



# Structured Data on Commons: A Second-Year Report For the Sloan Foundation

Prepared by Amanda Bittaker ([abittaker@wikimedia.org](mailto:abittaker@wikimedia.org)), Caitlin Virtue ([cvirtue@wikimedia.org](mailto:cvirtue@wikimedia.org)), and Jonathan Curiel, with contributions by Ramsey Isler, Sandra Fauconnier, Pamela Drouin, Brad Jorsch, Daniel Kinzler, Jonathan Morgan, Ben Vershbow, and Anne Gomez

November 1, 2018

Grant#: G-2016 7239

# Table of Contents

<b>Executive Summary</b>	<b>1</b>
<b>What We Did:</b>	<b>3</b>
Multi-Content Revisions	3
Multilingual File Captions	7
Architecting Search	15
Research	22
Other Design Work	26
Community and GLAM Stakeholder Input	28
<b>How We're Doing:</b>	<b>32</b>
Our Progress So Far	32
The Year in Review: A Timeline of Key Developments and Outreach	33
How We're Measuring Success: Deliverables Oct. 2017 to Sept. 2018	37
How We're Measuring Success: Deliverables Oct. 2018 to Sept. 2019	42
How We're Measuring Success: Targets and Measurement Methods	44
<b>What's Up Next:</b>	<b>49</b>
Rest of Grant	49
Beyond 2020	50
<b>Appendix:</b>	<b>52</b>
Staff working on Structured Data on Commons	52
Website links: Program and project pages	53
Glossary of Structured Data on Commons terms	54

# Executive Summary

Three weeks ago, British researcher Jamie Fawcett [posed a plea](#) on Twitter – asking for genuine ways to “encourage better participation in collaborative creation of openly licensed content.” Within minutes, Wikimedia UK – the Wikimedia Foundation’s British chapter – [tweeted back](#) a shout-out to our Structured Data on Commons project, telling Fawcett that when the project is done, “better tools will appear.”

Those tools are closer than ever. The Wikimedia Foundation has made good progress on the project, and we’re happy to report on a series of major developments from the past year, including:

- [Multi-Content Revisions](#) (MCR) is deployed and ready for structured data on Commons. This is the foundational engineering work that allows Structured Data on Commons to move forward. (See pages 3-6.)
- Our first project feature, [multilingual file captions](#), is currently being deployed to Beta Commons, and will be on production Commons later this month. Multilingual file captions are the first step to letting users in different languages utilize Commons as a truly global platform. (See pages 7-14.)
- We completed several major research surveys of Commons users, including those from GLAM institutions, and [the published results](#) gave us a new understanding of the needs and working habits of Commons users – and gave us fundamental insights into how to implement the project. (See pages 22-26.)
- We created an [ongoing focus group comprised of GLAM representatives](#) that is giving us regular input on the project. And we created a [community focus group](#) that is also providing valuable input. (See pages 28-31.)
- We’ve published several iterations of designs for the project, working with the Commons community to visualize the practical ways that Commons users can add captions, upload media files to Commons, and do other tasks. (See pages 7-14 and pages 15-22.)
- We started building a new search interface that gives Commons users a detailed way to search for Commons files. The interface integrates Commons very closely with Wikidata. (See pages 15-22.)

Each of these developments contributed to the project's overall momentum. And each one is detailed below – starting with our work on Multi-Content Revisions.

We're where we want to be. We've accomplished major milestones (see pages 33-43) and we plan to achieve the [outcomes](#) we set in our 2016 grant proposal.

Thank you for your support, which has taken Structured Data on Commons a long way in a relatively short time.

[The Structured Data on Commons team](#) – a 20-plus team of software engineers, product managers, designers, researchers, and others – has now worked together for more than a full year. [Last year's report](#) introduced you to a team that was newly formed. The project was also in its early stages – not quite a year into its funding from the Alfred P. Sloan Foundation. Your three-year grant has enabled us to move ahead with a project that will enable the metadata of every free media file on Wikimedia Commons – 50 million and counting – to be converted to a structured and machine-readable format, so that they become easier to view, search, edit, organize and re-use.

In the past year, we've dived into the features that we expect to be functional within the next 14 months. And we can already see beyond 2020 – to iterations that will even more closely involve GLAM institutions, and more closely involve Commons users from around the world – which we describe in this report's "What's Up Next" section on pages 50-52. But the majority of this report details the arc of our past 12 months. We went to where Commons users were – meeting them online and also in person, including at Wikimedia conferences and GLAM conference rooms. They welcomed our presentations, and we welcomed their feedback – part of a cycle of community feedback that has included nine community input requests and four online office hours. The past year included a robust discovery period, where we're making sure we meet all of our users' key needs. This cycle of feedback and response has been crucial for the project, and we detail it throughout this report.

As we've made advances on the project, more people are finding out about Structured Data on Commons and its potential to change the way people use and share media files, not just on Commons but across the Internet. "Exciting" is what Jamie Fawcett – a Master's student with the Oxford Internet Institute who's worked with the Open Data Institute – called the project after learning of its existence. In the following pages, we outline the progress of our past year.

# What We Did:

## Multi-Content Revisions

Multi-Content Revisions (MCR) is the most complicated engineering work of the project, and one of the most significant updates to the MediaWiki software in a decade. It's the engineering cornerstone of the project's first phase – and the only part that substantially changes the way MediaWiki works. One way to think of it: Multi-Content Revisions is a software infrastructure project within the Structured Data on Commons project. MCR changes the way data is stored and managed. Another way to think of it: Multi-Content Revisions is to Structured Data on Commons what a network of steel is to a groundbreaking high-rise building.

Multi-Content Revisions works on the backend of MediaWiki, the core software that all Wikimedia wikis use. Because it's on the backend, Commons users won't see the changes we've made. But Commons users will greatly benefit from these changes, which fundamentally improve the way that MediaWiki will store information about Commons' millions of media files, and will free up that information – including metadata – for better integration across Wikimedia sites and across the Internet.

As its name implies, Multi-Content Revisions is a layer of software that supports the multiple types of content in each revision. This allows multiple types of content to be managed on a single page. That's the key feature – without it, we could not have the wikitext descriptions and the new structured data descriptions side by side, fully integrated into the page and existing community workflows. Before Multi-Content Revisions, a page using MediaWiki software could only store *one* type of data: wikitext, or [JSON](#), or structured [Wikibase](#) data. Now, all the different components – including text, code additions, and structured data – are on a single page.

To manage these multiple content revisions under Structured Data on Commons, we needed to restructure part of the MediaWiki storage layer by adding a new level of “indirection” or linking between revisions and the different components. The introduction of this additional layer of indirection required a new storage interface in the server-side code, but more importantly it required a [data-base schema change](#) and [data migration](#).

In the past year, we began working on more than 20 major, high risk software engineering changes, where each change is an important part of the bigger MediaWiki restructuring. Among the granular engineering work that we've done:

- We [implemented](#) an MCR storage layer. This involved refactoring and implementing work needed in [MediaWiki core](#), which allows page revisions with multiple content objects to be stored and retrieved.
- We [made diff views](#) work with multiple [slots](#). A diff view details a revision of a page, and this engineering work lets us show diffs from multiple slots of a revision.
- We [implemented](#) basic multi-slot views. This work parallels our work on diff views, and shows the content of all slots when viewing a revision of a page.
- We [configured an extra slot](#) that lets us store Wikibase entities on MediaWiki in addition to wiki-text. [Wikibase](#), which is the software that runs Wikidata, was developed as a tool for collaborative semantic knowledge modeling – and it's foundational to Structured Data on Commons.
- We [ensured](#) that MediaWiki extensions were compatible with MCR. This ensures, for example, that the tools we use to prevent spam and other malicious behavior work with MCR.
- We made sure our internal search engine, [Cirrus Search](#), can crawl each slot, which will surface the information there to the widest possible audience. Doing this enables semantic search by structured data fields in our own search interface – a big part of the SDC project.

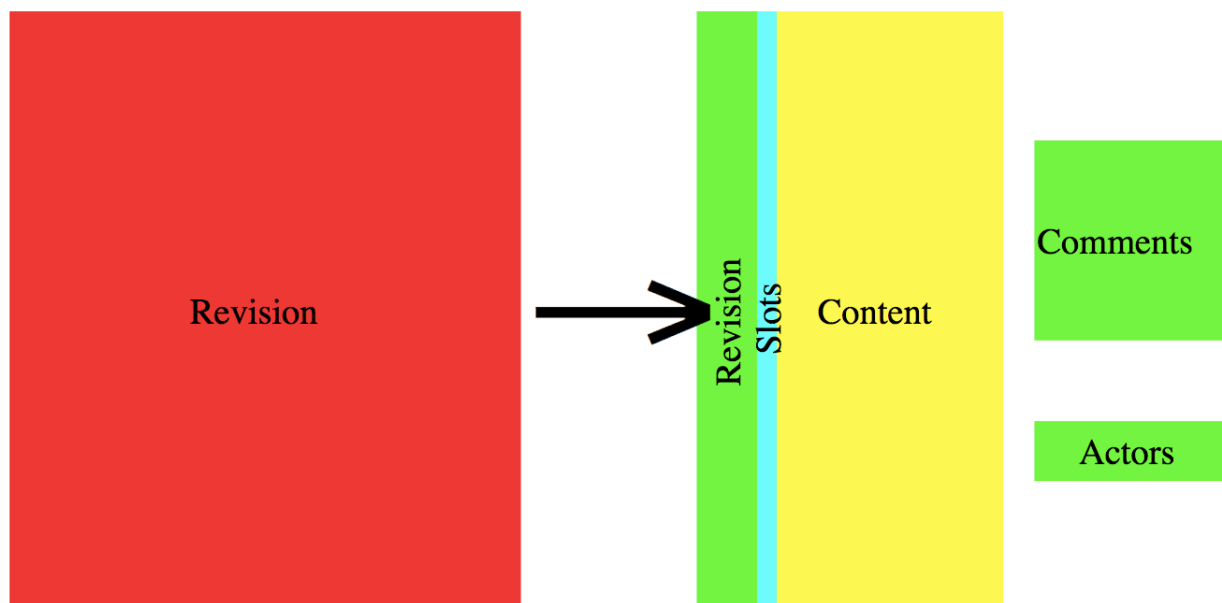
While we were doing the foundational engineering work this year on MCR, we encountered additional technical complexity and software issues – but we'd planned for such contingencies, and these issues became opportunities, segueing into other engineering work that has improved and strengthened the Structured Data on Commons project. One example: Orchestrating a new [Revision Table](#) schema, to improve scalability.

Our engineers went through months and months of iterations – figuring how to streamline the table, and how to separate less important data into auxiliary tables. We worked with the developer community on the Revision Table, and we combined both public feedback (see [here](#), for example)

and staff feedback (see [here](#)) as we progressively iterated new schema until we got it just right. An example of one improvement: We decided to move the edit summary into a separate table, as well as the user name. Moving these two major text strings out of the Revision Table will make it much easier to update the table – and much easier for users to get quick results. The engineering challenge wasn't the schema change as such, but the fact that we're doing the schema change on a table with literally billions of rows, without interrupting operations.

Below is a visual rendering on our work on the Revision Table. On the left: A representative figure of the old Revision table as it had grown in size. On the right of the arrow: A restructuring that streamlines the table – with new separate tables for [slots](#) and content, and new separate tables for Comments and Actors. “Actors” are user names or IP addresses that spur an action or event. (The introduction of slots was not primarily for streamlining the revision table. Rather, it adds a degree of freedom to MediaWiki's storage layer that allows structured data and wikitext to be stored on the same page, which is a key requirement for Structured Data on Commons.)

**Below, left: Revision table before our work. Below right: New configuration after our work.**



Moving the edit summary out of the main Revision Table also allowed us to resolve another long-standing request from users: Giving users more room to describe changes they make in an edit. The previous limit was 255 character bytes. This didn't provide enough space – especially for users who write in Chinese, Cyrillic, and certain other non-Latin alphabets, where some characters count as two or more character bytes. Even some accented characters in European languages use extra bytes. So we [extended the limit to 500 characters](#). The Wikimedia Foundation was planning to make this change eventually, but we took the opportunity to do that now – as part of Structured Data on Commons.

The work to finish all these engineering changes delayed our progress a bit – but we think it was valuable to give Commons users a more extensive package of improvements. Now, they'll be able to better describe their edits – which will help users better understand the changes that appear on Commons. Multi-Content Revisions will not only improve users' experience but make the infrastructure more robust. Each engineering feat in the past year laid the groundwork for this new, improved ability.

The storage component of Multi-Content Revisions is done. We will continue to work on seamless integration of the new multi-slot paradigm into all of MediaWiki for use cases beyond Structured Data on Commons. Further improvements of Multi-Content Revisions will have minimal impact on Structured Data on Commons, since we prioritized the part of MCR that does impact the project.

## **More About Multi-Content Revisions**

### **Phabricator pages that publicly log our engineering work:**

Overview of Phabricator tasks: <https://phabricator.wikimedia.org/tag/multi-content-revisions/>

Phabricator page, Multi-Content Revisions: <https://phabricator.wikimedia.org/T107595>

Phabricator page, "Implement MCR": <https://phabricator.wikimedia.org/T174022>

Phabricator page, "Deploy MCR": <https://phabricator.wikimedia.org/T174043>

Phabricator search page for open tasks with MCR as a tag: <https://phabricator.wikimedia.org/search/query/NG.AahqdUUpm/>

Phabricator search page for all tasks with MCR as a tag: <https://phabricator.wikimedia.org/search/query/q5-SuFH.jgTcC/>

### **MCR request for comment page on MediaWiki:**

[https://www.mediawiki.org/wiki/Requests\\_for\\_comment/Multi-Content\\_Revisions](https://www.mediawiki.org/wiki/Requests_for_comment/Multi-Content_Revisions)

### **Glossary of Multi-Content Revisions:**

[https://www.mediawiki.org/wiki/Multi-Content\\_Revisions/Glossary](https://www.mediawiki.org/wiki/Multi-Content_Revisions/Glossary)



## Multilingual File Captions

Multilingual file captions, which is deployed to Beta Commons, is our first Structured Data on Commons feature – and will be the first change that users see to Commons. It’s a front-end change – a practical, hands-on addition that Commons users can use to describe a Commons file. Like Wikipedia, Wikimedia Commons has a usership that speaks hundreds of languages – not just English – and multilingual file captions is a window into making Commons a truly international knowledge ecosystem that accommodates the languages of people around the world.

Captions are short descriptions of media that users enter. The captions are stored on Commons as structured data using an extension for Wikibase we named [MedialInfo](#). Structured captions let users work with an easy-to-use multilingual field for concisely describing files, help search algorithms prioritize text for search rankings, and provide a simple, consistently formatted string for machine reuse. Captions will live on file pages, where users can read captions added through the upload process, as well as add and edit captions to those file pages.

During the past year, we [introduced](#) three design alternatives for multilingual captions to the Commons community, which users tested and provided [feedback on](#) from May 22 to June 7. The design alternatives revolved around a Commons photo of a Seattle sunset, as seen on the following page. At the bottom of the Seattle sunset file, under “Media info,” is the space for “multilingual captions,” with a drop-down menu.



- Main page
- Welcome
- Community portal
- Village pump
- Help center
- Participate
- Upload file
- Recent changes
- Latest files
- Random file
- Contact us
- Print/export
- Download as PDF
- Tools
- What links here
- Related changes
- Special pages
- Permanent link
- Page information
- Cite this page
- Nominate for deletion

English PDrouin (WMF) Talk Preferences Beta Watchlist Uploads Contributions Log out

File Discussion

View Edit History More

Search Wikimedia Commons

## File:Downtown Seattle sunset.jpg

From Wikimedia Commons, the free media repository

Download Use this file Use this file Email a link Information



Size of this preview: 800 × 600 pixels. Other resolutions: 320 × 240 pixels | 640 × 480 pixels | 1,024 × 768 pixels | 1,280 × 960 pixels | 4,032 × 3,024 pixels. This file has annotations. Move the mouse pointer over the image to see them.

Add a note

Original file (4,032 × 3,024 pixels, file size: 2.44 MB, MIME type: image/jpeg); (request rotation); ZoomViewer: flash/no flash

Open in Media Viewer

### Media info

#### Multilingual captions

English Seattle skyline view from Smith Tower with Mount Rainier in background at sunset Seattle skyline view from Smith Tower with Mount Rainier in background at sunset Seattle skyline view from Smith Tower with Mount Rainier in background at sunset Seattle sky

See more languages

### Summary [edit]

**Description** English: Sunset view of Mt. Rainier from Smith Tower

**Date** 23 October 2017, 18:04:29

**Source** Own work

[mjunipero](#)

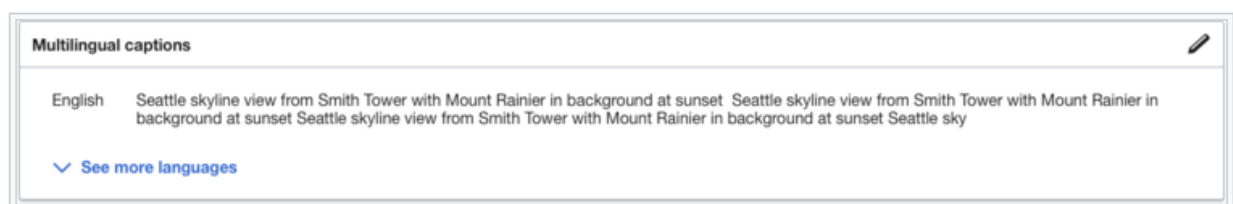
Displaying captions this way supports two design goals: making the preferred language visible to users, and making it clear to contributors how to add the short descriptive caption. Captions appear close to the image to give an idea of what the image is about, and are a piece of metadata that will appear in search results on Commons, in the MediaViewer, or other potential uses.

We tested three different states of displaying captions:

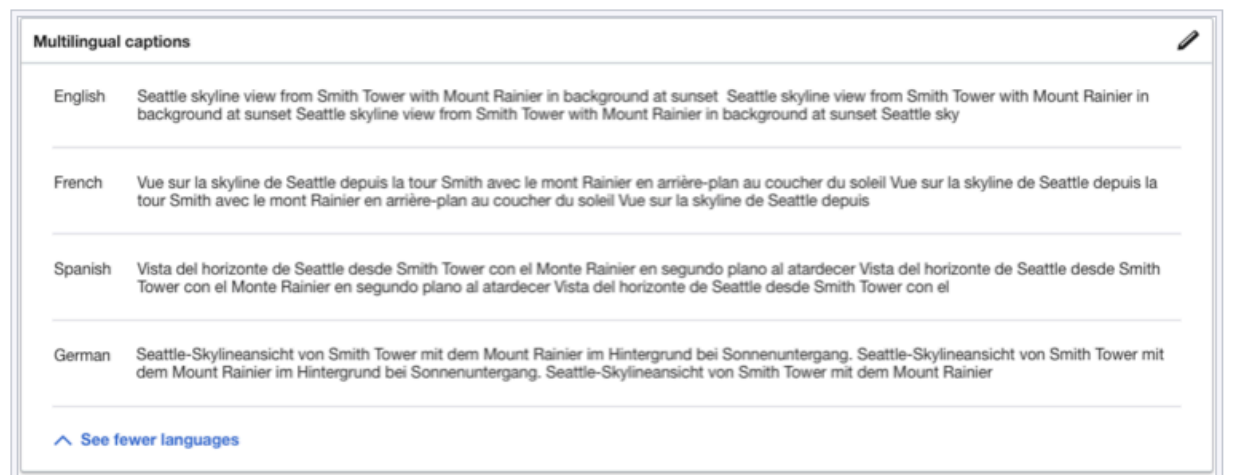
1. In this first image below, a file has no caption, and includes a clear call-to-action for someone to add a caption to it:



2. In the next image below, captions already exist, and the system defaults to the user's preferred language while collapsing the rest:

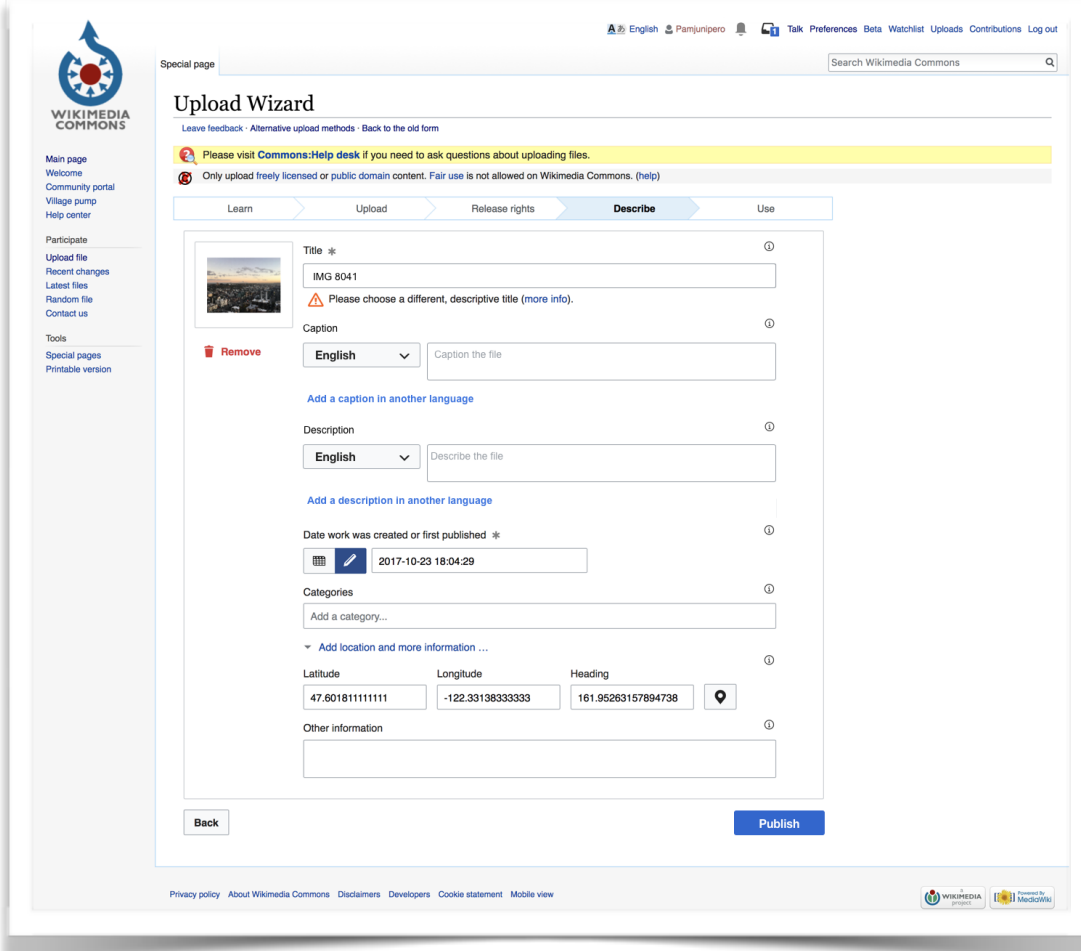


3. The third image below shows all the captions available in an expanded view:



We also [tested](#) the prototype to add multilingual captions during the media upload process, which resulted in [much feedback](#) from January 23 to February 7. The [Commons Upload Wizard](#) is designed to easily help Commons users upload media files. We tested three versions:

**Version 1:** Only a Caption field added; no further changes to the Upload Wizard. With version 1, the current design of the Upload Wizard can be maintained, with multilingual captions handled in the same way that multilingual descriptions are currently handled. There would be minimal changes to the user interface and workflow, and it would be the easiest to implement.



Special page English Panjunipero Talk Preferences Beta Watchlist Uploads Contributions Log out

Search Wikimedia Commons


## Upload Wizard

[Leave feedback](#) · [Alternative upload methods](#) · [Back to the old form](#)

**Please visit Commons:Help desk if you need to ask questions about uploading files.**

Only upload freely licensed or public domain content. Fair use is not allowed on Wikimedia Commons. (help)

Learn > Upload > Release rights > **Describe** > Use



Title \*

Please choose a different, descriptive title (more info).

Remove

Caption

English

[Add a caption in another language](#)

Description

English

[Add a description in another language](#)

Date work was created or first published \*



Categories

▼ Add location and more information ...

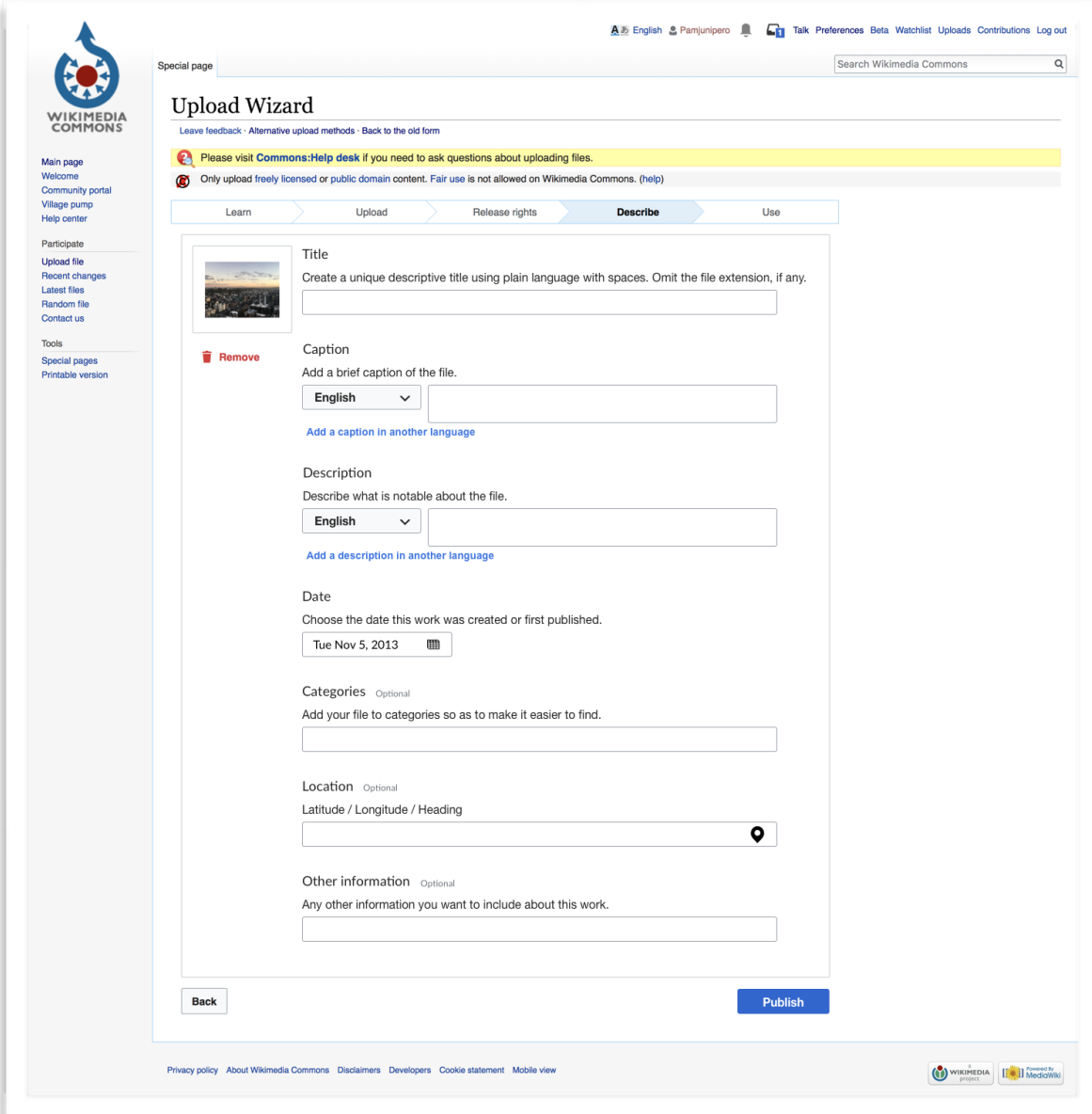
Latitude	Longitude	Heading	
<input type="text" value="47.601811111111"/>	<input type="text" value="-122.33138333333"/>	<input type="text" value="161.95263157894738"/>	<input type="button" value="📍"/>

Other information

[Privacy policy](#) · [About Wikimedia Commons](#) · [Disclaimers](#) · [Developers](#) · [Cookie statement](#) · [Mobile view](#)

**Version 2:** Caption field added; help text visible. Version 2 is a “middle of the road” approach. The current pattern for adding metadata in a multilingual field is maintained, but the help text is made more visible to users, especially new or infrequent users.

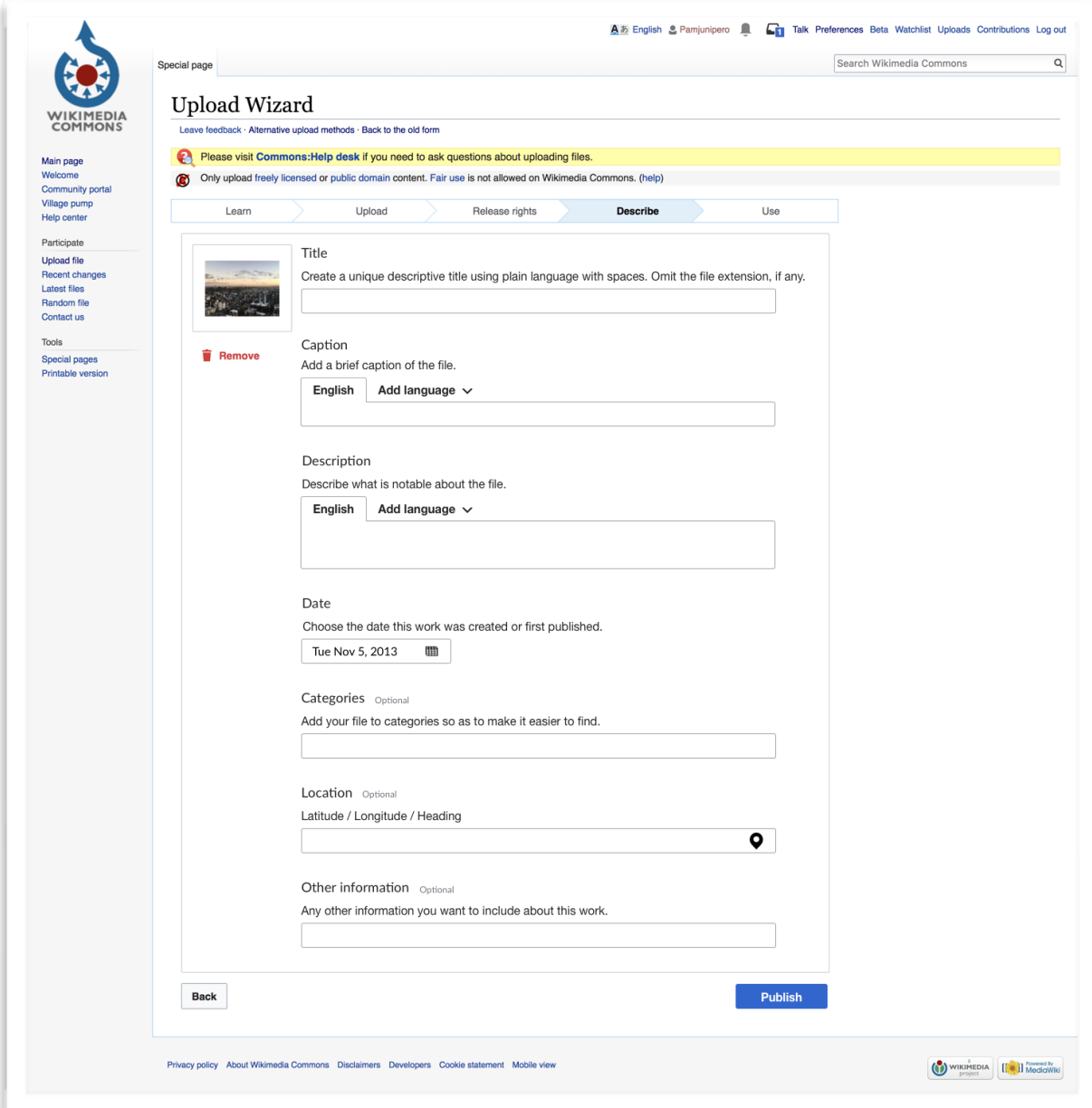


The screenshot shows the 'Describe' step of the Wikimedia Commons Upload Wizard. The page has a light blue header with the Wikimedia Commons logo on the left and navigation links on the right. The main content area is titled 'Upload Wizard' and includes a progress bar with steps: Learn, Upload, Release rights, Describe (current), and Use. The 'Describe' section contains several form fields:

- Title:** A text input field with the instruction: 'Create a unique descriptive title using plain language with spaces. Omit the file extension, if any.'
- Caption:** A text input field with a language dropdown set to 'English' and the instruction: 'Add a brief caption of the file.' Below the field is a link: 'Add a caption in another language'.
- Description:** A text input field with a language dropdown set to 'English' and the instruction: 'Describe what is notable about the file.' Below the field is a link: 'Add a description in another language'.
- Date:** A date picker set to 'Tue Nov 5, 2013' with the instruction: 'Choose the date this work was created or first published.'
- Categories:** A text input field with the instruction: 'Add your file to categories so as to make it easier to find.'
- Location:** A text input field with a location pin icon and the instruction: 'Latitude / Longitude / Heading'.
- Other information:** A text input field with the instruction: 'Any other information you want to include about this work.'

At the bottom of the form are 'Back' and 'Publish' buttons. The footer contains links for Privacy policy, About Wikimedia Commons, Disclaimers, Developers, Cookie statement, and Mobile view, along with logos for Wikimedia projects and the Internet Archive.

**Version 3:** Caption field added; help text visible; tabs for extra languages. Version 3 uses a different approach for adding multilingual fields during the upload process.



Special page English Pamjunipero Talk Preferences Beta Watchlist Uploads Contributions Log out

Search Wikimedia Commons


## Upload Wizard

[Leave feedback](#) · [Alternative upload methods](#) · [Back to the old form](#)

Please visit [Commons:Help desk](#) if you need to ask questions about uploading files.

Only upload [freely licensed](#) or [public domain](#) content. [Fair use](#) is not allowed on Wikimedia Commons. [\(help\)](#)

Learn > Upload > Release rights > **Describe** > Use



**Title**

Create a unique descriptive title using plain language with spaces. Omit the file extension, if any.

Remove

**Caption**

Add a brief caption of the file.

English Add language ▾

**Description**

Describe what is notable about the file.

English Add language ▾

**Date**

Choose the date this work was created or first published.

Tue Nov 5, 2013

**Categories** Optional

Add your file to categories so as to make it easier to find.

**Location** Optional



Latitude / Longitude / Heading

**Other information** Optional

Any other information you want to include about this work.

Back
Publish

[Privacy policy](#) · [About Wikimedia Commons](#) · [Disclaimers](#) · [Developers](#) · [Cookie statement](#) · [Mobile view](#)

The community preferred the second design option, as we learned a lot from [their feedback](#) in January and February and the dozens of comments and back-and-forth discussion. Version 2 offered the best mix of continuity of design and ease of use. Key points that we learned:

- “Help text” should be brought to the surface, where it can explain the differences between “title,” “caption,” and “description.”
- All entered captions should be visible on the screen at once.
- Coordinates located in one field are confusing and difficult to use, so we’ll keep coordinates unchanged from their current form.

In the bigger picture, the feedback on multilingual file captions prompted us to reconsider how we were naming things in the project – including the choice of the word “captions.” Subsequently, we did a [research study](#) on this (detailed in this report on page 23). The feedback on multilingual file caption exemplifies the outreach we’ve done – and are still doing – on the project. We’re listening to Commons users, who are helping shape how Structured Data on Commons is designed and implemented. We believe that this approach will lead to the best and most used version of Structured Data on Commons.

In the past year, users [tested](#) adding captions and other information on our [Structured Data on Commons test site](#) – a site that we noted in last year’s report. Users have since made hundreds of tests. They’re still making additions. Three weeks ago, [a Commons user](#) used the Upload Wizard to [add](#) a test image of a cat that was drawn by the 19th-century Dutch painter Jean Bernard. A caption was added in English: “Sketch of a rolled up cat by artist Jean Bernard.” A few days after the test upload, [another user](#) added a file caption in Dutch that reads this way: “Tekening van een opgerolde kat door kunstenaar Jean Bernard.” As of this writing, the image also has a caption in German: “Skizze einer eingerollten Katze des Künstlers Jean Bernard.”

You can see these three captions on the next page, along with a file history that shows a thumbnail version of Jean Bernard’s cat image:

## File history

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

	Date/Time	Thumbnail	Dimensions	User	Comment
current	<a href="#">21:41, 3 October 2018</a>		3,072 × 1,867 (1.54 MB)	TheRealNightcrafter (talk   contribs)	User created page with UploadWizard

- You cannot overwrite this file.

## File usage

There are no pages that use this file.

## Structured data

### Captions

English	Sketch of a rolled up cat by artist Jean Bernard
German	Skizze einer eingerollten Katze des Künstlers Jean Bernard
Dutch	Tekening van een opgerolde kat door kunstenaar Jean Bernard

[^ See fewer languages](#)

In testing, then, Structured Data on Commons is leading to captions in languages other than English. We expect to see this engagement carry over to Commons when the feature is launched.

## More About Multilingual File Captions

### Designs that we experimented with:

<https://wikimedia.invisionapp.com/share/VZGT8KX93DG#/screens/290767782>

### Community feedback:

[https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Feedback\\_requests/Multilingual\\_-\\_Captions\\_and\\_MediaInfo](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Multilingual_-_Captions_and_MediaInfo)

### Phabricator ticket that publicly tracks the design and technical changes:

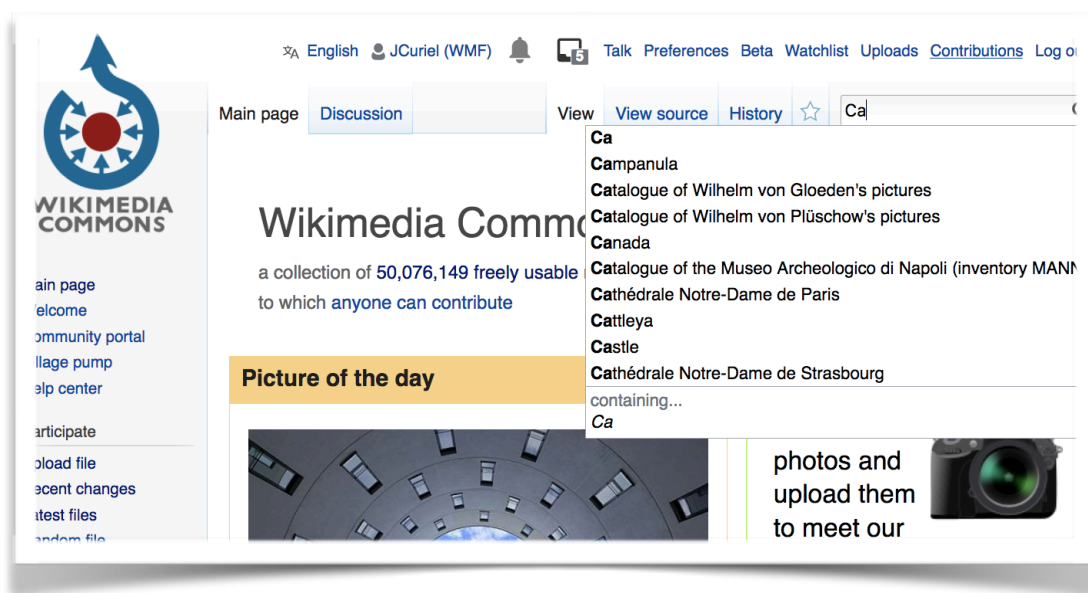
<https://phabricator.wikimedia.org/T201456>



## Architecting Search

How Commons users search for media files – how they look for photos, films, audio and other files – is a central element of Structured Data on Commons, and one that we’ve already advanced through design research, design ideas, and user feedback.

In the current state of Commons, users go to the search box in the upper right side and begin typing in characters – as in the search example below, where a user types in “Ca” and gets suggestions of named files with the same two letters.



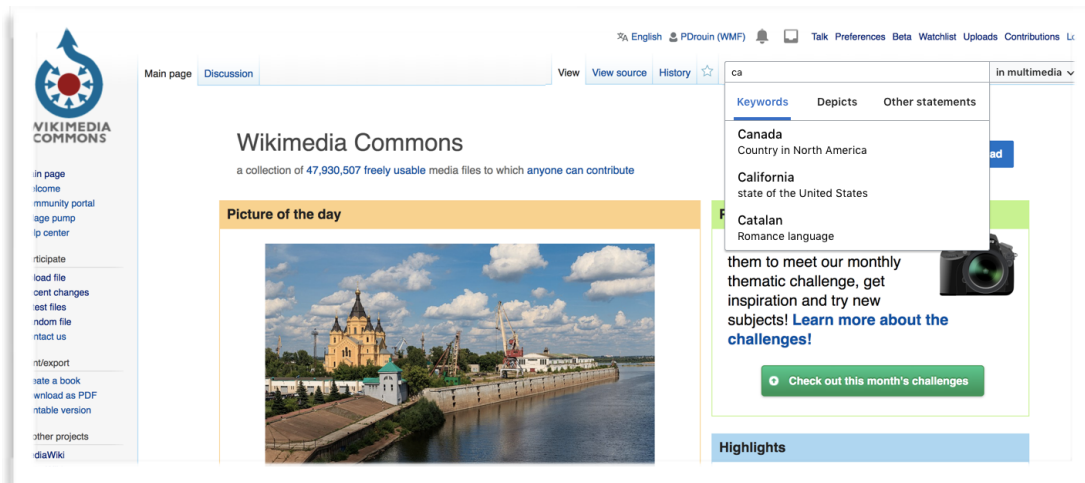
The search method approximates the searching that’s done on Wikipedia, which gives readers a list of suggested articles that gets more specific as users type in more letters – as in the search example on the following page using the letters “Ca” on English Wikipedia. This model isn’t right for media files, where the unique name of the file may not be exactly how someone searching would think about it.



Searching on Commons with structured data will be different. Keyword searching will work exactly as it does now, but we’re going to give users more efficient ways to search for media files, as in the design mockups below, which are also available [here](#). (We are working on an updated UI for adding and editing statements, but it is still going through internal feedback cycles.) This design prototype for search incorporates the new terms that Commons users will be able to use. The vocabulary for describing media files (such as creators, institutions, depicted people, places, animals, plants, buildings, and historical events) is drawn from Wikidata, and includes the three terms and concepts below:

- “[Statements](#).” A statement is a structured field for a key-value pair relating to a single concept, such as creator, media type, location, etc. By linking to each other, statements result in a networked data structure.
- “[Depicts](#).” This is the central and most important statement for a media file, listing what is depicted in a media file.
- “[Properties](#).” A property is the key in the key-value pair, and can be thought of as a category of data – for example the property “color” could have the value “blue.” Properties, when paired with values, form statements – which result in a linked data structure.

In the first image example below, a Commons user who's searching by keyword types "Ca" into the search box, and gets suggested media files for searching. We're maintaining this kind of searching:

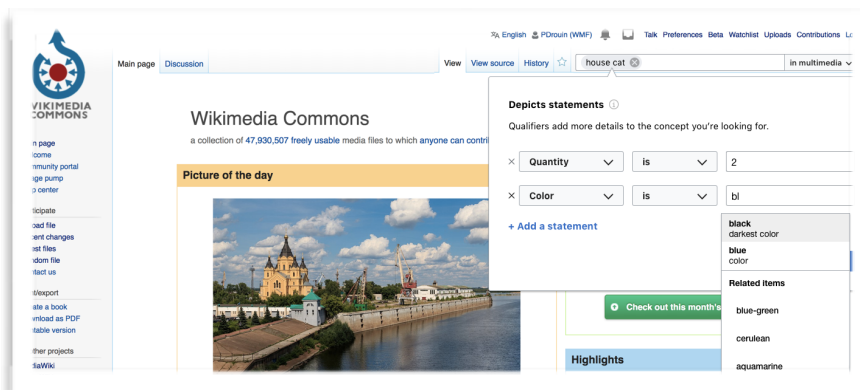
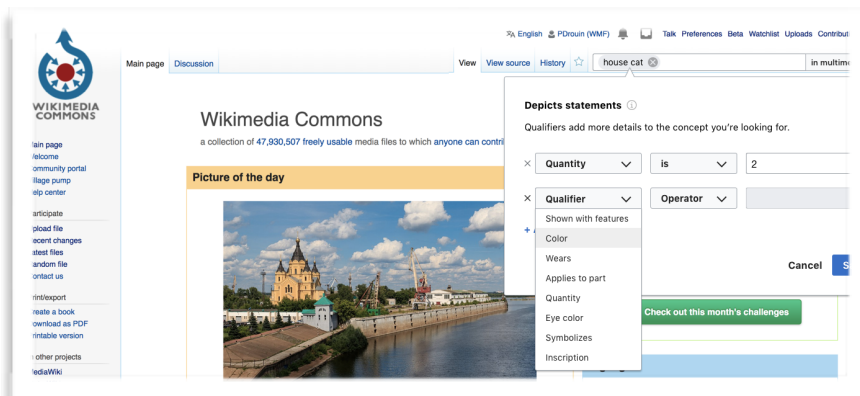
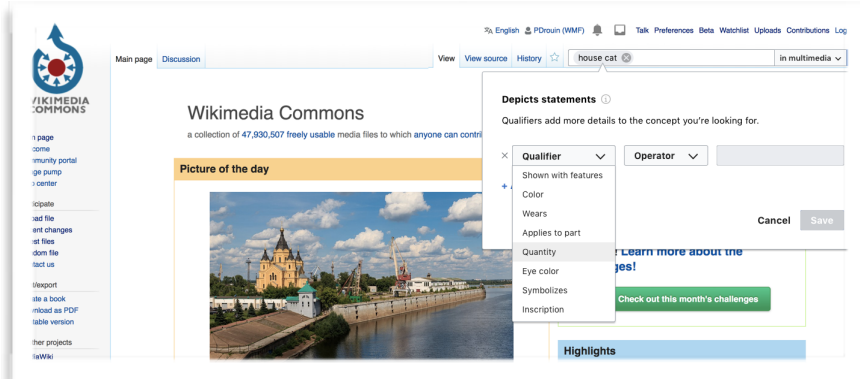


In the mockup below, the user has switched to the new search – “Depicts” search – and this searches Wikidata items. This is structured searching.



In the three examples on the following page, the Commons user narrows the search even more through a “qualifier” that lets the user better specify the search – in this case, the quantity of cats (2) and the color of cats (black).

### Three examples of narrowing a search through qualifiers:



As indicated by the previous prototype designs, we’re building a new interface that will let users specify depicts statements and qualifiers that apply to it. We’re also adding features in UploadWizard, the File page, and other areas to encourage users to add depicts statements with useful qualifiers so that this kind of search can work, perhaps with the assistance of suggestions from image content recognition systems. We’re building alpha working versions that allow us to test performance so we know exactly what’s feasible in terms of response time, server load, and other factors. We hope to have this deployed to Beta Commons by the end of 2018 and production Commons by February 2019.

We [published a document](#) that explains the new data types on Commons, the main types of structured data that can be added to Commons media files, and the technical indexing that will let users work with structured data to find what they’re looking for, whether it’s searching for a single statement, across multiple statements, or another type of search request. In August, when we [requested feedback](#) from Commons users about depicts, we published a helpful, 35-slide deck that explains depicts – including the features we’ll have in the initial release, later features, how depicts will give structure to media descriptions, and our UI requirements. We explain how, in later releases, depicts will aid batch uploads and editing, and how it will enable linked editing from Commons structured data to Wikidata structured data.

In our search interface, we’re incorporating qualifiers, like “quantity” and “color,” because these qualifiers help explain and contextualize each depict statement, as in the example of the cat search on the previous page, where the cat’s color and the number of cats are qualifiers that aid in searching. The number of qualifiers is currently limited, so to make search more effective on Commons, we plan to add more qualifiers – for example, “shape” or “eye color.”

As we move ahead with orchestrating search on Commons, we’re considering the many new factors that will go into structured searching on Commons, including the use of Wikidata items, Commons items, properties, qualifiers, and depicts statements. This is what we are now working out, exemplified by the Commons image of *The Art Gallery of Jan Gildemeester Jansz*, an 18th-century painting by the Dutch artist Adriaan de Lelie. The photo of the painting is [on Commons](#) and also anchors Wikipedia pages, including the [English-language article on the painting](#). De Lelie’s image is on the following page:



For structured searching, here's what we have to consider for this one painting:

- It's a scan of a painting that has its own Wikidata Q item
- The painting contains 15 other paintings that also have their own Wikidata Q items
- The painting also depicts numerous people, each with their own Wikidata Q items
- On Commons, the [Image Annotation feature](#) is used to [call out](#) each of these contained elements
- At a later date, Image Annotations will have to be tied into depicts as part of “structuring” them

For artwork on Commons, we'll need to reference existing depicts statements from Wikidata for paintings, statues, and other objects that have Q items. And we'll need to make it clear to the user that those additional depicts statements came from the artwork's Q item.

The intricate connection between Commons and Wikidata is driving our work on search, and is one focus in our outreach to Commons users as we get [their feedback](#). The [community feedback](#) on our search prototype design, which is influencing our designs as we move ahead, dovetails from [community feedback](#) about having depicts as a central naming and searching elements. Three weeks ago, for example, one longtime Commons user, [Jheald](#), [asked](#) whether a search for “all images of paintings by Leonardo da Vinci” would pull information entirely from Commons or whether Wikidata would also be involved. The answer is complex, we explained, and involves incorporating Q items, indexing statements for search engines, integrating them into the User Interface, and other factors. As one of the first platforms that is doing extensive indexing of federated data, we're solving for new, complicated issues. Structured Data on Commons [Product Manager Ramsey Isler](#) responded to Jheald right away, noting that our work on search is a work in progress.

“This question has a short answer and a long answer,” Ramsey noted. “The short answer is that, in this case, paintings by da Vinci would be denoted via depicts statements on the file, and for the artwork scenario the current plan (which could change) is to automatically place certain statements from the artwork’s Q item into the Elasticsearch index for the media file. Let’s say for example it’s Mona Lisa. The scan/photo on Commons will have a statement saying ‘depicts Mona Lisa.’ The Mona Lisa Q item’s statements will then automatically be added to the search index for that file. We don't know yet which statements this will apply to (maybe all? maybe some? still working on that).”

Then Ramsey elaborated. “The long answer is more complex,” he noted. “In addition to adding data to the search index, we have to answer whether there should indeed be some data on the M item (the MediaInfo item stored in Wikibase@Commons) referenced or copied from the Q item on Wikidata. There is certainly a reasonable argument to be made against this. However, there are some other things we have to consider too. There are no decisions made here yet, but one idea that has been lingering for a while is the concept of “faked statements”; they act like statements in the API and UI but are in actuality derived from another source and managed a different way. Theoretically, we might enhance Wikibase Federation with a faked statement that “links” the Creator data on the Wikidata Q item to the M item on Commons. The Wikidata item remains the canonical source but the faked statement provides a straightforward way to show relevant Wikidata data (and perhaps even provide a link to an on-Commons interface to edit said data), and we can play with UI designs to make a clear distinction between the data on the file itself and the data on the things it refers to. But, again, this is still stuff we’re speccing out and testing.”

Spec'ing out and testing. That's what we're doing on search, in conjunction with other layers of work that are giving us a real-time sense of what is working and what we still have to work on.

## More About Search

### Community Feedback on our Design Prototype:

[https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Get\\_involved/Feedback\\_requests/Search\\_prototype](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Get_involved/Feedback_requests/Search_prototype)

### Community Feedback on depicts:

[https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Feedback\\_requests/Depicts](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Depicts)

### 35-page slide deck that explains depicts:

[https://upload.wikimedia.org/wikipedia/commons/a/a5/Structured\\_Data\\_on\\_Commons\\_-\\_%E2%80%9CDepicts%E2%80%9D\\_CC\\_spec.pdf](https://upload.wikimedia.org/wikipedia/commons/a/a5/Structured_Data_on_Commons_-_%E2%80%9CDepicts%E2%80%9D_CC_spec.pdf)

### Phabricator ticket for implementing depicts search on Commons:

<https://phabricator.wikimedia.org/T191633>

### Phabricator ticket for design research for search:

<https://phabricator.wikimedia.org/T185059>

### Phabricator ticket for implementing searching of multilingual captions Commons:

<https://phabricator.wikimedia.org/T187438>

### Phabricator ticket for building search features for Structured Data on Commons:

<https://phabricator.wikimedia.org/T185025>

### Phabricator ticket for adding new search features and specifications on Commons:

<https://phabricator.wikimedia.org/T201450>

### Glossary of relevant terms from Wikidata:

<https://www.wikidata.org/wiki/Wikidata:Glossary>

## Research

For more than a decade, Wikimedia has had welcome pages and instructional pages on our sites that are geared toward GLAM representatives. These pages (like [this one](#) and [this one](#)) encourage and explain how to add content to our sites, including Commons – which Commons users have used to add images, and audio files, and film clips. The past decade has seen a steady growth in the number of Wikipedians who are working as [Wikipedians-in-Residence](#) with GLAM partners. The Wikimedia community's partnerships with GLAM institutions are more solid than ever. But the research that we've done on GLAM institutions in the past year for our Structured Data on Commons project has been the first to really articulate and detail GLAM institutions' use of Commons –



to suss out the complexities of that use, and examine it from all angles, from successes to frustrations.

Led by [Senior Design Researcher Jonathan Morgan](#), the research delved into GLAM use of Commons and into other contributors' usage of a site that we're transforming through Structured Data on Commons:

- [How GLAM institutions use Commons](#). We did surveys, workshops, and interviews from June 2017 to October 2017, then analyzed the findings to make [key observations](#), [detail users' challenges](#), and [make recommendations](#) on how to help GLAMs work better with Commons.
- [How organizations and individuals remix, repurpose, or otherwise incorporate visual media into their own creative works](#). The research resulted in a series of findings on reuse, the type of media that they use, and the way they use Commons.
- [How Commons contributors curate media](#). We studied the workflows of Commons contributors to identify opportunities to support these workflows better with new software or new metadata, and to avoid disruption of critical workflows during the transition to storing metadata on Commons.
- [How users follow our Upload Wizard instructions](#). We researched how to phrase instructions in the [Upload Wizard](#) interface, in order to reduce potential confusion between the "caption" and "description" fields.

During the past year, Jonathan also led the publication of [Commons' GLAM Personas](#) – profiles of four types of GLAM contributors that give us more insight into how we should retool Commons during this project. The studies and the personas have a definite practical application: They're informing the design of new software features and functionality for searching, browsing, downloading, and re-publishing Commons media. Each study revealed ways that Commons users like the current features and other ways that they dislike them. We needed to know that. And now we do – as with the study of GLAM institutions' use of Commons. We referenced this study in last year's report, and its completion in January 2018 is a major research accomplishment because of its intricate findings.

The report's three major findings:

- **GLAM projects are interested in impact.** Donating media to Commons is a means to an end. GLAM organizations and the volunteers who work with them want to know the media they upload is being used, and to be able to evaluate the impact of their donations against institutional goals.
- **Metadata types are (at least) as diverse as media types.** A few basic types of metadata are available for most collections (e.g., GLAM institution, license deed) and media items (e.g., creator, date of creation). Beyond that, there is a huge diversity of different kinds of metadata available for different kinds of media – and much of this metadata is considered to be vitally important. One size does not fit all.
- **There are different degrees of technical and Wikimedia literacy.** The project members involved in uploading and curating GLAM media on Commons vary widely in their level of familiarity with the Wikimedia Movement (its rules, norms, and community resources), and in their level of proficiency with technical tools and techniques for metadata preparation, batch upload, post-upload curation, and content monitoring.

Based on these findings, we articulated three major challenges as we continued with the project:

- **Tracking the impact of a media donation.** Although a variety of tools exist for keeping track of what happens to media after upload, these tools do not always capture the right metrics, at sufficient granularity, to demonstrate the impact of a donation or monitor the current status of an uploaded collection. These tools also suffer from a lack of visibility, and to some extent a lack of integration into the Commons platform.
- **Capturing metadata in (semi)structured ways.** Most GLAM projects use categories and templates to some extent to capture item- and collection-level metadata. However, their efforts are hampered by the complexity of the [category system](#), and its lack of standardization. In contrast, existing metadata templates are sometimes too standardized to capture the rich, diverse metadata associated with different media. Participants resort to clever workarounds to fit their important metadata into these existing structures as best they could.
- **Demonstrating and preserving media provenance.** Many GLAM participants expressed confusion and frustration with the process of demonstrating to [OTRS](#) (Commons' license-review system) that the GLAM institution could legally donate the uploaded media. Once a donation

was approved and uploaded, some GLAM participants expressed a desire for more flexibility in the mechanisms available for linking the collections on Commons back to the GLAM's website or institutional repository.

Based on these findings, we articulated three major ways we can help GLAM contributors:

- **Show the impact of a media donation.** Develop tools that enable GLAMs to easily see the metrics that are important to them – media use, metadata added, etc.
- **Make uploading metadata easier.** There are many well-reviewed upload tools, specialized for uploading large batches of media, in active use by GLAM organizations. These are built and maintained by volunteer developers within the Wikimedia movement. We plan to work with these developers to transition their tools to upload structured metadata in ways that make sense for GLAMs. In addition, in our redesign of Upload Wizard, we are making sure it will work for GLAMs who want to upload just a few images, with very rich metadata.
- **Consolidate and update batch upload help documentation.** Currently, help resources for the various steps in the batch upload workflow are scattered across namespaces and wikis, out of date, incomplete, or non-existent. Every uploader's workflow is different, but a single well-designed and well-maintained entry point that links to relevant information about process steps and common pitfalls will benefit everyone.

Lastly, our study on Upload Wizard instructions had a particularly practical result. We did the controlled study of the Upload Wizard with participants from [Amazon Mechanical Turk](#), and asked them to judge phrase instructions in the [Upload Wizard](#) interface. The main research aim: Investigate how they responded to the directional, “Add a one-line explanation of what this file represents, including only the most relevant information” versus a directional that said, “Add a short phrase to convey what this file represents, including only the most relevant information.” We wanted to discover whether different instructions led to captions of different lengths, and whether the quality of captions was impacted by the instructions. The result: The one-line explanation instruction produced better quality captions on average, so we will use the phrase one-line explanation in the Upload Wizard instructions for eliciting captions. It's important to capture good-quality captions from Commons contributors at the point of upload. Getting good captions in the beginning – rather than having Commons users edit them after they've been uploaded – is a small but significant factor in Commons' retooling for structured data.

## More About Research

**Research on supporting GLAM institutions' Commons contributions:**

[https://meta.wikimedia.org/wiki/Research:Supporting\\_Commons\\_contribution\\_by\\_GLAM\\_institutions](https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions)

**Research on Curation workflows:**

[https://meta.wikimedia.org/wiki/Research:Curation\\_workflows\\_on\\_Wikimedia\\_Commons](https://meta.wikimedia.org/wiki/Research:Curation_workflows_on_Wikimedia_Commons)

**Research on technical needs of external re-users of Commons media:**

[https://meta.wikimedia.org/wiki/Research:Technical\\_needs\\_of\\_external\\_re-users\\_of\\_Commons\\_media](https://meta.wikimedia.org/wiki/Research:Technical_needs_of_external_re-users_of_Commons_media)

**Research on Crowdsourced evaluation of Upload Wizard instructions:**

[https://meta.wikimedia.org/wiki/Research:Crowdsourced\\_evaluation\\_of\\_Upload\\_Wizard\\_instructions](https://meta.wikimedia.org/wiki/Research:Crowdsourced_evaluation_of_Upload_Wizard_instructions)

**Research on the four GLAM Persona types that use Commons:**

[https://commons.wikimedia.org/wiki/File:GLAM\\_Personas\\_for\\_Structured\\_Data\\_on\\_Commons\\_v2.pdf](https://commons.wikimedia.org/wiki/File:GLAM_Personas_for_Structured_Data_on_Commons_v2.pdf)

**Research on workarounds by GLAM partners to get media files on Commons:**

[https://meta.wikimedia.org/wiki/Research:Supporting\\_Commons\\_contribution\\_by\\_GLAM\\_institutions/Preserving\\_important\\_metadata\\_about\\_media\\_items](https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions/Preserving_important_metadata_about_media_items)

**Research on critiques of the Upload Wizard:**

[https://meta.wikimedia.org/wiki/Research:Supporting\\_Commons\\_contribution\\_by\\_GLAM\\_institutions/Functionality\\_and\\_usability\\_of\\_batch\\_upload\\_tools](https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions/Functionality_and_usability_of_batch_upload_tools)


## Other Design Work

In addition to the design work detailed on previous pages, we also completed a series of other design testing and mock-ups, including:

**Designs for Licenses:** We designed mockups of what Structured Licenses might look like, asking users how intuitive the designs were, whether Commons users thought more information should be added, whether they knew how to edit the licenses, and whether they knew what all the information meant. We received lots of [feedback](#), and some of the discussion is impacting decisions on where structured data is laid out on the page.

One example: Instead of scrolling down for the license, as we had originally planned, we will likely implement a tabbed interface – not just to reduce scrolling, but to keep depicts close to the image itself, which should make it easier for users to accurately describe files.

Image below: A view of a test image with accompanying file information, including a license at the end.




COMMONS

English Username Talk Preferences Beta Watchlist Uploads Contributions Log out

File Discussion
View Edit History More

**File:**

### Louvre at dusk.JPG



Louvre at dusk. Louvre pyramid, Cour Napoléon, Palais du Louvre, 1er arrondissement, Paris, France.

[Download this file](#)

[Re-use information](#)

[Share on internet](#)

[Use on a Wiki](#)

<b>Uploader</b>	<b>Type</b>
Username	image / jpg

**Public domain**

Free for use by anyone for any purpose without restriction under copyright law. [Full license](#)

**Direct URL**

<https://commons.wikimedia.org/wiki/File%3AL...>

**Attribution for re-use** REQUIRED

Username / Wikimedia Commons / "The Louvre a...

File information
Metadata

#### Metadata [ back to file info ]

**depicts** P180 · Edit

from Wikidata

**Louvre Pyramid** Q19675 · Make primary

Usage justification — de minimis

**usage rights** P## · Edit

**Creative Commons Attribution-Share Alike 3.0 Unported** Q##### · Make primary

Obligations — attribution, share alike


Usage justification — own work

Author — Michael Falbisoner

Attribution text — Photo by Michael Falbisoner

Attribution link — http://- - -.com

[Add another statement](#)



**What's metadata?**

---



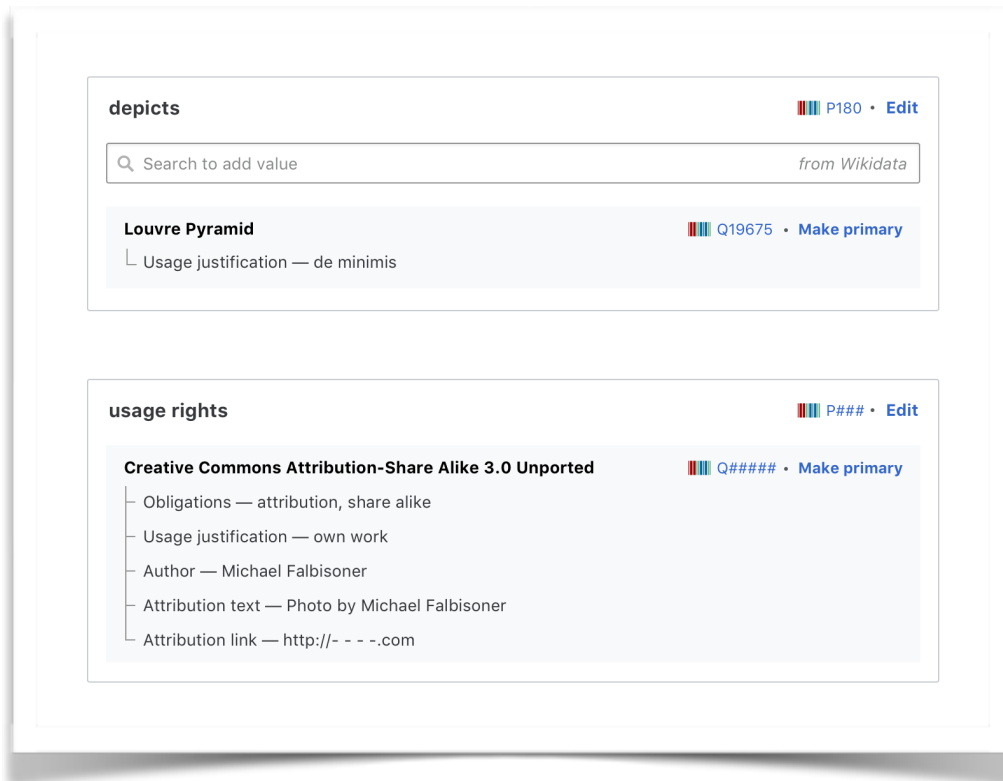
---



---

Learn more

Image below: A close-up of that license.



## Community and GLAM Stakeholder Input

Community input and input from GLAM stakeholders has been vital to every part of the Structured Data on Commons project – and has been a part of the project from the beginning. Through online channels and public talks, through regular meetings and regular requests for comments, we've ensured that people can contribute their ideas, ensured that we actively consider them, and ensured that Structured Data on Commons meets all of our users' key needs.

Community input has been essential because it's the community that will build the structured data content on Commons. It's the community's use that will herald a new era for Commons. In the past year, input channels included:

- Structured Data on Commons presentations, conversations and workshops at Wikimania, the EuropeanaTech Conference, and other key conferences and events.
- An [ongoing focus group comprised of GLAM representatives](#).
- An [ongoing Commons Community focus group](#).
- [Regular, public IRC chats](#), where the team takes questions and give updates on the project.
- Talk pages where Commons users make suggestions, leave critiques, and have online conversations with each other and with Structured Data on Commons staff members, as in the [talk page on Properties table](#).
- Request for comment pages, where Commons users respond to requests for feedback on new designs and other developments, as in [the feedback on ontology](#), where we proposed a three-part system of metadata storage.

By having regular communication channels open, and by doing public talks and proactive outreach to institutions around the world, we've garnered valuable buy-in to the project – so much so that some institutions have expressed interest in funding and building tools that would improve Structured Data on Commons. One example: The [Swedish National Heritage Board](#) will invest its own resources in a prototype to upload and ingest media and metadata to and from Commons. Based on the updated MediaInfo API, the tool will allow GLAMs to have back-and-forth synchronization of updated metadata between Commons and their own collection database.

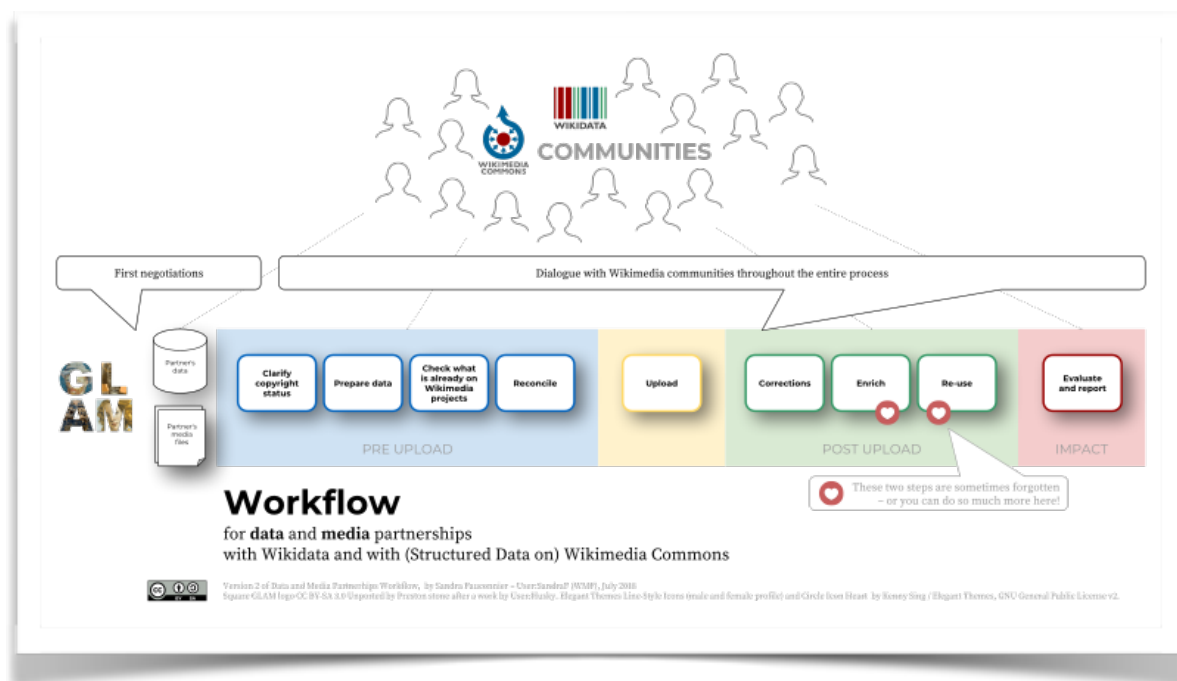
In fact, we're discussing with several institutions this concept of "round-tripping metadata". It would be another valuable reason for GLAMs to contribute media and metadata, since with a metadata round-tripping tool they could enrich the metadata of their own collections with Commons users' metadata additions (from id'ing people in images to naming objects). Cultural institutions have asked for this kind of interconnectedness for a long time. Now that our Structured Data on Commons project is becoming a full-fledged reality, they're taking the initiative themselves in

developing such technologies; without us even asking. “It’s exciting to see,” says [GLAM Program Officer Sandra Fauconnier](#). “It shows we’re pushing the right buttons.”

More “right buttons”: In the next year, we will mentor 5-8 GLAM pilot projects to use Structured Data on Commons. The pilot projects will be a diverse selection GLAM-Wiki projects that cover many typical project types. These will be described and documented extensively, so that other community members and GLAM staff members can replicate them and learn from them.

Based on the research by Jonathan Morgan on GLAM use of Wikimedia Commons, we also started preparing a conceptual workflow framework for documentation. We iterated several versions of a [general workflow for Wikimedia data and media partnerships for GLAMs](#), which aims to clarify this workflow for community members, tool developers and for GLAM staff, and which will serve as a basis for updated documentation on batch uploads to Wikimedia Commons in 2019.

**Image below: The workflow for Wikimedia data and media partnerships for GLAMs**





## More About Community Input

- Community input that helped us prioritize which tools are important for the Wikimedia Commons and Wikidata communities. The survey ran from December 5-22, 2017. [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Tools](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Tools) and <https://docs.google.com/spreadsheets/d/1GVR0jgh-BWuAGqJaT7KVXigMYWWNzdnrwl9nWqfJrCo/edit#gid=0>
- Community input on interesting media files (and other interesting phenomena) on Wikimedia Commons, as inspiration for modeling these cases in structured data in the future. December 8 to present. [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Feedback\\_requests/Interesting\\_Commons\\_files](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Interesting_Commons_files)
- Community input on how to name two new fields that will be added to Commons file pages. The fields' original names: "captions" and "descriptions." The survey ran from December 13, 2017 to January 3, 2018. [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Feedback\\_requests/Renaming\\_%27captions%27\\_and\\_%27descriptions%27](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Renaming_%27captions%27_and_%27descriptions%27)
- Community input on where Commons' metadata should go in the Structured Data on Commons system, spread across three main repositories: MediaWiki; Wikibase@Commons; Wikidata. The discussion opened on February 15 and closed on March 1. [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Feedback\\_requests/Ontology](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Ontology)
- March 15 to April 2, 2018 discussion about structured metadata for copyright and licenses on Wikimedia Commons. [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Get\\_involved/Feedback\\_requests/First\\_licensing\\_consultation](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Get_involved/Feedback_requests/First_licensing_consultation)
- April 19 to May 4, 2018 discussion about whether to prioritize mapping GLAM metadata schemes and ontologies to Wikidata and to structured data on Commons. [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Get\\_involved/Feedback\\_requests/GLAM\\_metadata\\_and\\_ontologies\\_mapping](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Get_involved/Feedback_requests/GLAM_metadata_and_ontologies_mapping). Follow-up summary written by Sandra Fauconnier: [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Get\\_involved/Feedback\\_requests/GLAM\\_metadata\\_and\\_ontologies\\_mapping#Summary\\_and\\_next\\_steps](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Get_involved/Feedback_requests/GLAM_metadata_and_ontologies_mapping#Summary_and_next_steps) and <https://phabricator.wikimedia.org/T192537>
- Design input on multilingual file captions, which opened on May 21: [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Get\\_involved/Feedback\\_requests/Multilingual\\_Captions\\_and\\_MediaInfo](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Get_involved/Feedback_requests/Multilingual_Captions_and_MediaInfo)
- Discussion that began on June 25 about how "properties" should be used on the project: [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Feedback\\_requests/Properties\\_for\\_-\\_Commons](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Properties_for_-_Commons)
- Follow-up discussion on "properties table" that began on August 29, 2018. [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Properties\\_table](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Properties_table)

# How We're Doing:

## Our Progress So Far

Structured Data on Commons is a multi-layered, multi-year project involving a [team](#) of software engineers, product managers, designers, researchers, and others (see also Page 52 of this report) – and we're driving our progress through quarterly and annual plans that have manifest a multitude of Phabricator tasks. In an overarching sense, though, we're measuring our progress in three fundamental ways:

- We have to build the infrastructure for the project.
- We have to build the features for the project.
- We have to get people to use the project.

In the past year, we built the infrastructure, which is Multi-Content Revisions. And we are already building the features for Structured Data on Commons. People are testing the first one, multilingual file captions. More features – and more testing – are on the horizon. We are sufficiently on track. In the past year. Here are the milestones we've achieved.

## The Year in Review: A Timeline of Key Developments and Outreach

Date	Development
Oct. 4-30, 2017	<b>GLAM survey released:</b> <a href="#">Senior Design Researcher Jonathan Morgan</a> releases the first survey of GLAM institutions, which asked them about their processes, wishes, and problems when contributing media to Wikimedia Commons. The survey parallels individual interviews that Jonathan did. Sixty-seven people started the survey and 44 respondents completed it. (Our 2017 SDC report to the Sloan Foundation cited the survey and interviews, but the survey and interviews were still in progress at that time.)
Oct 31, 2017	<b>Features for MVP determined:</b> Both the dev team and the Commons community wanted the first feature to be as small as possible while integrating upload, editing and search features. Thus, we chose and scoped multilingual file captions. We decided to release basic depicts and other statement functionality next, because it is the most fundamental feature.
Nov.-Dec. 2017	<b>GLAM survey analyzed:</b> Researcher Jonathan Morgan analyzes data from the survey of GLAM institutions.
Nov. 21, 2017	<b>IRC office hours:</b> In a live IRC session, four members of the Structured Data on Commons project – <a href="#">GLAM Program Officer Sandra Fauconnier</a> , <a href="#">Program Manager Amanda Bittaker</a> , <a href="#">Product Manager Ramsey Isler</a> , and <a href="#">Community Relations Specialist Keegan Peterzell</a> – answer the community’s questions about the project. <a href="https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2017-11-21">https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2017-11-21</a>
Dec. 3-Jan. 8	<b>Continuation of file info prototype:</b> We continue the preliminary work to prototype changes for file info pages. The work continued for a month. <a href="https://phabricator.wikimedia.org/T177005">https://phabricator.wikimedia.org/T177005</a>
Dec. 4-Jan. 29	<b>Research on adding files:</b> User Experience Designer <a href="#">Pamela Drouin</a> conducts research on design changes to the Upload wizard, which is the default method for uploading files to Wikimedia Commons. <a href="https://phabricator.wikimedia.org/T182019">https://phabricator.wikimedia.org/T182019</a> <a href="https://commons.wikimedia.org/wiki/Commons:Upload_Wizard">https://commons.wikimedia.org/wiki/Commons:Upload_Wizard</a>
Dec. 5-22	<b>Tools discussion:</b> Community input is taken that helped us prioritize what tools are important for the Wikimedia Commons and Wikidata communities. <a href="https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Tools">https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Tools</a>

Date	Development
Jan. 3 - ongoing	<p><b>Partner conversations:</b> Sandra Fauconnier, <a href="#">Community Programs Director Ben Vershbow</a>, and <a href="#">GLAM Strategist Alex Stinson</a> are having continued chats and discussions with the Linked Data for Libraries community, Smithsonian, Europeana, and several other partners.</p>
Jan. 9, 2018	<p><b>MCR changes database schema:</b> Added the architecture for multiple slots per revision to the database schema, so structured data and wikitext can exist on same page.  <a href="https://www.mediawiki.org/wiki/Multi-Content_Revisions/Database_Schema">https://www.mediawiki.org/wiki/Multi-Content_Revisions/Database_Schema</a></p>
Jan. 29, 2018	<p><b>GLAM survey published:</b> Led by Jonathan Morgan, we published results of the GLAM survey, writing about it a Jan. 29, 2018 blog post, and posting a video and slides that explain the work.  <a href="https://blog.wikimedia.org/2018/01/29/glam-multimedia-metadata-commons/">https://blog.wikimedia.org/2018/01/29/glam-multimedia-metadata-commons/</a>  <a href="https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions">https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions</a>  <a href="https://www.youtube.com/watch?v=Mx27yTqw7ro">https://www.youtube.com/watch?v=Mx27yTqw7ro</a>  <a href="https://commons.wikimedia.org/wiki/File:GLAM_SDC_research_presentation_Jan_2018.pdf">https://commons.wikimedia.org/wiki/File:GLAM_SDC_research_presentation_Jan_2018.pdf</a></p>
Feb. 13, 2018	<p><b>IRC office hours:</b> In a live IRC session, four members of the Structured Data on Commons project – Sandra Fauconnier, Amanda Bittaker, Ramsey Isler, and Keegan Peterzell – answer the community’s questions about the project.  <a href="https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2018-02-13">https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2018-02-13</a></p>
Feb. 15-March 1, 2018	<p><b>Metadata discussion:</b> Community input is received on where Commons’ metadata should go in the Structured Data on Commons system. <a href="https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Ontology">https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Ontology</a></p>
Feb. 28-March 1, 2018	<p><b>Wikidata Lesson Plans:</b> As part of an ongoing collaboration with the Wikidata team at Wikimedia Deutschland to develop teaching materials that can help GLAM and other communities contribute to Wikidata, Alex Stinson introduced a lesson plan for teaching Wikidata workshops, and wrote a guide for helping teach the Wikidata Query service.  <a href="https://www.wikidata.org/wiki/Wikidata:Planning_a_Wikidata_workshop">https://www.wikidata.org/wiki/Wikidata:Planning_a_Wikidata_workshop</a>  <a href="https://outreach.wikimedia.org/wiki/File:Wikidata_Query_Service_in_Brief.pdf">https://outreach.wikimedia.org/wiki/File:Wikidata_Query_Service_in_Brief.pdf</a></p>
March 15 to April 2, 2018	<p><b>License discussions:</b> First community discussion is held to identify the best way to build a model for structured license and copyright information. <a href="https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/First_licensing_consultation">https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/First_licensing_consultation</a></p>

Date	Development
April 21, 2018	<p><b>Wikimedia Conference:</b> Members of the Structured Data on Commons project – Sandra Fauconnier, <a href="#">User Experience Designer Pam Drouin</a>, and Alex Stinson – along with Wikimedia staff member Ben Vershbow host an SDC workshop at the Wikimedia Conference in Berlin. <a href="https://meta.wikimedia.org/wiki/File:WMCON_2018_-_Structured_Data_on_Commons_and_GLAM_pilots_session.pdf">https://meta.wikimedia.org/wiki/File:WMCON_2018_-_Structured_Data_on_Commons_and_GLAM_pilots_session.pdf</a> and <a href="https://meta.wikimedia.org/wiki/Wikimedia_Conference_2018/Program/47">https://meta.wikimedia.org/wiki/Wikimedia_Conference_2018/Program/47</a></p> <p>During the Learning Days at the conference, Sandra also workshopped a first iteration of the GLAM and structured data workflow diagram with a group of community members.</p> <p><a href="https://meta.wikimedia.org/wiki/Learning_and_Evaluation/News/Learning_Days_for_Wikimedia_Conference_2018/Thursday_sessions#Wikidata-Commons">https://meta.wikimedia.org/wiki/Learning_and_Evaluation/News/Learning_Days_for_Wikimedia_Conference_2018/Thursday_sessions#Wikidata-Commons</a></p>
May 10-12, 2018	<p><b>Armenia talks:</b> Sandra Fauconnier co-leads a Wikidata workshop at a GLAM Forum in Yerevan, Armenia, and also gives a presentation about GLAM work in The Netherlands and Belgium.</p> <p><a href="https://meta.wikimedia.org/wiki/GLAM_forum_in_Yerevan/Program">https://meta.wikimedia.org/wiki/GLAM_forum_in_Yerevan/Program</a></p>
May 14, 2018	<p><b>Europeana workshop in the Netherlands:</b> Sandra Fauconnier leads two workshops at the Europeana offices in the Hague, Netherlands – one on GLAM ontologies and vocabularies, and copyright modelling; the other titled, "Stronger Together: Wikidata as a vocabulary and authority control hub."</p> <p><a href="https://www.wikidata.org/wiki/Wikidata:Events/EuropeanaTech_2018_Wikidata_workshop_day">https://www.wikidata.org/wiki/Wikidata:Events/EuropeanaTech_2018_Wikidata_workshop_day</a></p>
May 18, 2018	<p><b>Wikimedia Hackathon:</b> At the year's biggest Wikimedia hackathon – the Wikimedia Hackathon in Barcelona – Sandra Fauconnier and Amanda Bittaker led a session devoted to Structured Data on Commons. And the SDC team led a discussion on creating "properties" for Structured Data on Commons.</p> <p><a href="https://etherpad.wikimedia.org/p/GLAM-SDC-wmhack18">https://etherpad.wikimedia.org/p/GLAM-SDC-wmhack18</a>  <a href="https://etherpad.wikimedia.org/p/Structured_Data_Commons_Hackathon_2018">https://etherpad.wikimedia.org/p/Structured_Data_Commons_Hackathon_2018</a>  <a href="https://phabricator.wikimedia.org/T194933">https://phabricator.wikimedia.org/T194933</a>  <a href="https://www.mediawiki.org/wiki/Wikimedia_Hackathon_2018/Structured_Commons">https://www.mediawiki.org/wiki/Wikimedia_Hackathon_2018/Structured_Commons</a></p>
May 31, 2018	<p><b>Personas Published on Commons:</b> Wikimedia's research team published four GLAM contributor personas to Commons. A persona is a fictional character that represents a specific kind of user, and the GLAM personas are four typical GLAM staff members that contribute to Wikimedia Commons. These personas can help anyone (including developers, designers, and organizers) better understand the people who contribute to Wikimedia Commons from a GLAM perspective.</p> <p><a href="https://commons.wikimedia.org/wiki/File:GLAM_Personas_for_Structured_Data_on_Commons_v2.pdf">https://commons.wikimedia.org/wiki/File:GLAM_Personas_for_Structured_Data_on_Commons_v2.pdf</a></p>

Date	Development
June 26, 2018	<p><b>IRC office hours:</b> In a live IRC session, four members of the Structured Data on Commons project – Sandra Fauconnier, Amanda Bittaker, Ramsey Isler, and Keegan Peterzell – answer the community’s questions about the project.  <a href="https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2018-06-26">https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2018-06-26</a></p>
July 15, 2018	<p><b>Workflows chart:</b> In the days before Wikimania, we publish a chart that explains the workflow for GLAM partnerships that involve structured data (Wikidata, and Wikimedia Commons in the future).  <a href="https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow">https://outreach.wikimedia.org/wiki/GLAM/Resources/Data_and_media_partnerships_workflow</a></p>
July 18-22, 2018	<p><b>Wikimania:</b> Members of the Structured Data on Commons project present talks and workshops at the year’s most important international gathering, which took place in Cape Town, South Africa.            July 18-19, Sandra Fauconnier – Hackathon:  <a href="https://wikimania2018.wikimedia.org/wiki/Hackathon/Structured_Commons">https://wikimania2018.wikimedia.org/wiki/Hackathon/Structured_Commons</a>            July 20, Sandra Fauconnier, Amanda Bittaker – overall presentation: <a href="https://wikimania2018.wikimedia.org/wiki/Program/Structured_Data_on_Wikimedia_Commons_and_knowledge_equity">https://wikimania2018.wikimedia.org/wiki/Program/Structured_Data_on_Wikimedia_Commons_and_knowledge_equity</a>            July 20, Sandra Fauconnier and Jonathan Morgan – presentation about research: <a href="https://wikimania2018.wikimedia.org/wiki/Program/Wikimedia_Commons_and_GLAM_needs_around_the_world">https://wikimania2018.wikimedia.org/wiki/Program/Wikimedia_Commons_and_GLAM_needs_around_the_world</a></p>
July 18, 2018	<p><b>Pilot talks:</b> At Wikimania, Sandra starts the first planning-oriented conversation with community members and GLAM representatives who are interested in working on GLAM pilot projects with Structured Data on Commons in 2019.</p>
Aug. 8, 2018	<p><b>MCR storage layer ready:</b> Completed code necessary to write to the new schema.</p>
Aug. 22, 2018	<p><b>Wikis start writing to old and new schema:</b> Began writing to both schemas to test new schema while remaining operational with old schema.</p>
Aug. 29, 2018 - present	<p><b>Properties consultation and table:</b> Facilitated conversation with Commons, Wikidata and GLAM communities. <a href="https://commons.wikimedia.org/wiki/Commons:Structured_data/Properties_table">https://commons.wikimedia.org/wiki/Commons:Structured_data/Properties_table</a></p>
Sept. 12, 2018	<p><b>MCR file caption ready:</b> Completed the code that was necessary for the first feature, file captions.</p>
Sept. 15, 2018	<p><b>File page multilingual captions alpha:</b> Completed alpha of new file page functionality for Structured Data on Commons. <a href="https://federated-commons.wmflabs.org/wiki/File:LighthouseinDublin2.jpg">https://federated-commons.wmflabs.org/wiki/File:LighthouseinDublin2.jpg</a>  <a href="https://phabricator.wikimedia.org/T201456">https://phabricator.wikimedia.org/T201456</a></p>
Sept. 21, 2018	<p><b>MCR statement ready:</b> Completed code that was necessary to support structured statements on Commons.</p>

Date	Development
Oct. 1, 2018	<b>MCR deployed to mediawiki.org:</b> Deployed the MCR code to the production cluster, and ensured that <a href="https://mediawiki.org">mediawiki.org</a> , a production wiki, is using the new code and schema.
Oct. 4, 2018	<b>Alpha version of Search with depicts:</b> Code developed on Labs, first working version of searching for media with structured data statements and qualifiers <a href="https://federated-commons.wmflabs.org/">https://federated-commons.wmflabs.org/</a> (search bar in upper right). <a href="https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Search_prototype">https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Feedback_requests/Search_prototype</a>
Oct. 4, 2018	<b>IRC office hours:</b> In a live IRC session, three members of the Structured Data on Commons project – Amanda Bittaker, Ramsey Isler, and Keegan Peterzell – answer the community’s questions about the project. <a href="https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2018-10-04">https://meta.wikimedia.org/wiki/IRC_office_hours/Office_hours_2018-10-04</a>
Oct. 8, 2018	<b>MCR deployed to other three wikis:</b> We deploy the new code and schema to three additional small- to medium-sized production wikis (igwiki, eswikivoyage, Itwikisource).
Oct. 31, 2018	<b>Integrate full stack:</b> On Beta Commons, we integrated and released the full backend and front-end stack (MCR and file captions).

## How We’re Measuring Success: Deliverables Oct. 2017 to Sept. 2018

In our 2016 proposal, we said that, by 2020, “the infrastructure will be in place to migrate Commons’ millions of media files to structured data.” We are on track for that, and are where we want to be. Our deliverables from the past year are on the following pages, broken down by Outcome, Output, What We Did, What We Learned, and What We’re Doing Next.

DELIVERABLES OCT. 2017 - SEPT. 2018				
Outcome	Output	What We Did	What We Learned — and Why That's Important	What We're Doing Next
Make editing metadata easier (and multilingual)	Multi-Content Revisions	Implemented MCR revision retrieval interface, MCR page update interface, MCR content metadata storage service. Tested and deployed MCR to Commons	This was complicated – a lot more complicated than we originally realized. It took more time, but in the end, we made a more robust update to MediaWiki.	Done! (MCR may be useful for other kinds of products as well. We will explore potential opportunities.)
Make editing metadata easier (and multilingual)	Multilingual file captions	Design, prototype, and alpha of adding, viewing, and editing file captions on the file page and in the upload wizard. Commons community consultations GLAM community consultations	Feedback on these designs confirmed that multilinguality is the most important benefit of SDC to most Commons users. We will continue to design and build with this goal paramount.	On Commons in Nov.



DELIVERABLES OCT. 2017 - SEPT. 2018				
Outcome	Output	What We Did	What We Learned — and Why That's Important	What We're Doing Next
Make editing metadata easier (and multilingual)	Depicts and other statements	<p>Research on Commons curators</p> <p>Basic design of displaying, adding, and editing "depicts" statements and other statements on the file page and in the upload wizard.</p> <p>Prototype of displaying, adding, and editing "depicts" and other statements on the file page.</p> <p>Design of the artwork scenario.</p> <p>Design of editing Wikidata statements from Commons.</p> <p>Community consult and feedback.</p> <p>Ontology Discussions.</p>	<p>It's hard to get people to build the ontology without concrete interface. Will push harder in December when "depicts" and other statements are on Beta Commons.</p>	<p>On Beta Commons in Dec., on Commons in Jan. or Feb.</p> <p>Incremental improvement of depicts statement editing features after initial release.</p>

DELIVERABLES OCT. 2017 - SEPT. 2018				
Outcome	Output	What We Did	What We Learned — and Why That's Important	What We're Doing Next
Make search more effective	<p>Search returns results with search term in file caption Search for media by depicts</p> <p>Search backend architecture</p>	<p>File captions indexed into search algorithm, works on alpha Designs and alpha of search by depicts feature</p> <p>Built ability to add a qualifier to a search term (e.g., " 'cat' that is 'in foreground', is 'black'); added a new data type for qualifiers, quantity, so now you can also search for " 'two cats' ", works on alpha</p> <p>Designed and implemented a method for extending CirrusSearch from MediaWiki extensions to facilitate storing and filtering search results based on structured data relationships.</p>	<p>Not much – sometimes things are useful but not illuminating</p> <p>Representing structured data relationships more than one step from the source in the search infrastructure requires both a strong understanding of the use cases involved, and specialized extraction routines for those use cases. Usage of structured data is only going to increase and moving forward we will need to think on how we can best expose this structured data in search.</p>	<p>On Beta Commons in Nov. On Beta Commons in Dec. Search for media by other types of statements on Beta Commons in Dec.</p>

DELIVERABLES OCT. 2017 - SEPT. 2018				
Outcome	Output	What We Did	What We Learned — and Why That's Important	What We're Doing Next
Satisfy GLAM use case	<p>Research</p> <p>List of Pain</p> <p>Pilot projects</p>	<p>Conducted generative user research of GLAM Commons users and their workflows</p> <p>Transformed research, expert advice, and community feedback into a prioritized backlog of tasks to alleviate pain points in GLAM workflows</p> <p>Ran preparatory workshops, identified projects, scoped work, explored timelines</p>	<p>We learned so much here, outlined in the research report, which informed what we would do and how</p>	<p>Research is done</p> <p>Prioritization is done but subject to change if new information emerges</p> <p>Confirm timelines, implement projects</p>
Allow license compliant reuse	Scoping structured licenses	Facilitating conversations around license ontology, display, editing, and reuse	<p>There are thousands of corner cases in license types, so it was difficult for this conversation to come to a consensus.</p> <p>We hope that having a concrete beta to play in will inspire people to tell us more about their core needs here.</p>	More community feedback on license needs, design, prototype, and build structured license feature.

DELIVERABLES OCT. 2017 - SEPT. 2018				
Outcome	Output	What We Did	What We Learned — and Why That’s Important	What We’re Doing Next
Support tool development	<p>Transitioning key tools</p> <p>Supporting development of new tools</p>	<p>Contacting key tool developers Gathering requirements for API and components from tool developers</p> <p>Scoped work with movement-aligned institutions to build additional structured data tooling</p>	<p>We had to go to the developers – on git, at Hackathons and conferences. Reaching out personally was very important here. These opportunities were a lovely surprise for us. It was only because our roadmap was sufficiently flexible that we were able to integrate them.</p>	<p>Support first tool transitions and new tool building in November.</p>

## How We’re Measuring Success: Deliverables Oct. 2018 to Sept. 2019

Our deliverables for the next year are on the following pages, broken down by Outcome, Output, What We’ll Do.

DELIVERABLES OCT. 2018 - SEPT. 2019		
Outcome	Output	What We'll Do
Allow license compliant reuse	Building structured license display, adding, editing, and search. Build other reuse/ embedding features.	Since structured licenses will probably heavily rely on default statement functionality, this is most likely a task about facilitating community decisions about the exact structure of licensees (e.g., which things are qualifiers, do we need “make primary” rules in the case of multiple licenses, etc.). There may be a little bit of dev work required to implement some front-end features based on user input.
Make editing metadata easier (and multilingual)	Image Annotations	The Image Annotation extension allows users to draw a rectangular region on an image and add descriptive metadata. An update to this system would include general workflow/UI improvements, integration with IIIF “relative position with image” spec, and possibly depicts statements, so we can define exactly what region of an image contains a depicted element.
Make editing metadata easier (and multilingual)	Machine-assisted depicts tagging	Using off-the-shelf platforms (either commercial or open source), implement a computer vision system that would suggest potential depicts tags based on the content of the image. A.I. isn't perfect, so we would mitigate errors by initially making this suggestion only, and perhaps limiting the suggestions to certain topics (objects for instance, instead of types of people)

DELIVERABLES OCT. 2018 - SEPT. 2019		
Outcome	Output	What We'll Do
Satisfy GLAM use case	Metadata download Metrics tools	Metadata download would involve at the very least allowing GLAMs to access raw JSON for download and import (perhaps for multiple files at once). Metrics tools could focus on providing usage and curation data either on-page, via API, or both (TBD).
Support tool development	Evangelizing Batch editing Campaigns	In these cases we will make sure the volunteer developer community knows the API tools available and some potential workflows they could enable. For batch editing tools, we hope to provide UI design for devs to start from.

## How We're Measuring Success: Targets and Measurement Methods

We've also developed targets and measurement methods based on the outcomes outlined in the grant proposal:

Outcome 1	Target 1	Measurement Method 1
<p>Commons contributors can edit metadata more easily, in any MediaWiki-supported language</p>	<ol style="list-style-type: none"> <li>1. Commonists add structured metadata to 5 million media files.</li> <li>2. Commonists add metadata to media files 2 months after the media is uploaded 10% more frequently.</li> <li>3. Commonists add metadata in structured fields in non-English languages to 1% of Commons media files.</li> <li>4. Reduce the percentage of Commons contributors reporting editing is 'very' or 'somewhat' difficult by 20%</li> </ol>	<ol style="list-style-type: none"> <li>1. Baseline/quarterly comparison of metadata on files with a common template, such as the information template.</li> <li>2. Baseline/quarterly comparison of when metadata on file pages are edited.</li> <li>3. Quarterly measurement media containing structured fields using non-English languages.</li> <li>4. Baseline/endline surveys of Commons contributors.</li> </ol>

Outcome 2	Target 2	Measurement Method 2
<p>People searching for media on Commons can find more relevant results faster</p>	<ol style="list-style-type: none"> <li>1. People searching for media on Commons get more relevant results with less time and effort.</li> <li>2. Default Commons search is more usable by searchers unfamiliar with Wikimedia projects after search using “depicts” feature is released.</li> <li>3. People searching for media on Commons can find relevant results in 10 example queries.</li> </ol>	<ol style="list-style-type: none"> <li>1.1 Click-through rate</li> <li>1.2 Clicks or scrolls to more results</li> <li>2. Usability testing</li> <li>3. Test search using the example queries</li> </ol>



Outcome 3	Target 3	Measurement Method 3
<p>GLAM institutions have more reliable and less challenging workflows for sharing media files and metadata on Wikimedia Commons as part of content partnerships, allowing for increased GLAM participation.</p>	<p>1. Over half of participants in identified pilot projects describe batch uploads to Commons as having a reasonably low barrier of entry. Note: We would like to be able to measure the number institutions and community members involved in GLAM batch uploads to Wikimedia Commons, and the number of files uploaded as part of these partnerships as a way of evaluating support of these kinds of projects (if it's easier, then we should see more). However, the current infrastructure does not allow us to establish a baseline. During this year we will work with community members to implement a process for community adoption of a practice using structured data to track this.</p>	<p>1. Exit interview with GLAM participants of the pilot case studies.</p>

Outcome 4	Target 4	Measurement Method 4
Reuse/Embedding	Since we do not yet know the form reuse and embedding tools will take, we don't yet have a target	The measurement method will follow the target.

Outcome 5	Target 5	Measurement Method 5
Developers who create and maintain key Commons tools, including volunteer developers, can transition those tools to work with structured data on Commons.	Tool developers transition key Commons tools related to metadata editing.	Identify which tools are most used and/or most essential, then track when they are transitioned.

# What's Up Next:

## Rest of Grant

Now that the major infrastructure is ready, we can deploy the major user-facing feature sets outlined in our roadmap. Although it took longer to build the infrastructure than we expected, we still plan to release the originally planned features, and features that we identified through generative research as key to the GLAM use case.

Originally planned features we will do next:

- Structured licenses
- Reuse and embedding

New features we will also do:

- Image annotations
- Machine-learning-assisted depicts statements
- Metrics tools
- Batch editing

There are some features we originally planned to do, but we think are less important than the newly identified features. We'll still do them if we have time:

- Related images: This will be an important discovery path, but we need more structured data before it is useful. Thus, we are prioritizing features to add more structured data, more easily.
- Updated MediaViewer: This will be useful for other wiki projects, but again, we think the other planned features to use structured data are sufficient for now, and we should build more features to create structured data first.

Also, the Wikimedia Foundation's [Android team](#) wants to develop mobile micro-contribution features, and structured file caption translations are planned to be one of the first ones, released before June 2019. We think it will encourage multilinguality through translations. While not part of the grant, this is work the grant has enabled through the tools we're building.

Finally, in engaging people with the new structured data features, we will also encourage and mentor community and GLAM adoption of structured data on Commons in 2019. We are planning to actively support 5 to 8 GLAM-Wiki pilot projects to test the initial implementation of structured data. We are selecting a deliberately heterogeneous set of projects: they will be diverse in terms of content (from digitized publications to historical maps, photographs and images of artworks), partnerships (both community-driven projects and collaborations with smaller and larger GLAMs) and in approaches (bulk uploads to Wikimedia Commons, crowdsourcing campaigns, and tool development).

The organizers of these pilot projects will be actively mentored by Sandra Fauconnier. The experiences and lessons gained from these projects will form the basis for renewed documentation for GLAM contributions to Wikimedia Commons. Wikimedia Sweden has agreed to assist in this work as part of their [FindingGLAMs](#) project, in which part of the planned case studies for GLAM collaborations will also focus on structured data on Wikimedia Commons.

We are also exploring partnering with community organizers to initiate structured data creation/migration/translation campaigns.

## Beyond 2020

Beyond 2020, we are thinking of the day when millions of people are using Commons as much more than a repository. We're thinking of the day when they are using Commons and Wikidata as the central place on the Internet to share their collections and educational media. The "they" are GLAMs – museums, university libraries, archives, and other institutions. The "they" are also other knowledge stakeholders, such as grassroots and community-based projects working to document living knowledge and underrepresented heritage. The "they" are everyday Commons users we spoke with in the past year, and who sometimes refer to themselves as "Commonists." And the "they" are people who aren't regular users of Commons but who could be. They may not even know about Commons. At least not yet.

Structured Data on Commons is a game-changer. It's disrupting Commons, and sets up Commons – and sets up the Wikimedia movement – for bigger things. We want to build upon the base of user value we've created. We want to make Commons and Wikidata essential infrastructure of the GLAM ecosystem and build out our GLAM contributor community. We want to elaborate on the reuse and embedding features we build for other knowledge services so that more free knowledge emanates from the Wikimedia movement. We want to make sure marginalized languages and knowledge find a home on Commons and become part of the free media and metadata movement. We want people from these marginalized communities to share rich representations of knowledge, and take advantage of – and link to – the vast body of knowledge already amassed in the Wikimedia projects. Structured Data on Commons will make free media and metadata more accessible, more editable, and more shareable. We want our work to become the catalyst for positive, systemic change within the free knowledge ecosystem. We have new momentum. And we have a clearer vision of what we can do within the wider Wikimedia movement.

# Appendix:

## Staff working on Structured Data on Commons

Project leader

- [Amanda Bittaker](#), Program Manager

Wikidata and Wikibase:

- [Lydia Pintscher](#), Product Manager for Wikidata
- [Adam Shorland](#), Software Developer for Wikidata
- [Léa Lacroix](#), Community Liaison for Wikidata

Multimedia:

- [Ramsey Isler](#), Product Manager
- [Mark Holmquist](#), Lead Software Engineer
- [Cormac Parle](#), Senior Software Engineer
- [Matthias Mullie](#), Software Engineer

Community Programs:

- [Ben Vershbow](#), Director, Community Programs (GLAM/Education)
- [Sandra Fauconnier](#), Program Officer for GLAM and Structured Data, GLAM engagement
- [Alex Stinson](#), GLAM Strategist

Community Engagement:

- [Keegan Peterzell](#), Community Relations Specialist

Design:

- [Nirzar Pangarkar](#), Design Manager
- [Jonathan Morgan](#), Senior Design Researcher
- [Pamela Drouin](#), User Experience Designer
- Niharika Ved, Design Research Intern

Search Platform:

- [Deborah Tankersley](#), Product Manager
- [Erika Bjune](#), Engineering Manager

- [Stas Malyshev](#), Senior Performance Engineer
- [Erik Bernhardson](#), Senior Software Engineer

MediaWiki Platform:

- [Daniel Kinzler](#), Principal Platform Engineer for Wikidata
- [Brad Jorsch](#), Senior Software Engineer

Audiences Engineering (at large):

- [Adam Baso](#), Director, Engineering
- [Brion Vibber](#), Lead Software Architect
- [Gergő Tisza](#), Senior Software Engineer (Infrastructure)
- Piotr Miazga, Software Engineer (Web)

## Website links: Program and project pages

A summary of web links for program and project pages:

- The project's main page: [https://commons.wikimedia.org/wiki/Commons:Structured\\_data](https://commons.wikimedia.org/wiki/Commons:Structured_data)
- Overview of Phabricator tasks for Multi-Content Revisions: <https://phabricator.wikimedia.org/tag/multi-content-revisions/>
- Community feedback on our design prototype for searching Commons: [https://commons.wikimedia.org/wiki/Commons\\_talk:Structured\\_data/Get\\_involved/Feedback\\_requests/Search\\_prototype](https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Get_involved/Feedback_requests/Search_prototype) Phabricator log of design and technical changes for multilingual file captions: <https://phabricator.wikimedia.org/T201456>
- Research on supporting GLAM institutions' Commons contributions [https://meta.wikimedia.org/wiki/Research:Supporting\\_Commons\\_contribution\\_by\\_GLAM\\_institutions](https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions)
- GLAM Focus group that is advising and providing regular feedback: [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/GLAM\\_focus\\_group](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/GLAM_focus_group)
- Community Focus group that is advising and providing regular feedback: [https://meta.wikimedia.org/wiki/Global\\_message\\_delivery/Targets/Structured\\_Commons\\_focus\\_group](https://meta.wikimedia.org/wiki/Global_message_delivery/Targets/Structured_Commons_focus_group) See also: [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved/Community\\_focus\\_group](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Community_focus_group) Volunteer group that is translating content related to Structured Data on Commons: [https://meta.wikimedia.org/wiki/Structured\\_Data\\_on\\_Commons/Newsletter/Translators](https://meta.wikimedia.org/wiki/Structured_Data_on_Commons/Newsletter/Translators)

- Archive of Structured Data on Commons newsletters: [https://meta.wikimedia.org/wiki/Structured\\_Data\\_on\\_Commons/Newsletter](https://meta.wikimedia.org/wiki/Structured_Data_on_Commons/Newsletter)
- How Commons users can get involved in the project: [https://commons.wikimedia.org/wiki/Commons:Structured\\_data/Get\\_involved](https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved)
- A link to last year's Structured Data on Commons report: [https://upload.wikimedia.org/wikipedia/commons/a/ad/Structured\\_Data\\_on\\_Commons\\_Nov.\\_2017\\_Report\\_to\\_Alfred\\_P.\\_Sloan\\_Foundation\\_-\\_From\\_the\\_Wikimedia\\_Foundation.pdf](https://upload.wikimedia.org/wikipedia/commons/a/ad/Structured_Data_on_Commons_Nov._2017_Report_to_Alfred_P._Sloan_Foundation_-_From_the_Wikimedia_Foundation.pdf)

## Glossary of Structured Data on Commons terms

### Three main types of structured data that can be added to Commons media files:

- **Depicts:** Description of what's depicted. Depicts statements on Commons are structured data elements that are essentially the same as [Wikidata depicts statements](#) but with a few additional user-facing functionalities on Commons. Depicts statements on Commons can be thought of as a special kind of topical "tag" for multimedia, similar to how tags work on sites like Flickr and WordPress.
- **Other statements:** These key-value pair structured data elements are also basically the same as the [statements that exist on Wikidata](#). Examples: P462=Q1088 (color=blue), P2048= 60 inches (height=60 inches)
- **Multilingual file captions:** These are plain text strings that are essentially stored like [Wikidata labels](#) behind the scenes, but on Commons they have a different UI.

### Other terms:

- **Properties** (identified with a unique ID number in the form of PXXX). A property describes the data type of a statement and can be thought of as a category of data – for example "color" for the data value "blue." Properties, when paired with values, form a statement [in Wikidata](#).
- **Items:** [In Wikidata](#), a concept, topic, or object is represented by an item – identified with a unique ID number in the form of QXXX. Example: the planet Earth on Wikidata is Q2.
- **Statements:** [Similar to Wikidata](#), a statement is how the information we know about an item – the data we have about it – gets recorded in structured data. This happens by pairing a property with at least one value. Color=blue is an example of a statement, but in structured data it would be P462=Q1088.
- **Qualifiers:** Qualifiers are properties used to further describe or refine the value of a property given in a statement. Example: Berlin (Q64) can have a property called population, and for that



population property there's a qualifier of "point in time" that clarifies when that population number was measured/estimated.

- **Entity:** The content of a Wikidata page, which may be either an "item" (in the main namespace) or a "property" (in the property namespace). Every entity is uniquely identified by an entity ID, which is a prefixed number, for example starting with the prefix Q for an item and P for a property. <https://www.wikidata.org/wiki/Wikidata:Glossary>