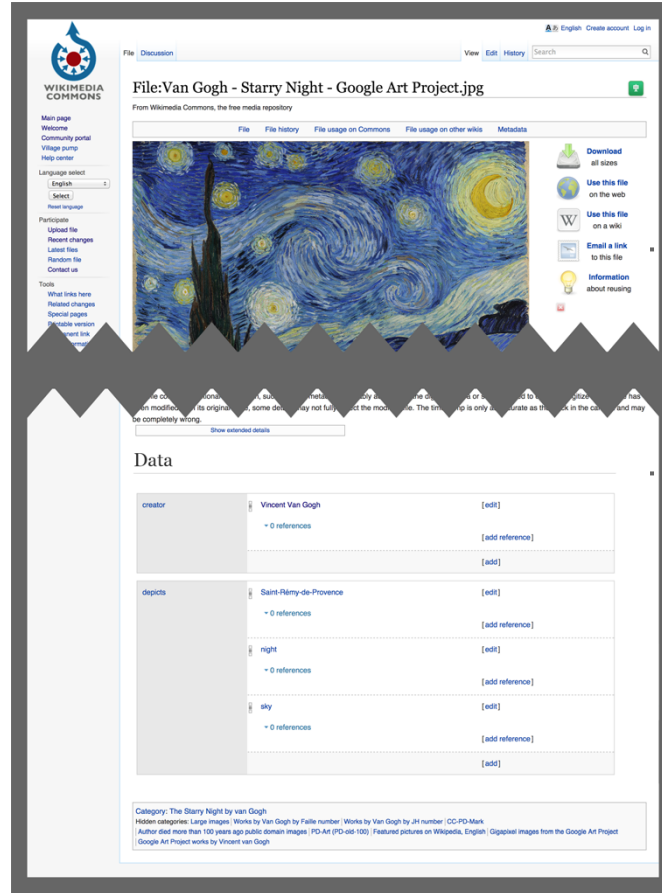
- make it easier for **users** to **read & edit file info**
- enable **developers** to **build better multimedia tools**

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

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

Data on file pages

First concept



The screenshot shows a Wikimedia Commons file page for 'File:Van Gogh - Starry Night - Google Art Project.jpg'. The page features the Wikimedia Commons logo, navigation links, and a central image of the painting. Below the image is a 'Data' section with a table of metadata.

Property	Value	Actions
creator	Vincent Van Gogh	[edit]
	• 0 references	[add reference]
		[add]
depicts	Saint-Rémy-de-Provence	[edit]
	• 0 references	[add reference]
	night	[edit]
	• 0 references	[add reference]
	sky	[edit]
	• 0 references	[add reference]
		[add]

Category: The Starry Night by van Gogh
Hidden categories: Large images | Works by Van Gogh by Falla number | Works by Van Gogh by JH number | CC-PD-Mark
Author died more than 100 years ago public domain images | PD-Art (PD-old-100) | Featured pictures on Wikipedia, English | Guggenheim images from the Google Art Project | Google Art Project works by Vincent van Gogh

File page






Data view

Starry Night by [Vincent Van Gogh](#), from Google Art Project. Public domain. Via Wikimedia Commons.

Data view

First concept

The image shows a user interface for viewing metadata of a photograph. The main content is a table with the following rows:

	Title	English ▼	Lake Laanemaa	 2
	Description	Lake Laanemaa at Orkjärve Nature Reserve, Estonia. Sunrise. Intense colors were present for just couple of minutes.		 2
	About	lake		
	Type of work	Photograph	More about the same work?	
	Author	<i>Heino Ruiso?</i>		
	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			

On the right side, there is a panel titled "Title in other languages" showing the title in Estonian: "Laanemaa järv". It includes edit and delete icons and a button to "Add in another language".

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

Data editor

First concept

The image shows two versions of a data editor interface, illustrating a transition from a simple single-language editor to a structured multilingual editor.

Top Left (Single Language Editor): A form titled "Description in" with a dropdown menu set to "English". The text area contains "Painting about a big lake.". At the bottom right, there is a small icon of a star and the number "2" next to a plus sign. At the bottom, there are two buttons: "Cancel" and "Create description".

Top Right (Structured Multilingual Editor - Draft): A form titled "Description in two languages". It shows the "English" version: "Painting about a big..." with edit and delete icons. Below this is a section for adding more languages, labeled "Language" with a star icon. A dropdown menu is open, showing "French", "German", and "Spanish". A green "Add" button is visible to the right. An orange arrow points from the "Spanish" option in the dropdown to the bottom right version.

Bottom Left (Single Language Editor): Identical to the top left version.

Bottom Right (Structured Multilingual Editor - Final): The form now shows two entries: "English" with "Painting about a big..." and "Spanish" with "Pintura sobre un gran...". At the bottom, there is a section "Add another language" with a star icon and a search icon.

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

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.](#)

Next steps

- **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.](#)

File Metadata Cleanup Drive

a.k.a. "OMG Think Of The Robots!!!"

Machine-readable structured data

Description	A red bird
Author	Puzzly
Source	Own work
Rights	CC-by-sa-3.0

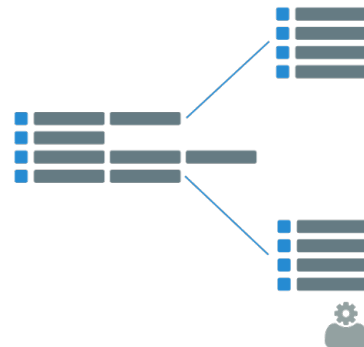


what you edit

Description	A red bird
Author	Puzzly
Source	Own work
This file is under the Creative Commons Attribution Share-Alike 3.0 license.	



what you see



what robots see

Machine-readable templated wikitext (2006–present)

```
{{information
| description = A red bird
| author = Puzzly
| source = Own work
}}
{{CC-by-sa-3.0}}
```



Description A red bird

Author Puzzly

Source Own work

This file is under the Creative Commons Attribution Share-Alike 3.0 license.



Mostly templated wikitext

```
{{information
| description = A red bird
| author = Puzzly
| source = Own work
}}
{{CC-by-sa-3.0}}
```



Description A red bird

Author Puzzly

Source Own work

This file is under the Creative Commons Attribution Share-Alike 3.0 license.



Plain wikitext (2001–present)

I took this picture of a red
bird and it's under cc by sa.



I took this picture of a red bird and
it's under cc by sa.



How many files are we talking about?

Meet Mr Metadata

tools.wmflabs.org/mrmetadata

File metadata cleanup drive

The [File metadata cleanup drive](#) is an effort started in September 2014 by the Wikimedia Foundation. Its goal is to fix file description pages and tweak templates to ensure that multimedia files consistently contain machine-readable metadata across Wikimedia wikis.

[See how to fix the pages >>](#)

Global tally

42% complete

1,466,688 files to go

- Total number of files: 2,495,716
- Files with missing machine-readable metadata: 1,466,688

Incubator

95% complete

1 file to

Mediawiki

18% complete

2,496 files to go

Meta

23% complete

3,293 files to go

Species

92% complete

1 file to go

Strategy

45% complete

143 files to go

wikivoyage:en — File metadata cleanup drive

The [File metadata cleanup drive](#) is an effort started in September 2014 by the Wikimedia Foundation. Its goal is to fix file description pages and tweak templates to ensure that multimedia files consistently contain machine-readable metadata across Wikimedia wikis.

The files listed below don't contain machine-readable metadata. Please help fix them!

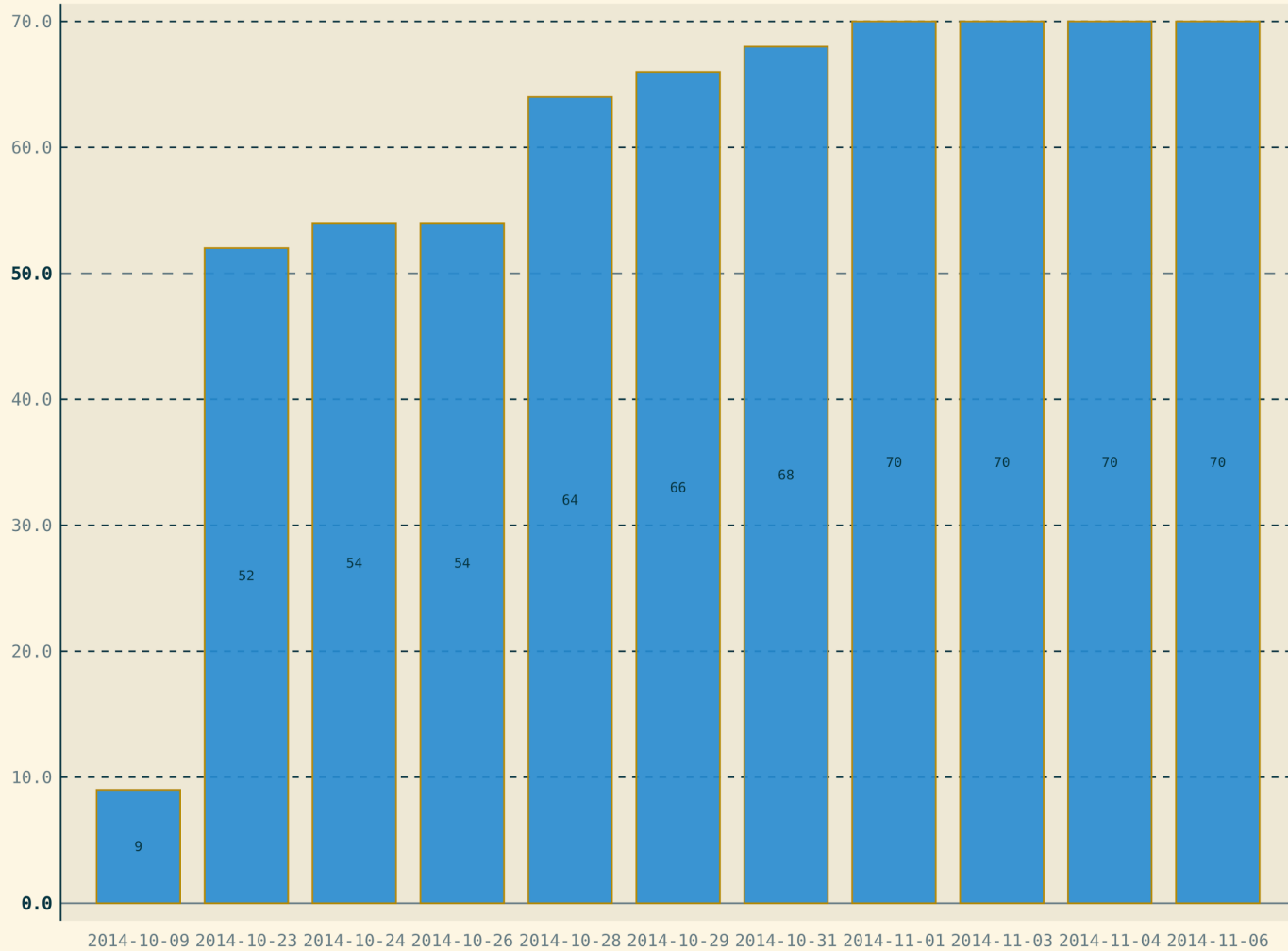
70% complete

224 files to go

- Total number of files: 733
- Files with missing machine-readable metadata: 224
- Last updated on: 2014-11-06

[See how to fix the metadata »](#)

#	File	Description	Author	Source	License	License URL
1	File:1024 Main Page.png	missing	missing	missing	ok	ok
2	File:11.png	missing	missing	missing	ok	ok
3	File:1366 Main Page design.png	missing	missing	missing	ok	ok
4	File:Abu Dhabi in evening.JPG	missing	missing	missing	ok	ok
5	File:Accra lighthouse.jpg	ok	missing	missing	missing	missing
6	File:Akihabara Ad Manga.JPG	missing	missing	missing	ok	ok
7	File:AmundsenWEB.jpg	missing	missing	missing	ok	missing
8	File:Anniversary banner.png	missing	missing	missing	ok	ok



文件元数据清理/如何修复元数据

File metadata cleanup drive

This page is a translated version of the page File metadata cleanup drive/How to fix metadata and the translation is 77% complete.

Other languages:	العربية • català • Chavacano de Zamboanga • Deutsch • English • español • فارسی • suomi • français • हिन्दी • magyar • Bahasa Indonesia • italiano • Lëtzebuergesch • Bahasa Melayu • norsk bokmål • Nederlands • پشتو • português • português do Brasil • русский • Basa Sunda • svenska • தமிழ் • Türkçe • українська • 中文
-------------------------	---

- 文件元数据清理
- 进展
- 如何修复文件

Many files on Wikimedia wikis don't contain **machine-readable metadata**. You can help fix them!

為沒有信息模板的文件提供信息模板

對於每個頁面，我們需要找出所有沒有信息模板的文件，並為它們補充信息模板。
使用這個工具來查看哪些文件缺少達到標準的信息模板。這個工具將會試圖推測如何添加合適的信息模板，並自動生成一個模板供您編輯（示例）。
 這個工具默認列出維基共享資源中的文件，不過您也可以使用它來列出其他項目的文件，比方說**中文維基文庫**。

加入机器可读标记至信息模板

对于维基共享资源以外的wiki，我们需要检查**跨wiki链接列表**以查看您的wiki是否已拥有信息模板。

您的wiki尚未拥有信息模板 [\[edit source\]](#)

- 直接复制一个简化版本至您的wiki；此模板已拥有机器可读标记。
- 為了使您項目的其他人可以找到它，請添加它到維基內部連結列表。
- 将模板加入您wiki的文件中；实验室工具也可以帮助您。

您的wiki已经拥有信息模板 [\[edit source\]](#)

如果您的wiki的模板是最近几年中从共享资源导入的，它可能已经有了标记。否则，您可以手动加入它们：

- 請在包含實際信息的 `<td>` 之前的 `<td>` 元素中以 HTML ID 的形式添加這些標記。這意味著您所標記的 `<td>` 元素將包含此字段的標題，而不是此字段本身。比如：

```
<td id="fileinfotpl_desc">说明: </td>
<td>{{{描述}}}</td>
```

- 在描述中将 `id="fileinfotpl_desc"` 加入至 `<td>` 元素中。
- 在来源中将 `id="fileinfotpl_src"` 加入至 `<td>` 元素中。
- 在作者中将 `id="fileinfotpl_auth"` 加入至 `<td>` 元素中。

Contents [\[hide\]](#)

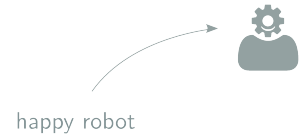
- 為沒有信息模板的文件提供信息模板
- 加入机器可读标记至信息模板
 - 您的wiki尚未拥有信息模板
 - 您的wiki已经拥有信息模板
- 加入机器可读标记至版权模板
 - 包含多许可的模板
 - Assessments that are not licenses (Public domain, fair use)
- 了解您的进度
- 获取帮助
- 阅读更多

了解您的進度

Once you've added the markers to the templates, you need to wait a little for the markers to propagate to the files. After a few days, you should see a change on the **progress dashboard** for your wiki.

获取帮助

- 元维基：在讨论页留下消息，一些人会帮助您！您可以使用您的语言，如果您不使用英语。
- 通过email：订阅技术大使列表然后提问。
- 通过IRC：加入freenode上的#wikimediaconnect频道然后提问；当没人可询问时你可以ping guillom。



happy robot

Make the robots happy

And when the time comes, they'll do all the migration work for you.

Join us, and start fixing files and templates on your wiki:

meta.wikimedia.org/wiki/File_metadata_cleanup_drive

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)