



Structured Data on Commons: A Nine-Month Report for the Sloan Foundation

Prepared by Caitlin Virtue (cvirtue@wikimedia.org), Amanda Bittaker (abittaker@wikimedia.org) and Jonathan Curiel

November 1, 2017

Table of Contents

Executive Summary	1
Who's Who on the Project	4
Amanda Bittaker, Program Manager	4
Sandra Fauconnier, Community Liaison and GLAM Liaison	5
Nirzar Pangarkar, Design Manager	6
Jonathan Morgan, Senior Design Researcher	6
Pam Drouin, UX Designer	7
Daniel Kinzler, Technical Lead at Wikimedia Deutschland	7
Lydia Pintscher, Product Manager for Wikidata	8
Ramsey Isler, Multimedia Product Manager	9
Alex Stinson, GLAM-Wiki Strategist	10
Planning Structured Data on Commons: Creating a Stellar Action Plan	12
Software Engineering: Laying the Groundwork for Structured Data on Commons	16
Test Federation:	16
Multi-Content Revisions:	20
Community Input: The Work of Jarek Tuszynski	23
Nine Months in Review: A Timeline of Key Developments	25
Research and Interviews: Finding the Sweet Spot for Structured Data on Commons	28
Research-Driven Design: How Structured Data on Commons Is Coming to Life	31
Researching Usage: How to Measure the Effectiveness of Structured Data on Commons	39
Deliverables: What We Achieved in Engineering and Other Project Work	41
Where People Are Learning About — and Discussing — Structured Data on Commons	50
Appendixes	51

Executive Summary

On the morning of Saturday, August 12, as Sandra Fauconnier was setting up for her talk at Wikimania – Wikimedia’s biggest annual conference of the year – Commons users, Wikipedia contributors, software engineers, and others began crowding into her Montreal conference room, eager to hear what she would say. Sandra’s talk, titled “Structured Data on Wikimedia Commons: What’s Coming,” would offer a preview of what is Commons’ biggest change in its 13-year history, and seats in the room quickly filled up. Many people stood in the back, or sat in the aisles. And when Sandra – the project’s Community Liaison, and a veteran Wikipedia and Commons contributor who first edited in 2003 – finally began, she detailed the work and the progress to come.

“The software that powers Wikidata, called Wikibase, will be integrated into Wikimedia Commons,” she said.¹ “We will have better search functionalities, and APIs . . . so external parties can do large-scale reuse of images on Commons.”

Sandra also said this: “It has been fun so far.”²

It’s been fun because we are rebuilding, repurposing, and reimagining what Commons can be. What is already the world’s largest repository of freely licensed educational media will become much more useful to Commons contributors, Wikipedia editors, casual users, and major institutions – including museums, archives, libraries, and universities. The transition is centering on a straightforward change: transform the information throughout Commons’ media files from free text into machine-readable data.

The past nine months have laid the foundation for Commons’ dramatic evolution because we hired key project staff members like Sandra, because we created essential software that lets users do very early preview testing of the project, and because we’ve done outreach – from in-person talks at Wikimania to online announcements to Wikimedia Foundation web pages that outline the project and invite feedback at any time. We are conducting in-depth interviews with Commons users –

¹ See 35:00 of Sandra’s Wikimania talk on YouTube: <https://www.youtube.com/watch?v=OwM9u6BTveA>. Sandra’s talk begins at exactly 32:00 of the video.

² See 33:04 of Sandra’s Wikimania talk on YouTube: <https://www.youtube.com/watch?v=OwM9u6BTveA>

both individual users and representatives of large institutions – to get a precise and proactive understanding of their needs and their ideas. From the beginning, Structured Data on Commons has been a collaborative project that lets our users participate in planning Commons’ future.

Among the highlights from the past nine months:

- From July to September 2017, Jonathan Morgan, the Wikimedia Foundation’s Senior Design Researcher, conducted 11 detailed Structured Data on Commons interviews with key members from GLAM institutions and the Wikipedia community.
- Since January, we made good progress on Multi-Content Revisions, which is an engineering prerequisite for building Structured Data on Commons.
- On July 6, 2017, we announced that engineers had developed the first functional version of “federation” for testing by the community. It allows users to test an important building block for Structured Data on Commons. They can test how they will use Wikidata’s items and properties to describe media files on Commons in the future.
- Around the same time, we made several key hires, including Sandra as the project’s Community Liaison, Amanda Bittaker as the project’s Program Manager, and Ramsey Isler as the new product manager on the multimedia team, whose primary focus will be on implementing Structured Data on Commons.
- With Sandra’s talk and other talks at Wikimania 2017, we introduced Structured Data on Commons in a very public and personal way – which set a good tone for a project where Wikimedians’ input is critical to the project’s success.

Thank you. The Alfred P. Sloan Foundation is funding the Structured Data on Commons project, and your support has helped spur the beginnings of Commons’ new beginning. The first nine months have created a momentum that’s reflected in the work we are doing every day on the project. One of the most important developments: Community members are already contributing to the project. Even in our initial stage, people are subscribing to updates, commenting on talk pages, testing pages, contributing ideas, and influencing how we are going forward. As expected, we are getting lots of questions as well as constructive feedback from some users – and that’s great. The questions and feedback mean people are engaged with the project, and want to shape its future.

The active collaboration that’s happening between the community and the project mirrors the active collaboration that’s happening between the Wikimedia Foundation and Wikimedia Deutschland (Germany), whose semantic web database experts are doing crucial engineering work

that is integrating the software behind Wikidata into Commons. Hiring the right people – people with expertise in Commons, Wikipedia, data projects, and stewarding ideas from stage to stage – and grouping them with like-minded staff members was a key milestone of the first six months.

The hiring spilled over into the beginning of the 2017-2018 fiscal year – and it has been the last three months, after the team was more completely assembled, that we moved beyond just essential infrastructure building. We’ve had more conversations with the community, and introduced the work more thoroughly. We’ve begun more extensive interviews with GLAM institutions. We’ve made concrete outlines of the project that clarified to us – and to the community – the best ways to move forward. The symbiosis between the first six months and the last three is clear: The hiring and early work, while slower than we’d envisioned, set up the bursts of work that are happening now. While project expenses were much lower than expected in the 2016-2017 fiscal year, it was money well spent as we assembled the team that is now functioning at a very high level. And staff will be working on the project more than originally budgeted in the 2017-2018 fiscal year, which will expedite our work and use the grant’s funding allocation more expansively.

Each staff member is a public face of the project. Each staff member is bringing a personal touch to the project. At her Wikimania talk, Sandra veered between being funny and serious. She made people laugh while also demonstrating her deep community knowledge. She put them at ease.

“I started out as a volunteer” on Commons, Sandra told the audience.³ “We have now reached 40 million, 900,000 freely usable files, which is kind of amazing. I think I contributed a thousand of those.”

In the next year, Commons may have more media files than Wikipedia has articles. The time is right for Structured Data on Commons, and we are initiating a new vision of Commons that will change the way people search for those millions of images and change the way people use those millions of images. The progress we’ve made so far – including staff hirings, software developments, and engagement with the Wikimedia community – is just the start of our much bigger plan for Structured Data on Commons.

³ See <https://www.youtube.com/watch?v=OwM9u6BTveA>, 35:10 to 35:26

Who's Who on the Project

The staff members who are working on Structured Data on Commons bring years of expertise to their positions, and have a history of achievement in working on Wikipedia, Wikimedia Commons, Wikidata, and other projects. Collectively, they are bringing veteran experience and innovative outlooks to Structured Data on Commons. By collaborating with engineers at Wikimedia Deutschland (Germany) who have years of experience with Wikidata, the Wikimedia Foundation is ensuring that the project has an engineering depth that is foundational and transformative. The team⁴ is led by Amanda Bittaker, who is Program Manager, and is made up of product managers, designers, engineers, researchers, GLAM experts and community engagement experts from across four departments in two organizations, including Design, Multimedia, Community Programs, Search Platform, Research, and MediaWiki Platform, and Wikidata.

Amanda Bittaker, Program Manager

Amanda, who has been with the Wikimedia Foundation since 2014, was named the Structured Data on Commons program manager in June, chosen for her expertise in nonprofit program management and her breadth of global experience, evaluation design, and project implementation. When Amanda joined the Wikimedia Foundation, she worked initially with Learning and Evaluation⁵ to help the foundation's program organizers and our communities design, manage, and evaluate their programs. During that time, she also partnered with engineers to build program tools such as the Program and Events Dashboard⁶, which helps manage online and offline programs (including editathons and workshops) and works in any language or Wikimedia project.



[Photo of Amanda](#) by Myleen Hollero, CC BY-SA 3.0

⁴ https://commons.wikimedia.org/wiki/Commons:Structured_data/Development/Team

⁵ https://meta.wikimedia.org/wiki/Learning_and_Evaluation

⁶ <https://blog.wikimedia.org/2016/12/02/programs-events-dashboard/>

Before joining the Foundation, Amanda worked for five years in program design, monitoring and evaluation, and finance in the international development industry, which included a position as research advisor with Mercy Corps in Medellin, Colombia, and as monitoring and evaluation coordinator with Mercy Corps in La Paz, Bolivia. She was also a finance officer with the organization in Portland, Oregon. In 2009 and 2010, she was Director of Research with the Rubicon Global Group. Amanda studied economics, statistics, mathematics, and political science at Reed College.

Amanda's Wikimedia profile: [https://wikimediafoundation.org/wiki/User:Abittaker_\(WMF\)](https://wikimediafoundation.org/wiki/User:Abittaker_(WMF))

Amanda's LinkedIn profile: <https://www.linkedin.com/in/amanda-bittaker-8b9a8010/>

Sandra Fauconnier, Community Liaison and GLAM Liaison

A longtime Wikipedia and Commons editor who is also a Wikidata contributor, Sandra joined the Wikimedia Foundation in July 2017 and is supporting collaboration between the Wikimedia communities (Commons, Wikidata, GLAM, etc.) and the product development teams at the Wikimedia Foundation and Wikimedia Deutschland. Sandra is also driving the engagement with new individual content contributors, existing and new GLAM organizations, and developers interested in exploring the possibilities of the new platform.

Sandra did her first edits on English Wikipedia in 2003 – just two years after Wikipedia began. After a hiatus to start a family, she became an active Commons editor, was a board member and volunteer for Wikimedia Nederland, and has been enthusiastically contributing to Wikidata for several years. Sandra has a Master of Arts in art history and has worked on a variety of Internet and video projects in the Dutch cultural sector. Twice, she has been a [Wikipedian in Residence](#) – a position reserved for veteran Wikipedia editors who work with institutions to facilitate Wikipedia entries related to that institution's mission, who encourage and assist it to release material under open licenses, and who develop the relationship between the institution and the Wikimedia community. Based in Rotterdam, Netherlands, Sandra speaks English, Flemish, Dutch and slow French, and she knows basic German, Spanish and Italian.



[Photo of Sandra](#) by Victor Grigas, CC BY-SA 3.0

Sandra's Wikimedia profile: [https://meta.wikimedia.org/wiki/User:SFauconnier_\(WMF\)](https://meta.wikimedia.org/wiki/User:SFauconnier_(WMF))

Sandra's LinkedIn profile: <https://www.linkedin.com/in/sandrafauconnier/>

Sandra on Twitter: <https://twitter.com/sanseveria?lang=en>

Nirzar Pangarkar, Design Manager

Nirzar has been the Wikimedia Foundation's Design Manager since December 2015, joining the foundation a year earlier as a UX Designer. As Design Manager, Nirzar oversees a team of designers⁷ who are regularly improving the user interfaces of our websites and apps through research, community outreach, and design orchestration. Before joining the Wikimedia Foundation, Nirzar was a Senior Product Designer with Cleartrip, an online travel company in Bangalore, India, where he led the company's mobile design team in its products for iOS, Android and Windows. Nirzar has a Bachelor of Computer Science from Fergusson College in Pune, India.



[Photo of Nirzar](#) by Myleen Hollero, CC BY-SA 3.0

Nirzar's Wikimedia profile: [https://www.mediawiki.org/wiki/User:Npangarkar_\(WMF\)](https://www.mediawiki.org/wiki/User:Npangarkar_(WMF))

Nirzar's LinkedIn profile: <https://www.linkedin.com/in/nirzar/>

Jonathan Morgan, Senior Design Researcher

Jonathan has been with the Wikimedia Foundation since 2011. As a Wikimedia Research Fellow, he performed research to understand the new editor experience of Wikipedia, and helped design and launch the Wikipedia Teahouse ([WP:Teahouse](#)), a new-user peer-support hub on English Wikipedia. Since then, he has served as the Research Lead in the Community Engagement department, and since 2015 as the Senior Design Researcher.

Jonathan has a PhD in Human Centered Design & Engineering from the University of Washington, and a BA in philosophy and classics from St. John's College. His doctoral thesis, "Coordinating the Commons: Diversity & Dynamics in Open Collaboration,"⁸ examined the role of openness, and the potential antecedents and consequences of formalization, within Wikipedia through an analysis of three distinct but interrelated social structures: community-created rules within the Wikipedia policy environment; coordination work and group dynamics within self-organized open teams (WikiProjects); and the socialization mechanisms that Wikipedia editors use to teach new community

⁷ <https://www.mediawiki.org/wiki/Design>

⁸ <https://digital.lib.washington.edu/researchworks/handle/1773/25074>

members how to participate. He has edited Wikipedia since 2008.

Jonathan's Wikimedia page: <https://wikimediafoundation.org/wiki/User:Jtmorgan>

Jonathan's website: <http://jtmorgan.net/>

Jonathan's Google Scholar profile: <https://scholar.google.com/citations?user=gDRyMhcAAAAJ>

Jonathan's LinkedIn Page: <https://www.linkedin.com/in/jonathan-t-morgan-4019058/>



Photo of Jonathan by Matthew Roth, CC BY-SA 3.0

Pam Drouin, UX Designer

Pamela, who joined the Wikimedia Foundation as a UX Designer in September 2017, has a background in design and library science that gives her a unique insight into the ways that users access and engage with information. Before Wikimedia, she worked as a UX designer at Bocoup, a web platform consulting firm in Boston, where she focused on user research, information architecture, and interaction design. She also played an integral role in the creation of Bocoup's Open Design Kit,⁹ open-source design software that anyone can use to make better online designs. Before that, she was a technical product manager at ProQuest, the long-time information-content company, and was a UX Researcher and Content Specialist at the University of Washington. She has a Masters in Library Science from the University of South Florida, and a Bachelor of Arts in English Literature from Florida State University.



Photo by Pam Drouin

Pamela's LinkedIn profile: <https://www.linkedin.com/in/drouinp/>

Daniel Kinzler, Technical Lead at Wikimedia Deutschland

Daniel is the technical lead at Wikimedia Deutschland, and is overseeing the integration of Wikidata into Structured Data on Commons. Daniel joined Wikimedia Deutschland in 2008, then became the organization's Lead Engineer on Wikidata in 2012 before he was promoted to Principal Platform Engineer in January 2017. Fluent in English and German, Daniel has a diploma in Informatics and Logics from the University of Leipzig, where his thesis was about the extraction of a multilin-

⁹ <http://opendesignkit.org/>

qual thesaurus from Wikipedia. He also has a pre-diploma in informatics from the University of Stuttgart.

Daniel is an expert in integrating MediaWiki with other services, and in using Wikipedia content for semantic integration and multilingual information retrieval. He has expertise in managing large data sets in MySQL, which is an Open Source data-management system, and in building high performance LAMP services, which revolve around four Open Source components. Daniel is a member of Wikimedia Technical Committee¹⁰, which acts as an extension of the Wikimedia Foundation's Chief Technology Officer and oversees the integrity, consistency, stability and performance of the software supporting the Wikimedia projects. He lives and works in Berlin.

Daniel's Wikimedia page: [https://meta.wikimedia.org/wiki/User:Daniel_Kinzler_\(WMDE\)](https://meta.wikimedia.org/wiki/User:Daniel_Kinzler_(WMDE))

Daniel's LinkedIn Profile: <https://www.linkedin.com/in/brightbyte/>



Photo of Daniel by Wikimedia Deutschland, CC BY-SA 4.0

Lydia Pintscher, Product Manager for Wikidata

Lydia has been the Product Manager for Wikidata since October of 2013, leading oversight of the knowledge base that is a central storage for the structured data of the Wikimedia Foundation's websites. Lydia's oversight includes managing the development of the project in an open and collaborative way, coordinating with community and other stakeholders, and managing communication and community interaction around Wikidata's development. Lydia is widely recognized in the Free Software community, with two regarded books to her credit:¹¹ *Open Advice*, a collection of essays by 42 prominent Free Software contributors; and *20 Years of KDE: Past, Present and Future*, which narrates the history of the KDE free software community¹² – one of world's biggest free software communities.

¹⁰ https://www.mediawiki.org/wiki/Wikimedia_Technical_Committee

¹¹ <http://www.lydiapintscher.de/books.php>

¹² <https://en.wikipedia.org/wiki/KDE>

In her spare time, she is the president of KDE e.V., a German nonprofit that supports the international KDE free software community. She has led the organization since 2014, and since 2008 has overseen the mentoring of more than 200 students every year in KDE software development. Lydia, who is based in Berlin, has a diploma in computer science from Germany's Karlsruhe Institute of Technology, where her focus was on robotics in medicine, language/speech and innovation.



[Photo of Lydia](#) by Victor Grigas, CC BY-SA 3.0

Lydia's Wikimedia profile: [https://meta.wikimedia.org/wiki/User:Lydia_Pintscher_\(WMDE\)](https://meta.wikimedia.org/wiki/User:Lydia_Pintscher_(WMDE))

Lydia's website: <http://www.lydiapintscher.de/about.php>

Lydia's LinkedIn profile: <https://www.linkedin.com/in/lydiapintscher/>

Ramsey Isler, Multimedia Product Manager

Ramsey, who has a two-decade-long background as a successful developer, designer, writer, and technical director, joined the Wikimedia Foundation in August 2017 as Product Manager for the multimedia team, whose now-primary focus is implementing Structured Data on Commons. Before joining Wikimedia, Ramsey was a product consultant who helped major Japanese companies (Wacom, Nikon, Pioneer Electronics, and others) build new web, app, and hardware products for Western audiences. Prior to joining the Product Manager ranks, he was an interactive developer and startup co-founder with experience in front-end development, the LAMP stack, and Flash for the web and mobile devices. Among his achievements: Ramsey conceptualized, designed, and led the development of GoTunes – a fitness app that combined jogging, Spotify streaming music, beat-per-minute algorithms, and GPS event-triggering features. Ramsey, who studied at the Rochester Institute of Technology and is now based in Los Angeles, has also been an analyst for the Department of Defense. He is also an author and screenwriter, and worked for years as a writer with IGN, the video game and entertainment media company. Among the many people he has interviewed: Leonard Nimoy and JJ Abrams.

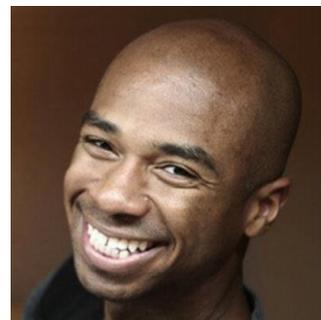


Photo by Ramsey Isler

Ramsey's Wikimedia Profile: [https://meta.wikimedia.org/wiki/User:RIsler_\(WMF\)](https://meta.wikimedia.org/wiki/User:RIsler_(WMF))

Ramsey's LinkedIn profile: <https://www.linkedin.com/in/ramsey-isler-a514943/>

Ramsey's website: <http://ramseyisler.com/>

Alex Stinson, GLAM-Wiki Strategist

As GLAM-Wiki Strategist since May 2016, Alex supports the Global GLAM-Wiki community, focusing on helping cultural heritage organizations (principally Galleries, Libraries, Archives and Museums) share, create content, and otherwise engage with Wikimedia's projects. Before this role, Alex was project manager with [The Wikipedia Library](#), focusing on developing relationships with publishers, research libraries and other cultural organizations to give editors more library resources to improve Wikipedia articles. Alex has a Masters degree in English Literature from Kansas State University, with research focused on cultural studies and the digital humanities. Prior to the Wikimedia Foundation, he worked as a digital humanist at Kansas State University, where he developed [en:Digital Humanities](#) projects in which he created partnerships with educators and cultural institutions. He has a bachelor's degree in history and English from James Madison University. As a volunteer Wikipedia editor ([en>User:Sadads](#)), he has made more than 100,000 edits.



Photo by Myleen Hollero, CC BY-SA 3.0

Alex's Wikimedia profile: [https://meta.wikimedia.org/wiki/User:Astinson_\(WMF\)](https://meta.wikimedia.org/wiki/User:Astinson_(WMF))

Alex's Wikipedia page: <https://en.wikipedia.org/wiki/User:Sadads>

Alex's LinkedIn profile: <https://www.linkedin.com/in/alex-stinson-bb0602b6/>

Other Staff Members Working on the Project:

Audiences:

Toby Negrin, Chief Product Officer

Wikidata and Wikibase:

Adam Shorland, Software Developer for Wikidata

Katie Filbert, Software Developer for Wikidata

Léa Lacroix, Community Liaison for Wikidata

Thiemo Mättig, Software Developer for Wikidata

Multimedia:

Adam Baso, Director, Engineering

Mark Holmquist, Lead Software Engineer
Cormac Parle, Senior Software Engineer
Matthias Mullie, Software Engineer

Community Programs:

Ben Vershbow, Lead Programs Manager, Community Programs

Design:

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:

Cindy Cicalese, Product Manager
Brion Vibber, Lead Software Architect
Brad Jorsch, Senior Software Engineer

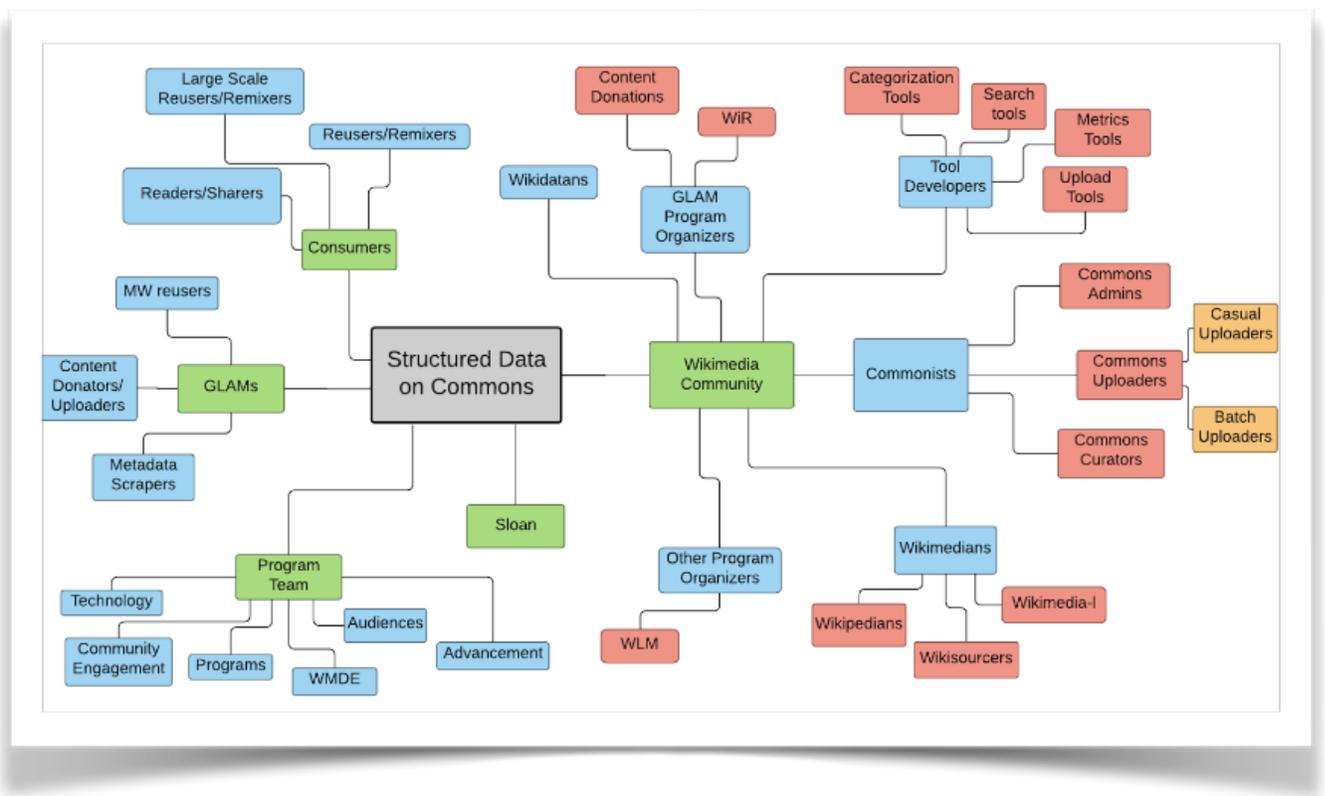


Photo by Joel Aufrecht

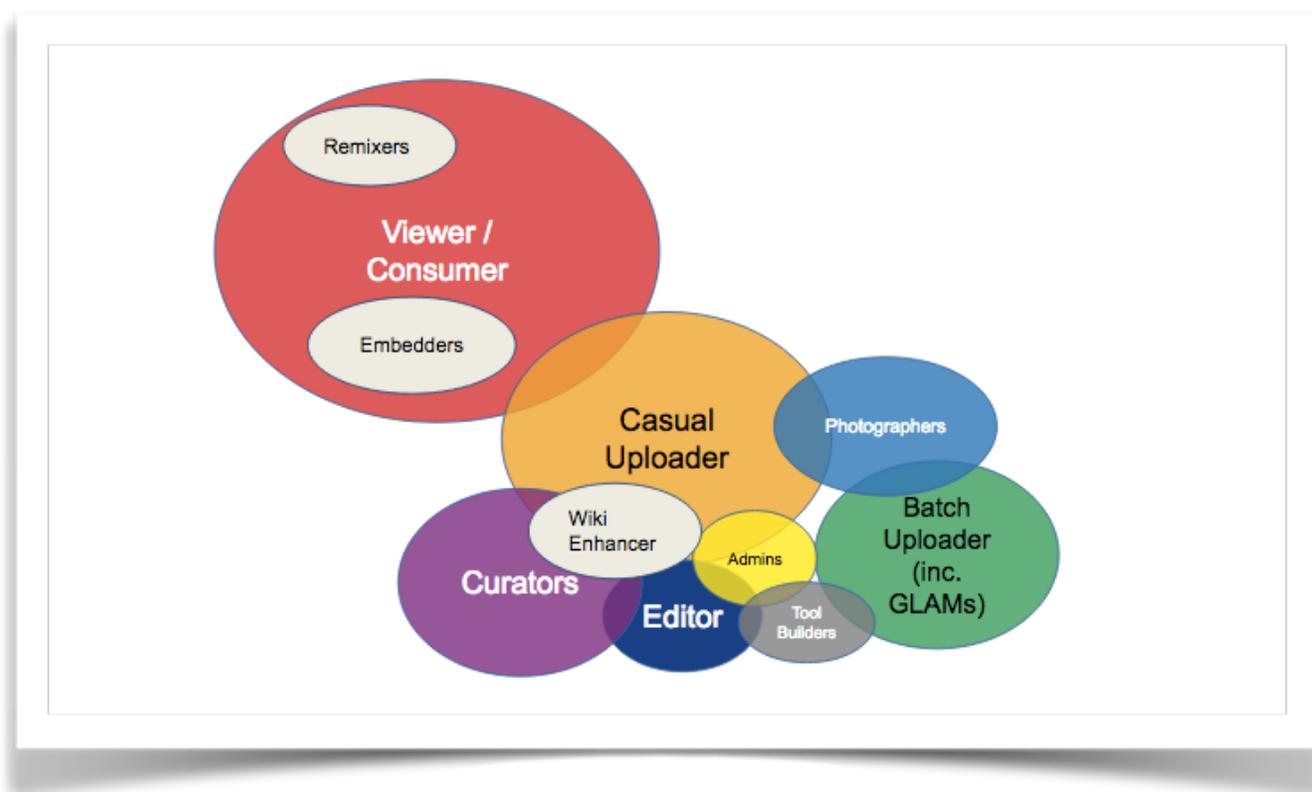
Photo above right: The Structured Data on Commons team during a project meet-up at Wikimania 2017

Planning Structured Data on Commons: Creating a Stellar Action Plan

Once we assembled the Structured Data on Commons team, the next stages included detailing our 2017-2018 roadmap, and articulating a “Stakeholders” chart that reflects exactly who the project is reaching. The Stakeholders map (see below) contextualized the project’s big picture – showing us the entire Wikimedia community and the interconnectedness among the community members.



From this Stakeholders map, we derived the different user types that are set to benefit from Structured Data on Commons. Ramsey Isler, the Multimedia team’s Product Manager, led the research that qualified the needs and interests of nine user types. Ramsey created detailed “user stories” that describe the users’ practices and motivations – and the ways that the Structured Data on Commons project can aid their use of Commons. Each user type has distinct needs and workflows:



Viewers: They mainly visit Commons to find particular free files. They most likely never edit. Within this type are **Remixers** (people who will download, edit, and remix the images in their own new art) and **Embedders** (bloggers, reporters, etc. who embed unaltered work on other pages).

Casual Uploader: They actively (on average at least once every month) upload one image at a time (which may or may not be their own). These are amateurs who probably take most photos with their phones.

Batch Uploader: They upload 20 or more images at a time using a batch upload tool. They are typically someone associated with a GLAM project, but not always.

Wikimedia enhancers: They are users on various wiki projects who search for images to use on those projects (Wikipedia, WikiVoyage, etc).

Photographers: They are pros or semi-pros who actively upload their own images to Commons. They are generally using DSLR/mirrorless cameras but may use phones in a pinch.

Editors: They actively edit media information on Commons for the sake of accuracy, completeness, or maintaining site quality.

Curators: They actively categorize, group, and label images to make things organized and easy to find. They may also be involved in picking featured, quality, and valued images.

Tool Builders: They are volunteer developers who write and release software to supplement Commons functionality or fill in functional gaps.

Admins: They are users who have special abilities to enforce the rules and primarily act as site enforcers of copyright policies and social norms. They may or may not actively donate media.

From these nine distinct user types, we decided to do in-depth research on the needs and workflows of three major Commons groups that we don't know enough about yet: GLAM users, curators, and reusers – with GLAM users an initial priority.

We also laid out an engineering roadmap for the 2017-2018 fiscal year that evolved to center on five major areas:

Multi-Content Revisions: [Multi-Content Revisions](#) is an update to the core MediaWiki software (the software that all Wikimedia wikis run on) that will allow multiple types of content, such as wiki-text, JSON, and structured data to live on the same wiki page. With this, Commons users can more efficiently add, edit, and manage structured Wikibase metadata associated with Commons

media files. **Short-term goal:** By December 2017, test MCR implementation, and integrate MCR code with existing MediaWiki features.

MedialInfo Wikibase Extension: This is [an extension](#) to [Wikibase](#) that adds the new MedialInfo entity type for handling structured data for multimedia files. The extension connects with a file description page and adds a link to a MedialInfo page storing supplemental metadata about the file. The metadata can include the author, detailed license information, and concepts that an image actually depicts. **Short-term goal:** Deploy the first version by December 2017.

Metrics and Evaluation: [The Search Platform Team](#) is gathering statistical data and analysis to better understand usage patterns on Commons and areas where Structured Data might provide improvement. **Short-term goal:** By December 2017, the team will capture a snapshot of Commons usage now, such as the number of media used across projects, the types of uploads, and the breakdown by file types. The team will also identify behaviors that can help improve searching on Commons. This information will help us determine how effective search is now, and how to make it more effective.

Community Outreach and Feedback: We are reaching out to community members from every realm, including GLAM users, administrators, curators, and others. **Short-term goal:** By December 2017, build up and facilitate a focus group of Commons and Wikidata contributors for long-term feedback and input on Structured Data on Commons. And by December 2017, update our info portal that details the work we're doing.

Define and Finalize the Minimum Viable Product: With so many features to consider, we have defined the project's Minimum Viable Product and the first feature that users will employ in the project.

The work on Multi-Content Revisions has been ongoing from January 2017. As noted in the Executive Summary, Wikimedia Deutschland's semantic web database experts have been doing the engineering work that is integrating the software behind Wikidata into Commons. They were doing this work even before we hired Amanda Bittaker as Program Manager and before we assembled the full team that is now working together. Their engineering, which is detailed in the next section, has laid the early groundwork for Structured Data on Commons.

Software Engineering: Laying the Groundwork for Structured Data on Commons

Software engineering is one of the project's main pillars, and since January we've focused on two areas:

1. We've tested Wikibase "federation," which allows users to link to and view structured data from Wikidata on a test version of Commons.
2. We're creating "Multi-Content Revisions," which streamlines the backend such that Commons users can more efficiently add, edit, and manage structured data.

While the test system was set up and announced for testing in July 2017, the work on Multi-Content Revisions is ongoing and expected to be finished by the end of the calendar year. In tandem, federation and Multi-Content Revisions comprise the major infrastructure updates necessary to enrich Commons with structured data.

Test Federation:

In January 2017, Wikimedia Deutschland engineers continued the work to develop a test version of Wikibase "federation," which allows users to test a basic version of Structured Data on Commons. The federation, which debuted in July 2017, lets Wikidata's items and properties describe media files on Commons until Multi-Content Revisions allows us to store the structured data on the file page. The information for each file on the test site is spread across two pages:

- One page that's similar to a Wikidata page, which allows users to note the media item's structured data "statements" such as what's depicted, the item's photographer, and its coordinates (which allow geo-locating of the object). See <http://structured-commons.wmflabs.org/wiki/MediaInfo:M13>.

- One page that's similar to a Commons page, which contains the actual image and allows users to note the file's standard elements such as the date/time the image was taken and the photographer. See <http://federated-commons.wmflabs.org/index.php/File:LighthouseinDublin.jpg>

Below: The [test Federation page](#) that depicts the image and non-structured information



File Discussion

File:LighthouseinDublin.jpg

File File history File usage Metadata



Size of this preview: 800 × 600 pixels. Other resolutions: 320 × 240 pixels | 2,048 × 1,536 pixels.
 Original file (2,048 × 1,536 pixels, file size: 375 KB, MIME type: image/jpeg)

Here is some unstructured text

File history

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

	Date/Time	Thumbnail	Dimensions	User	Comment
current	12:42, 5 July 2017		2,048 × 1,536 (375 KB)	Lydia Pintscher (talk contribs)	

• You cannot overwrite this file.

File usage

There are no pages that link to this file.

MediaInfo:M13

Metadata

This file contains additional information, probably added from the digital camera or scanner used to create or digitize it. If the file has been modified from its original state, some details may not fully reflect the modified file.

Below: The [Structured Data page](#) associated with that image that allows Commons users to add and test more structured metadata associated with that image



[Main page](#)
[Recent changes](#)
[Random page](#)
[Help](#)

[Tools](#)
[What links here](#)
[Related changes](#)
[Special pages](#)
[Printable version](#)
[Permanent link](#)
[Page information](#)
[Concept URI](#)

[In other languages](#)
[Add links](#)

MediaInfo Discussion

Poolbeg Lighthouse (M13)

File:LighthouseinDublin.jpg
worm's eye view of a lighthouse [edit](#)

[In more languages](#) Configure

Language	Label	Description
English	Poolbeg Lighthouse	worm's eye view of a lighthouse

[All entered languages](#)

Statements

Photographer [edit](#)

Lydia Pintscher [add reference](#)

[0 references](#)

[add](#)

license [edit](#)

CC-BY [add reference](#)

[0 references](#)

[add](#)

Coordinates [edit](#)

53°20′31.56″N, 6°9′6.12″W [add reference](#)

[0 references](#)

[add](#)

Date taken [edit](#)

22 September 2014 [add reference](#)

[0 references](#)

[add](#)

Both test pages link to the corresponding other test page. Both test pages are allowing users to experiment with uploading a media file and creating a corresponding Structured Data page – as on July 7, 2017, when the veteran Turkish editor [User:HakanIST](#) added an image of Istanbul's Üsküdar University and a [corresponding Structured Data page with statements](#).

Below: [User:HakanIST's test addition](#) of a Structured Data on Commons image from July 7, 2017:



File Discussion

File:Uskudar university.jpg

File File history File usage Metadata



Original file (2,898 × 1,932 pixels, file size: 1.13 MB, MIME type: image/jpeg)

One of the previous 500 × 333 pixels. Other resolutions: 300 × 212 pixels | 2,438 × 1,652 pixels.

Uskudar University front gate in winter

File history

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

	Date/Time	Thumbnail	Dimensions	User	Comment
current	06:37, 7 July 2017		2,898 × 1,932 (1.13 MB)	HakanIST (talk contribs)	Uskudar University front gate in winter

• You cannot overwrite this file.

File usage

There are no pages that link to this file.

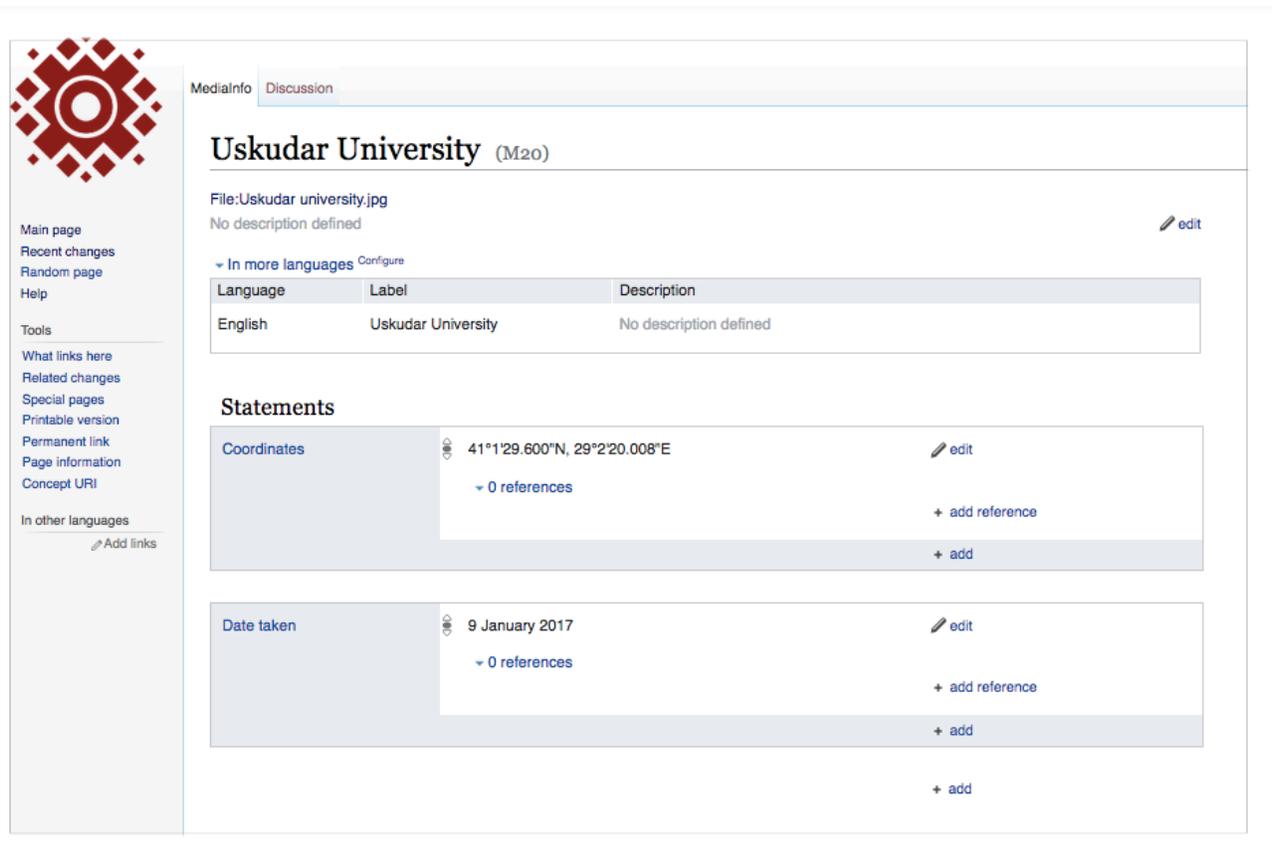
MediaInfo:M2o

Metadata

This file contains additional information, probably added from the digital camera or scanner used to create or digitize it. If the file has been modified from its original state, some details may not fully reflect the modified file.

Camera manufacturer	Canon
Camera model	Canon EOS 70D
Exposure time	1/250 sec (0.0015625)
F Number	f/16

Below: User:HakanIST's [test addition](#) of a Structured Data on Commons page from July 7, 2017:



The screenshot shows a Wikimedia Commons page for 'Uskudar University'. The page includes a sidebar with navigation links, a main content area with a title, a file upload section, a table of translations, and a 'Statements' section with two entries: 'Coordinates' and 'Date taken'. Each statement has an edit icon, a reference count, and options to add references or new statements.

Uskudar University (M20)

File:Uskudar university.jpg
No description defined [edit](#)

[In more languages](#) Configure

Language	Label	Description
English	Uskudar University	No description defined

Statements

Coordinates [edit](#)
[+ add reference](#)
[+ add](#)

Date taken [edit](#)
[+ add reference](#)
[+ add](#)
[+ add](#)

More information about the Test Federation: Phabricator page that heralds the federation's software beginnings: <https://phabricator.wikimedia.org/T156114>

Multi-Content Revisions:

Currently, a page using MediaWiki software can only store one type of data: wikitext, or [JSON](#), or structured Wikibase data. Our work on Multi-Content Revisions is one of the engineering cornerstones of the project's first phase. Multi-Content Revisions is an update to the core MediaWiki

software (the software that all Wikimedia wikis run on) that will allow multiple types of content, such as wikitext, JSON, and structured data to live on the same wiki page. With this, Commons users can more efficiently add, edit, and manage structured Wikibase metadata associated with Commons media files.

Multi-Content Revisions streamlines the MediaWiki editing process, amalgamating all the different editing components – text, code additions, structured data, etc. – so that a single page can hold and display the media, structured, and unstructured data. This improves the users’ experience, but also makes the infrastructure more robust, because MCR provides the support to reduce the likelihood of data entry error or malformed input.

The main objectives of Multi-Content Revisions:

- Allow auxiliary information that is currently embedded in wikitext, like the content in information templates and the coding for [category names](#) (the categories that define Wikipedia’s breadth of articles), to be separated and managed in their own “slots.” The slots will maintain the shared history and page-level functionality associated with the main content – and will allow the auxiliary information to be edited at the same time as the main content.
- Allow auxiliary information that is currently managed on associated pages, like [template documentation](#) (the coding and headings that editors use to create pages) and [gadget styles](#) (the MediaWiki extension that allows certain JavaScript coding or Cascading Style Sheets coding), to share the edit history and page-level functionality associated with the main content – and will allow the auxiliary information to be edited at the same time as the main content.
- Investigate a generic way to associate derived data like maps with the same, single revision.

Now, metadata about media files has to be encoded in free-form wikitext in a semi-structured way (on file description pages), which makes it hard to query and re-use. With MCR, we can separate the meta-data from the free-form text and encode it as JSON (using the wikibase schema), while still treating both kinds of content as parts of the same "page". This means it is treated as a single unit for administrative operations like watching, protecting, renaming, or deleting a page. This allows us to introduce machine readable meta-data with minimal disruption for the existing community workflows on Wikimedia commons.

Structured Data on Commons could proceed without Multi-Content Revisions, but it would be much less efficient. Users would have to update data in multiple locations (a file page, a file, and a structured data page). They would have to subscribe to multiple notifications to monitor changes to different files. Multi-Content Revisions helps economize the MediaWiki backend of Commons, and lays the foundation for a more efficient and effective user experience. Though this change will likely be largely invisible to editors on the Wikimedia projects, it allows for a number of future software changes that make contributing files and other content more streamlined. The entire histories of files will be stored more efficiently. Multi-Content Revisions will make it easier to transition from the current method of free text and semi-structured data to a Commons that is grounded in Structured Data. It's designed to add a degree of freedom to MediaWiki, which will let us implement new features, and allow some features that have been hacked in in the past to work much more efficiently, smoothly, and in a user-friendly way. And MCR will lead to many other benefits and improvements – like putting related data on the same page, as with video and subtitle files.

Daniel Kinzler, the technical lead for Wikimedia Deutschland, is leading the engineering work on Multi-Content Revisions, helped by the MediaWiki Platform engineering team. The engineering work on Multi-Content Revisions has been ongoing, and we are looking to finish the core foundation of the work by December 31. Its advent will be an important milestone for the whole project.

More information about Multi-Content Revisions: A PDF that explains, with visuals, how Multi-Content Revisions will streamline and de-clutter the editing process: <https://upload.wikimedia.org/wikipedia/commons/3/37/WikiDev2017-MCR.pdf>. “Request for Comment” that explains the technical and engineering reasoning behind the changes, and includes a video of an MCR engineering strategy session: https://www.mediawiki.org/wiki/Requests_for_comment/Multi-Content_Revisions. Phabricator page, started in July 2015, that began the work on Multi-Content Revisions: <https://phabricator.wikimedia.org/T107595>.

Community Input: The Work of Jarek Tuszynski

Even in this early stage of Structured Data on Commons, software engineers who are longtime members of the Wikimedia community have embraced the project, and have been doing significant preparation work to integrate Wikidata links into Commons. A good example: volunteer [Jarek Tuszynski](#), a [top engineer](#) with the Virginia-based company Science Applications International Corporation ([SAIC](#)), has been rewriting infoboxes on Commons' files in the Lua programming language

to pull metadata from Wikidata, and has been linking Commons pages with Wikidata, often using a tool called [mix-n-match](#) developed by [Magnus Manske](#) (another prolific volunteer contributor).

At Wikimania 2017, Jarek [gave a talk](#), “[New Frontier: Using Wikidata on Commons](#),” where he explained to other developers and community members the work he and other volunteers are doing.¹³ Jarek specializes in reworking information from infoboxes and templates. Jarek and other volunteers have already entered Wikidata links for 60,000 art-related media files on Commons, and Wikidata links for tens of thousands of other Commons files.

“Hopefully,” he said about Structured Data on Commons, “the lessons learned from migrating the current infoboxes to Wikidata will be helpful.”

The Wikimania audience applauded Jarek’s talk, which preceded Sandra’s talk and took place in the same room. At Wikimania, some of the most prolific Commons tool developers said they were on board with updating their tools to work with Structured Data on Commons. Their enthusiasm – and the work they are already doing – is a poignant reflection of the project’s collaborative nature.

¹³ Jared’s talk begins at 3:15

Nine Months in Review: A Timeline of Key Developments

Here are key developments from the past nine months:

Date	Development
Jan. 9, 2017	<p>Sloan’s Grant Announced: The Wikimedia Foundation announces the three-year grant from the Alfred P. Sloan Foundation. We announce the grant on our blog, at https://blog.wikimedia.org/2017/01/09/sloan-foundation-structured-data/; at the Commons pages devoted to the project, at https://commons.wikimedia.org/wiki/Commons:Structured_data/Sloan_Grant; and at the talk page of our Structured Data project, at https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Archive_2017#Grant_to_fund_work_on_Structured_Data_on_Commons.</p>
Jan. 10, 2017	<p>Engineering Work on Multi-Content Revisions Continues: At a San Francisco developer summit, engineers from the Wikimedia Foundation and Wikimedia Deutschland discuss how to implement Multi-Content Revisions, one of the engineering cornerstones of the project’s first phase. The discussions segue from a Wikimedia Deutschland “Request for Comment” about Multi-Content Revisions that was initiated in August 2016. See https://www.mediawiki.org/wiki/Wikimedia_Developer_Summit/2017/Multi-Content-Revisions. See also https://www.mediawiki.org/wiki/Requests_-_for_comment/Multi-Content_Revisions and https://phabricator.wikimedia.org/T107595.</p>
Jan. 24, 2017	<p>Building of Test Federation Continues: Wikimedia Deutschland continues the work to develop a test “federation” by setting up a system that would allow users to test a basic version of Structured Data on Commons. The federation will let Wikidata’s items and properties describe media files on Commons. Work on the federation began in 2016. See https://phabricator.wikimedia.org/T156114.</p>
June 26, 2017	<p>Program Manager Hired: Amanda Bittaker is named the Wikimedia Foundation’s new program manager on the Structured Data program. See https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Archive_2017#Welcome_Amanda_Bittaker_as_program_Manager_for_Structured_Commons.</p>

Date	Development
July 3, 2017	<p>Community Liaison Hired: Sandra Fauconnier joins the Wikimedia Foundation as the new community liaison focusing on the Structured Data program. Sandra is supporting the collaboration between the communities (Commons, Wikidata, GLAM, etc.) and the product development teams involved at the Wikimedia Foundation and Wikimedia Germany. See https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Archive_2017#Welcoming_Sandra_Fauconnier.2C_our_new_Structured_Data_community_liaison.</p>
July 5, 2017	<p>Research Interviews Begin with GLAM Institutions: Senior Design Researcher Jonathan Morgan begins researching GLAM institutions' batch upload workflows, and conducting interviews with Commons contributors at museums, universities, and other institutions. The interviews will take place over the next four months. See https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions.</p>
July 6, 2017	<p>Test Federation Debuts: We announce that engineers have developed a test "federation," which allows users to test a very basic preview version of Structured Data on Commons. The test lets Wikidata's items and properties describe media files on Commons. See https://commons.wikimedia.org/wiki/Commons_talk:Structured_data/Archive_2017#New_step_towards_structured_data_for_Commons_is_now_available:_federation and https://www.mail-archive.com/commons-l@lists.wikimedia.org/msg03559.html.</p>
July 19, 2017	<p>First Newsletter Published: The Wikimedia Foundation publishes the project's first newsletter, which announced updates on the project, including the federation, research aims, and new hires. (The newsletter already has more than 150 subscribers.) See https://www.mail-archive.com/commons-l@lists.wikimedia.org/msg03565.html.</p>
Aug. 9-13, 2017	<p>Presentations at Wikimania: Staff working on Structured Data on Commons make presentations at Wikimania in Montreal. More than 900 people attend the year's most important gathering for those involved in the Wikimedia projects. Among the talks: Community Liaison Sandra Fauconnier's talk, "Structured Commons: What Changes Are Coming?" and Design Manager Nirzar Pangarkar's discussion about potential user-interface designs for the Structured Data on Commons. At Wikimania, Senior Design Researcher Jonathan Morgan also interviewed Commons users about their use of the sites and what particular functions they might need. See https://wikimania2017.wikimedia.org/wiki/Submissions/Structured_Commons:_what_changes_are_coming%3F and https://wikimania2017.wikimedia.org/wiki/Submissions/Structured_Data_on_Commons_design_discussion and https://upload.wikimedia.org/wikipedia/commons/4/44/Structured_Data_on_Commons_-_What_changes_are_coming%3F_-_presentation_at_Wikimania_2017.pdf and https://wikimania2017.wikimedia.org/wiki/Programme.</p>

Date	Development
Aug. 21, 2017	Product Manager Hired: Ramsey Isler joins the Wikimedia Foundation as the new product manager on the multimedia team, whose primary focus will be on implementing user-facing editing, upload, and search features for Structured Data on Commons.
Aug. 24, 2017	Scoping of Multi-Content Revisions: Daniel Kinzler, the technical lead for Wikimedia Deutschland, streamlines the engineering task-management of Multi-Content Revisions, creating subtasks and parent tasks that will drive the engineering work for the next four months. See https://phabricator.wikimedia.org/T174022 .
Sept. 8-10, 2017	Wikidata Product Manager speaks at WikiCon: Lydia Pintscher, PM for Wikidata, attends WikiCon in Leipzig, Germany. In this conference of German-speaking Wikipedians, Lydia is part of an important session about Commons, “Ist Commons noch zu retten?,” which also explained the history, background, and current state of the Structured Data on Commons project to attendees. See https://de.wikipedia.org/wiki/Wikipedia:WikiCon_2017 and https://wikicon.regulus.uberspace.de/de/wiki-con2017/public/events/55 .
Sept. 11, 2017	New User Stories Begin: Ramsey Isler begins researching new user stories that describe the users’ practices and motivations – and the ways that the Structured Data on Commons project can aid their use of Commons. Among other things: The user stories are helping us with design research and deciding what first features to build. See https://phabricator.wikimedia.org/T175589 .
Sept. 25, 2017	Designer Hired: Pam Drouin joins the Wikimedia Foundation as a UX Designer on the multimedia team. She will specialize in user and contribution experiences related to Structured Data on Commons.
Oct. 13, 2017	<p>Invitation Extended for Focus Group: Sandra puts on-wiki a formal invitation to join the focus group of Wikimedia community members for Structured Commons. This invitation notes that the group “is not for decision-making: that is a process that needs to be done by the community at large! It’s more for (smaller) conversations and consultations.” See https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved/Community_focus_group and https://phabricator.wikimedia.org/T174142.</p> <p>GLAM Users Invited to Take Survey: Jonathan Morgan adds an announcement to the main Structured Data on Commons page, asking GLAM users to take a 15-minute survey on how they upload images to Commons. See https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions/Survey.</p>
Oct. 25, 2017	Second Newsletter published: The second newsletter is published. Among its updates: An announcement that Commons User:Rama has published an article about Structured Commons in Arbido, a Swiss online magazine for archivists, librarians and documentalists. See https://meta.wikimedia.org/wiki/Structured_Data_on_Commons/Newsletter/2017-10-25 .

Date	Development
Oct. 29, 2017	Community Liaison Speaks at WikidataCon: Sandra Fauconnier speaks at WikidataCon in Berlin, in a talk titled, “Structured Data on Wikimedia Commons: What’s Coming and How to Be Involved as Wikidatans.” See https://media.ccc.de/v/wikidata-con2017-10038-structured_commons_and_wikidata and https://commons.wikimedia.org/wiki/File:Structured_Common_and_Wikidata_-_WikidataCon_2017.pdf and https://www.wikidata.org/wiki/Wikidata:WikidataCon_2017/Submissions/Structured_Data_on_Wikimedia_Commons:_what%27s_coming,_and_how_to_be_involved_as_Wikidatans .

Research and Interviews: Finding the Sweet Spot for Structured Data on Commons

In the first six months we identified three types of users who are central to the success of the structured data on Commons program – but who we need to know more about:

1. GLAM users
2. Commons curators
3. Commons media reusers

We have scoped generative user research to investigate the workflows and needs of these groups, including interviews, surveys, and focus groups, which will culminate in research reports and personas for each group.

This research will reveal these groups' practical use of Commons, and the ways we can help them get the most from Structured Data on Commons. For Structured Data on Commons to succeed, it needs to integrate with core volunteers' workflows, eliminate barriers and pain points to their participation, and optimize their ability to contribute. We need to fundamentally understand users' habits, users' aspirations, and users' frustrations, and we are doing just that through in-depth user research.

We've already spoken with a cross-section of Commons users from different parts of the Wikimedia community, and we've already initiated surveys that will elicit more important feedback.

Among the research and interviews from the past year:

- From July to September 2017, Jonathan Morgan, the Wikimedia Foundation's Senior Design Researcher, conducted 11 detailed Structured Data on Commons interviews with key members from GLAM institutions and the Wikipedia community. Jonathan did the interviews – which averaged more than an hour each – through face-to-face video chats.
- In August at Wikimania 2017 – the year's largest Wikimedia gathering of users and Wikimedia staff members – Community Liaison Sandra Fauconnier and Design Manager Nirzar Pangarkar gave talks that collectively filled more than 100 seats, and many participants offered feedback on their vision of what Structured Data on Commons should be.
- Also at Wikimania, Jonathan Morgan interviewed Commons contributors about their use of the sites and what particular functions they might need. And Jonathan Curiel – the Wikimedia Foundation's Senior Development Communications Manager – interviewed six veteran Commons users in sit-down sessions that revealed important suggestions for the project.
- Later in August, Sandra met with staff from Europeana to discuss the organization's aspirations for Structured Data on Commons and other Wikimedia projects. Among Europeana's representatives in the meeting: Europeana's Wikimedian in Residence Liam Wyatt; R&D manager Antoine Isaac; Data R&D Coordinator Valentine Charles; Search Engineer Timothy Hill; and Nuno Freire, Senior Data Specialist of the European Library. They discussed how the Wikimedia Foundation can work with Europeana on Structured Data – perhaps through a pilot program or through formal Wikidata training sessions. They also discussed Europeana's biggest interest:

The ability of Structured Data on Wikimedia projects to increase the interoperability of their collections.

- In October 2017, we [invited](#) Commons users to join a focus group of Wikimedia community members who will consult with us on Structured Data on Commons. (We are still formulating the structure of the volunteer group.)
- Also in October 2017, we [announced a new survey](#) for GLAM users of Commons. The 15-minute survey, which Jonathan Morgan designed, focuses on how GLAM users upload images to Commons. Over 100 GLAM users have taken the survey so far. (Sandra provided valuable feedback on survey design and disseminating it through channels where the right people could see it – which led to a high response rate.)

Beyond this work, the foundation's GLAM outreach team is writing a larger narrative strategy to bring in even more people from the GLAM-Wiki community, including an advisory group that would include members of two of the world's biggest digital aggregators of collections from heritage organizations: Europeana and DPLA (Digital Public Library of America). The two organizations supported the Wikimedia Foundation's funding application for Structured Data on Commons, and their further involvement would solidify our project's ability to succeed.

With Structured Data on Commons, cultural heritage organizations can leverage the connections between Commons, Wikidata and individual objects within their collections. Libraries and museums are increasingly using Wikidata to complement the items in their collections. With Structured Data on Commons, we are doing work that parallels their own digital enrichment projects. We are involving GLAM institutions in our early stages of Structured Data on Commons to ensure structured data on Commons is implemented in a way that satisfies their needs and encourages them to contribute..

Over the next two years, we plan to improve the development of a better community of support to help GLAMs take advantage of the linked data ecosystem developing around Wikidata and Commons. Alex Stinson, the Wikimedia Foundation's GLAM-Wiki Strategist, who has academic training as a digital humanist – and previously supported on our Wikipedia Library project – has already established tangible relationships with Europeana, DPLA, and other prominent cultural organizations, like the Smithsonian and National Archives of the United States, and those connections have been crucial as we continue to go forward with Structured Data on Commons. Europeana [already encourages](#) its thousands of member museums, libraries and archives to add their collections' meta-

data to Wikidata. (See also [here](#).) Europeana has worked closely with many Wikimedia affiliates and projects over the years, including development of the GLAMwiki Toolset mass upload tool. On October 28-29, 2017 in Berlin, Valentine Charles of Europeana [spoke](#) at Berlin's [Wikidata conference](#), in a talk titled "Structured Data: How Can GLAMs Grab the Low-Hanging Fruit?" We are seeing increased interest in Wikidata-related initiatives like Europeana's across the sector, including [a project by the Association of Research Libraries](#). By supporting the development of these multi-organizational conversations about how to use Wikidata, WMF staff and community members are preparing the GLAM community for using a Commons with Structured Data as well.

Research-Driven Design: How Structured Data on Commons Is Coming to Life

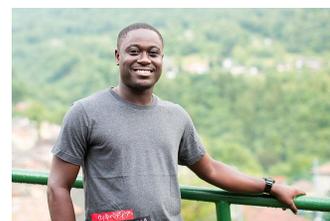
The design of Structured Data on Commons will help determine the success of the project. We want the design to be functional and inviting – to give new contributors an ease of use that also recognizes the demands of veteran contributors who may want an interface that's more elaborate. And similar to the entire project, the design portion is emphasizing community feedback – through research interviews, on-wiki discussions, workshops, seminars, and conferences. We aren't just building technology for technology's sake – we're understanding user motivations, why people are using Commons, and what they want from it in the foreseeable future.

Our research is ongoing, but we've already spoken with more than 60 Commons users, led by Jonathan Morgan's detailed Structured Data on Commons interviews with 11 diverse members from GLAM institutions and the Wikipedia community. In each interview, Jonathan is asking about

users’ workflows – i.e., the practical mechanics of their Commons contributions that takes them from an idea in their head to an action on our website – and asking how we can better those workflows. (His report on those interviews is forthcoming, and will be available to read at https://meta.wikimedia.org/wiki/Research:Supporting_Commons_contribution_by_GLAM_institutions.)

Among the people whom Jonathan interviewed:

- Esther Sole, an art historian who’s a former Wikipedian-in-Residence with Fundació Joan Miró
- João Alexandre Peschanski (photo top right¹⁴), who is communications supervisor with the Research, Innovation and Dissemination Center for Neuromathematics at the University of Sao Paulo
- Mauricio Genta, a Wikipedian in Residence with Wikimedia Argentina
- Netha Hussain (photo right¹⁵), a prominent Wikipedian from India
- Olaf Janssen, who is Open data and GLAM-wiki coordinator with the National Library of the Netherlands/Koninklijke Bibliotheek
- Susan Tolich, New Zealand’s first Wikipedian in Residence, with the Auckland Museum
- Jesse de Vos, GLAM-wiki coordinator with the Netherlands Institute of Sound and Vision
- Delyan Drumev with the Bulgarian Archives State Agency
- Felix Nartey (photo below right¹⁶) of the Ghana National Archive and who is the 2017 Wikimedian of the Year



A small example of the interviews: João Alexandre Peschanski told Jonathan that the Research, Innovation and Dissemination Center for Neuromathematics at the University of Sao Paulo wanted more freedom to upload images to Commons. They could “only upload 30 images at a time,” he said. “It took two weeks for an upload because we have to write up all the information. Sometimes

¹⁴ Photo by Miguel Galves, CC BY-SA 3.0 https://commons.wikimedia.org/wiki/File:374_1060625628855_8610_n.jpg.

¹⁵ Photo by Adam Novak, CC BY-SA 3.0 https://commons.wikimedia.org/wiki/File:Netha_Hussain-925.jpg

¹⁶ Photo by Ruby Mizrahi, CC BY-SA 3.0 https://commons.wikimedia.org/wiki/File:Felix_Nartey_2.jpg

we need to do research to find out how the object came into the collection. Interesting aside: during the initial uploads, they were all eliminated. We didn't understand [OTRS](#)," which is the system that is used to verify and archive licensing permissions.

Another example: [Netha Hussain](#), a longtime Wikimedian who's a medical doctor from Kerala, India, told Jonathan about the laborious efforts she undertook to upload her hospital's pathology slides to Commons. The slides show rare conditions that would help other doctors diagnose patients. The original slides were in a glass case in her hospital's pathology lab. The lab had no access to the Internet, so the director let Netha take home digital files on a computer. From her house, Netha used her cellular data plan to upload the files to Commons – but the files had little metadata. The slides weren't accompanied by details that linked the patients' conditions to the images or their magnification. So Netha did the magnification herself, and added the complex descriptions, which includes the type of glass used in the slide since different types of glass refract light differently, which affects how the tissue sample is displayed in a digital image. The glass type is an important facet of describing the images – but where does that go in a regular Commons file? Structured Data on Commons would make Netha's pathology slides – and the [more than 700 other images she's uploaded](#) – much more useful, and she's excited about the project's potential to inform people and save lives.

Jonathan's interviews, which last more than an hour each, are providing valuable insight into the needs and wants of Commons users. "We are conscious that there's a lot we don't know about how Commons works," he says. "GLAM-Wiki creates opportunities to bring in new types of media on Commons. With each new type of opportunity, we'll discover new challenges."

because

Pam Drouin, who in September joined the Wikimedia Foundation as a UX Designer and specializes in user and contribution experiences related to Structured Data on Commons, will work with Jonathan to do more in-depth interviews in the fall with Wikimedia Commons curators.

The additional interviews will help dictate the design of the project and our User Interface, but Design Manager Nirzar Pangarkar did early mock-ups and his team has gotten additional feedback on how the user interface could improve the users' experience and ability to contribute. These very early designs, which have emphasized clean lines and layouts, are giving users a better sense of the project, as with the designs on the following pages:

On the following three pages: Early mock-ups of what Structured Data on Commons pages could look like



COMMONS

English Nzr12345 Talk Preferences Beta Watchlist Uploads Contributions Log out

File Discussion
View Edit History More



Mona Lisa

monalisa-2618.jpg

The painting is thought to be a portrait of Lisa Gherardini, the wife of Francesco del Giocondo, and is in oil on a white Lombardy poplar panel. It is believed to have been painted between 1503 and 1506; however, Leonardo may have continued working on it as late as 1517.

Uploader

[NirzarP \(WMF\)](#)

Owner

Ansel Adams

 **Public domain**
Free for use by anyone for any purpose without restriction under copyright law. [know more](#)

About Camera License [+ Add](#)

Instance

✎

Artist

Leonardo Di Vinci

Below and on the following two pages: Early mock-ups of what Structured Data on Commons pages could look like



COMMONS

English Nrr12345

Talk Preferences Beta Watchlist Uploads Contributions Log out

File Discussion

View Edit History More

Search Wikimedia Commons



Uploader: NirzarP (WMF) | Type: image / jpg

Dimension: 2000 x 3000 px | Size: 3.2 megabytes

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%3A%3C...>

Attribution for reuse REQUIRED
Nirzar Pangarkar / Wikimedia Commons / "Lüding..."

Nighthawks by Edward Hopper 1942.png

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

About

Camera

Location

Usage

+ add section

instance of

Painting ✎

Genre Painting ✎

+ add another

inception

1942 ✎

1 reference imported from [DBpedia](#)

+ add another

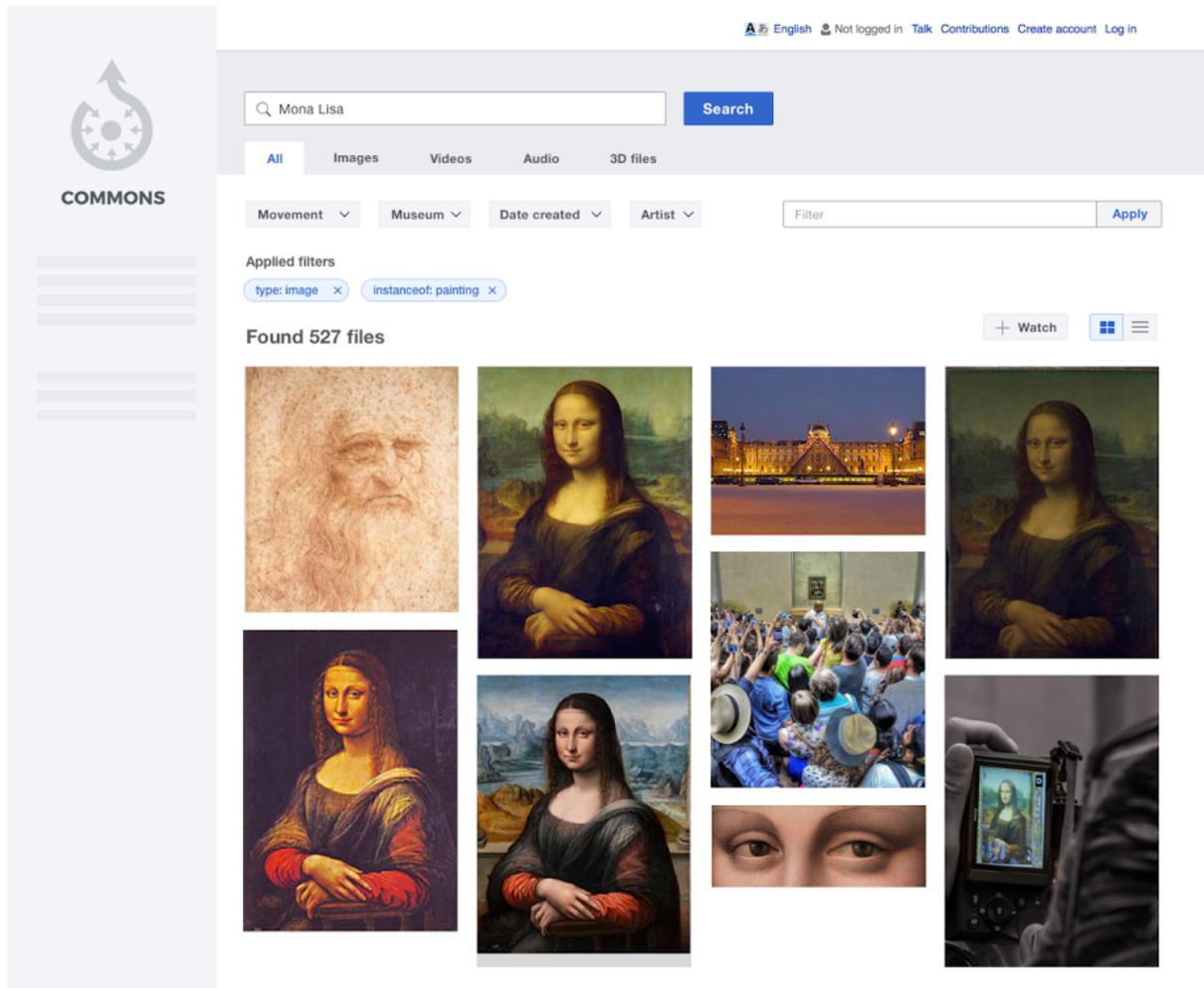
movement

Realism ✎

+ add another

+ add new property

Below: An early mock-up of what a Structured Data on Commons page could look like



At a design workshop at Wikimania 2017 in Montreal, where Nirzar Pangarkar introduced mock-up Structured Data on Commons designs to a crowded room of interested Commons users, attendees [made a number of helpful suggestions](#) that Program Manager Amanda Bittaker – who also spoke – noted on large pieces of drafting paper. Leading up to those suggestions, Nirzar talked about “smart filters” and tags that could allow Commons users to quickly filter images through search terms that narrowed a search to specific subjects like “museums with Monet paintings” and specific dates like “paintings made between 1900 and 1910.”

“We can apply algorithms,” said Nirzar, whose talk focused on the file page editing and search potential of Structured Data on Commons. “We can surface those properties that are relevant to a particular search term. . . .We are navigating around how we can make it easier to digest and use so much information.”

Nirzar also discussed the software designs that will help users manage the media files that they get. And he emphasized that we can design for complexity and create different interfaces for different actions.

Through Wikidata and text that is already on Commons, Commons already has enormous amounts of structured data on image files, Nirzar noted. In essence, Structured Data on Commons will make it even easier to search for media files on Commons than to search on Google. But that’s in the future. At Wikimania, the attendees peppered Nirzar with questions.

“I would have a suggestion,” said an attendee who asked the first question. “Something I’d be very interested in is to differentiate pictures depicting Monet and done by Monet. Just putting ‘Monet’ in a search doesn’t exactly capture the role we’d like to see.”

“Yeah,” said Nirzar. “We are thinking about what are properties and what are search phrases. You are absolutely right. . . . We have to design for all cases.”



Photo by Jonathan Curiel

Above: Nirzar Pangarkar (center at podium) answering questions at Wikimania 2017, with Amanda Bittaker (at right) writing down attendees' responses for everyone to see

Researching Usage: How to Measure the Effectiveness of Structured Data on Commons

Potential Progress Measurements:

As part of the project, we are also researching how to monitor our progress – how to gauge how people are using Structured Data on Commons. We want to use “before” and “after” metrics that quantify how people are searching on the new Commons, how people are using the files, and how they’re doing other activities. And we want to do this in multiple languages – not just English – since one objective of the project is increasing the multilingual usability of Commons, whose meta-data are now mostly in English.

One example of a possible metric: How many Commons searches are made in a particular language both before and after the completion of Structured Data on Commons and how many file pages currently contain info in those languages. We could do targeted outreach to promising language communities that may be unaware that Commons offers so many freely usable media files. So a more useful Commons could lead to increased usage among different language communities.

Other possible measurements:

- **Relevance:** Did users find what they wanted? We could measure that through such components as “satisfied clicks” or satisfaction surveys.
- **Time:** Did users quickly find what they wanted? We could measure that through quantitative metrics.
- **Effort:** How many clicks did it take for users to find what they wanted? We could also measure that through quantitative metrics by test sample of searches.

The Wikimedia Foundation’s [Search Platform analysts](#) are now [researching these possible metrics](#). Their work will take them at least through the end of this calendar year. The search habits that they identify will let us create snapshots of usage that – at a glance – will tell us how the project is doing and whether we want to make any changes in emphasis.

Identifying the MVP:

This quarter, we are also identifying what should be the project’s first “Minimal Viable Product.” Guided by Ramsey Isler, the teams decided on a likely first feature to present to the Commons community: multilingual file captions, which would let users in hundreds of languages use Commons with an ease they’ve never had before. Netha Hussain, for example, is active on both the English and Malayalam Wikipedias, but her Commons uploads in Malayalam – a language spoken by almost 40 million people – are not really searchable in Malayalam or other languages, so her pathology slides aren’t nearly as useful as they could (and should) be.

Why Reuse Is Important:

Our project is adapting multiple ways to improve Commons. At the same time that we’re working with GLAMs to strengthen the work that they do, we’re improving the way that individual users – from veteran editors to casual uploaders – can use Commons. And we’re working on ways to encourage reuse of Commons media – to make it easier to share or repurpose Commons’ millions of media files, which would better integrate those millions of files into the wider Internet and better integrate those files into the lives of our users.

The Structured Data on Commons projects is more than just “data,” more than the engineering work we’re completing, more than the outreach we’re doing, and more than the collaboration that’s already happening between the Wikimedia Foundation, GLAM institutions, and the Wikimedia Foundation and individual users. It’s about creating a more involved community of Commons users that’s better connected across the Internet. Everyone who uses the Internet for media files, or a library or cultural institution for research, can be a Commons user and a Commons contributor – can discover, can learn, and can use media files that are important to them.

Structured Data on Commons will expand the Internet and give people a more expansive way to understand the world. That’s why the project is so important to us. And why we have marshaled so many people from across the Wikimedia Foundation and Wikimedia Deutschland to work on the project. Even small contributions are making a big difference. The past year has shown that, we believe. One step at a time, we can indeed see the future of Commons.

Deliverables: What We Achieved in Engineering and Other Project Work

Our 2016 proposal divided the Structured Data on Commons project into three years. Some of the early project goals that we envisioned a year ago – like having a good technical foundation to store metadata by June 30, 2017 – were delayed as hiring was delayed, but these goals are now in progress, and will be complete by the end of this year. The project’s overarching timeline is still in tact.

In our proposal, we said that, by 2020, “the infrastructure will be in place to migrate Commons’ millions of media files to structured data.” We are still on track for that. We will still accomplish our overarching goals. With a more detailed roadmap, we’re finding places where we can work in parallel and deliver milestones more quickly. We’re also bringing in more engineering resources to make up for the time lost at the beginning of the project. Each year’s goals are designed to bring us closer to our vision of Structured Data on Commons.

DELIVERABLES JAN. 2017 TO SEPT. 2017					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Program Management	Multi-team collaboration	Aligned team on commitments and deliverables, set up team norms and processes for communication. Held team offsite in August 2017.	There were a lot of possibilities of what to work on, and not enough clarity on who would be responsible for what. At offsite, gained clarity on most important outcomes and who would be responsible for them.	Building a first feature that incorporates all the infrastructure, editing, upload, and search, so that we can learn to effectively hand off work between 5 teams by July 2018.	✓
Program Management	Key personnel hired	All key personnel hired.	It takes time to hire the right people. But, the right people are worth the wait.	Now we put them to work.	✓
Program Management	Aligned teams on program components	Assessed program components and documented findings including a stakeholder map, stakeholder analysis, communication plan, roadmap, and risk management doc.	This project is complex enough to require a lot of programmatic structure, which is a heavy upfront investment but will pay off in effective design, community collaboration, time management, and improved impact.	Create an M&E plan that identifies the metrics of success for the desired program outcomes by Feb 2018.	✓
Make editing meta-data easier	Wikibase Federation built and tested	Test site up at structured-commons.wmflabs.org/wiki/Main_Page	As we develop Federation, we've learned that many other institutions and knowledge bases could benefit from using Federated Wikibase. While this isn't part of this project, it is an interesting area of potential future work to support open knowledge and access.	Integrate with MCR by June 2018.	✓

DELIVERABLES JAN. 2017 TO SEPT. 2017					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Make editing metadata easier	Define schema for integrating structured data into Commons	New schema defined		Build and test MCR by December 2017.	
Make editing metadata easier	Test and prototype metadata as structured data, integration of structured data interface, review labels and vocabulary	Designed and provided an initial implementation of two new database tables that support the MCR feature: the comment table and the actor table.	The design work in this period led to a prototype of the functionality. This prototype was the starting point that resulted in the implementation that was pursued in the next quarter.	Build and test MCR by December 2017.	

DELIVERABLES JAN. 2017 TO SEPT. 2017					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Make editing metadata easier	Initial file page and search designs re-viewed with community	Initial file page and search designs reviewed with community in design feedback discussion at Wikimania 2017.	So far, community members have been very willing to give constructive feedback on initial designs, when they are appropriately framed as straw designs. As we continue the pattern of clearly framed requests and early and often feedback, we hope to strengthen reinforce this dynamic and improve our tools and increase the adoption of them by the community.	Prototype and test version of file and search pages starting in Jan 2018.	
Make editing metadata easier	Review concepts & data models with the Wikimedia Commons and Wikidata communities	Review concepts & data models with the Wikimedia Commons and Wikidata communities	Much of the research done by community members and staff over the past 3 years is directly applicable to the how structured data on Commons could be modeled.	Applying research review to design of new research and new feature designs.	

UPCOMING DELIVERABLES					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Make editing meta-data easier	Database storage layer complete	Work started in Oct 2017.	Nothing yet, work started in Oct. 2017.	Implementing MCR revision retrieval interface, implementing MCR page update interface, implementing MCR content metadata storage service.	
Make search more effective	Explore features for Search & Query	Scoped research to learn more about needs of searchers and reusers in August 2017. Defined success metrics for improving search effectiveness in Sept 2017.	We have a much deeper understanding of GLAM workflows and how they need to transform their data to integrate with Commons now. We learned that our current data model is very fragile and when it breaks it often leaves GLAM uploaders lost. With this deeper understanding we can ensure our upload tools work with existing GLAM data models and that in the case something does fail, it does so in a way that is easily understood and worked around.	Gathering baselines for metrics in October 2017.	

UPCOMING DELIVERABLES					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Satisfy GLAM use case	Review data models with partners	Scoped research to learn more about needs of GLAM institutions, how they store their data, how they want to integrate with Commons. Interviews completed in September 2017, survey launched in September 2017, and began compiling findings into a research report, to be finished by December 2017. Initiated working relationships with Creative Commons and IIF regarding integration and interoperability.	We have a much deeper understanding of GLAM workflows and how they need to transform their data to integrate with Commons now. We learned that our current data model is very fragile and when it breaks it often leaves GLAM uploaders lost. With this deeper understanding we can ensure our upload tools work with existing GLAM data models and that in the case something does fail, it does so in a way that is easily understood and worked around.	Define key GLAM user stories in conjunction with key GLAM stakeholders by Dec 2017. Investigate integration and interoperability with Creative Commons and IIF through Sept 2018.	
Make editing meta-data easier	Integration of structured data interface and Wikidata UI concepts in pages and tools. (ex Upload Wizard, Media Viewer, Visual Editor)	Initial straw designs presented to community at Wikimania in August 2017.	So far, community members have been very willing to give constructive feedback on initial designs, when they are appropriately framed as straw designs. As we continue the pattern of clearly framed requests and early and often feedback, we hope to strengthen and reinforce this dynamic and improve our tools and increase the adoption of them by the community.	Prototyping and testing of upload wizard starting in April 2018.	

UPCOMING DELIVERABLES					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Make search more effective	Search & Query: tested	Not testing yet, plan to begin in Jan 2018.	Nothing yet, plan to begin in Jan 2018.	Plan to begin search backend and search prototyping and testing in Jan. 2018.	
Effective program management	Feature launch feedback	Established several channels of communication with community including Commons and Meta talk pages, newsletter, IRC office hours, and focus groups, and mapped which stakeholders we can find in each channel.	People want to know what info will be available where, and how often. Setting clear expectations around this increases transparency and lowers stress for community members.	Reply to feedback on announcement of initial feature in November 2017. Facilitate communication around first feature development through July 2018.	
Support tool development	Support migration of tools	We are identifying key tools and tool builders in order to support their transition to supporting structured data on Commons. A comprehensive list should be ready by March 2018. Scoping support of first key tool, Monumental, in October 2017.	Because of the decentralized nature of volunteer built tools for Commons and Wikidata, it has been difficult to get a comprehensive list of important tools. To be as comprehensive as possible, we are using a mix of qualitative (interviews, consultations) and quantitative (tool lab usage metrics) to determine which tools our communities care about.	Determine the tool builder's needs and scope and schedule how to support them by May 2018.	
Allow license compliant reuse	Media Viewer License Display	<i>(Work to be done after Oct. 2018)</i>			

UPCOMING DELIVERABLES					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Make search more effective	Search & Query: deployed	<i>(Work to be done after Oct. 2018)</i>			
Support tool development	Tools: iteration and maintenance	<i>(Work to be done after Oct. 2018)</i>			
Make search more effective	Search and Query: iteration and maintenance	<i>(Work to be done after Oct. 2018)</i>			
Effective program management	Feedback and curation support	<i>(Work to be done after Oct. 2018)</i>			
Support tool development	Serve Commons community: Encourage tool development for engagement, API usage for usage, migration and contribution	<i>(Work to be done after Oct. 2018)</i>			

UPCOMING DELIVERABLES					
Goal	Objective	What We Did	What We Learned — and Why That's Important	What We're Doing Next	
Support tool development	Serve Commons community: Encourage tool development for engagement, API usage for usage, migration and contribution	<i>(Work to be done after Oct. 2018)</i>			

Where People Are Learning About — and Discussing — Structured Data on Commons

Here are the main places online where people are learning about and discussing Structured Data on Commons. In the forums for online discussions, staff from the Wikimedia Foundation are participating in conversations that are illuminating ways that the project can move forward.

- **Main Structured Data on Commons project page:** https://commons.wikimedia.org/wiki/Commons:Structured_data
- **Talk page of the main Structured Data on Commons project page:** https://commons.wikimedia.org/wiki/Commons_talk:Structured_data
- **Talk page of the Structured Data on Commons community focus group:** https://commons.wikimedia.org/wiki/Commons_talk:Structured_data
- **Phabricator, where we are detailing the project's work tasks – both engineering and non-engineering:** <https://phabricator.wikimedia.org/tag/structured-data-commons/>
- **Newsletter that features updates about the engineering and non-engineering work of the project:** https://meta.wikimedia.org/wiki/Structured_Data_on_Commons/Newsletter
- **“Get involved” page that outlines how people can participate in the project:** https://commons.wikimedia.org/wiki/Commons:Structured_data/Get_involved

Appendixes

As separate attachments, the Wikimedia Foundation is submitting four Appendixes. They are:

Appendix 1: Stakeholder Matrix, which analyses the different stakeholders (curators, uploaders, etc.) who will be engaged with Structured Data on Commons.

Appendix 2: User Stories, which details the many types of users and their motivations for using Commons, and how Structured Data on Commons can help them.

Appendix 3: Risk Management, which analyses the different project risks and our strategies to counteract them.

Appendix 4: Q2 Road Map, which outlines our planned work in Quarter 2 of the 2017-2018 Fiscal Year, from October to December 2017.