Structured Data on Wikimedia Commons

GLAM pilot projects workshop

Wikimedia Conference 2018
Hello!

Pam Drouin
User:PDrouin (WMF) pdrouin@wikimedia.org

Sandra Fauconnier
User:SandraF (WMF) sandraf@wikimedia.org

Alex Stinson
User:AStinson (WMF) astinson@wikimedia.org

Ben Vershbow
User:BVershbow (WMF) bvershbow@wikimedia.org

Liam Wyatt
User:Wittylama liamwyatt@gmail.com
Goals

1. **You** learn about Structured Data on Commons (SDC);
2. The **SDC team** learns what you need in order to be able to work with SDC;
3. We **all** leave this workshop with a list of ideas and potential pilot projects!
# Schedule

<table>
<thead>
<tr>
<th>Activity</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured Commons and GLAM: intro</td>
<td>30 minutes</td>
</tr>
<tr>
<td><strong>Workshop exercise 1:</strong> List your upcoming GLAM projects</td>
<td>5 minutes</td>
</tr>
<tr>
<td><strong>Workshop exercise 2:</strong> Project highlights and issues</td>
<td>5 minutes</td>
</tr>
<tr>
<td><strong>Workshop exercise 3:</strong> Changes and needs</td>
<td>60 minutes</td>
</tr>
<tr>
<td>Group summaries</td>
<td>10 minutes</td>
</tr>
<tr>
<td><strong>Wrapping up, conclusions and follow-up</strong></td>
<td>10 minutes</td>
</tr>
</tbody>
</table>
Parking lots
WMF Support for SDC Pilot Projects
Structured Data on Commons
Structured Data on Commons

converts metadata on Commons to a **structured & machine-readable format** making Commons files easier to view, search, edit, organize and re-use, in many languages.
Starry Night by Vincent Van Gogh, from Google Art Project. Public domain. Via Wikimedia Commons.
5 outcomes
(and one corollary outcome)

1. Make editing metadata easier
   (and enable multilingual statements)
2. Make search more effective
3. Satisfy GLAM use case
4. Allow license compliant reuse
5. Support tool development
Structured Commons - What lives where? (sample, doesn't cover everything)

You will see the things in this blue box on Commons

Some properties and items pulled from Wikidata

---

### Wikimedia Commons

<table>
<thead>
<tr>
<th>MediaWiki</th>
<th>Wikibase</th>
</tr>
</thead>
<tbody>
<tr>
<td>File name</td>
<td>File ID (M12345678)</td>
</tr>
<tr>
<td>Resolution</td>
<td>Title / Caption</td>
</tr>
<tr>
<td>Categories</td>
<td>Description text (string)</td>
</tr>
<tr>
<td>Wikitext templates</td>
<td>EXIF metadata (?)</td>
</tr>
<tr>
<td>Uploader user page</td>
<td>Non-notable contributors (smart URI)</td>
</tr>
</tbody>
</table>

- Notable contributors
- Copyright and licensing
- Things "depicted"
- Dates (datatype date/time)
- Geo location (datatype coordinate)

---

### Wikidata

- Notable contributors (items/properties)
- Copyright, licensing (items/properties)
- Things "depicted" (items/properties)
- Dates (properties)
- Geo location (properties)

---

Smart URI (new type of link to Wikimedia user page, Flickr user page, ...)

Uses Wikidata items/properties via federation
Upcoming

- **A first feature**
  - Multilingual captions (Summer 2018)

- **Full structured data rollout**
  - 'Depicts' in Autumn 2018

- **Search**

- **Upload**

- **New data modelling**
  - by Commons and Wikidata communities
    - What is depicted in media
    - Contributors
    - Copyright and licensing
Commons: Structured data

Newsletter

Phabricator board

Community focus group
SDC and GLAM
Research

- Focus group @ 2017 European GLAMwiki coordinators meeting
- Interviews w/ GLAMers from 5 continents
- Survey through GLAM monthly newsletter

meta:Research:GLAM and Commons
Example 1:

GLAMs and Wikimedians need good documentation

→ Embark on GLAM pilot projects, starting April 2018

- to be well documented
Example 2: Data wrangling before uploads is hard!
→ Work on mapping GLAM metadata schemes to Wikidata and Commons
What does SDC make possible for new GLAM-Wiki projects?
New ways to search Commons and to discover interesting files
Much better crowdsourcing ...

... with data that can be given back to an institution.

Rijksmuseum – Naturalis Biodiversity Center – Wikimedia Nederland & COMMIT/ SEALINCMedia present a unique birdwatching event

Birds are everywhere. In your own garden, in nature, and also in art. Among the Rijksmuseum’s 1.2 million collection objects are many prints, paintings and artefacts that have bird species depicted on them. Among the 37 million objects in the Naturalis collection are many birds from all over the world that have been collected in the last 200 years, as well as historical drawings of plants and animals in which many birds are depicted.

Wikimedia Commons

Some of the depicted birds are easily identified. Others require a trained eye to determine which species the artist has pictured. The Rijksmuseum and Naturalis are looking for experienced bird watchers and other avian enthusiasts to join an expedition through their digital collections and help the museums identify bird species in works of art.

The Rijksmuseum and Naturalis are currently in the process of donating large parts of their digitized collections of bird images to Wikimedia Commons, Wikipedia’s open multimedia library. Participants of the birdwatching day are challenged to collaboratively identify as many species as possible in this unique opportunity to contribute to the historical record of biodiversity.

Birdwatching Rijksmuseum

On Sunday October 4th, World Animal Day, the Rijksmuseum will host a birdwatching day in collaboration with Naturalis Biodiversity Center, Wikimedia Netherlands and the COMMIT SEALINCMedia project.
More advanced tools (based on a much improved Commons API)
Bringing together information that has never been combined before (across heritage collections!)
New types of campaigns

Ajapaik rephoto application displays historic images taken near to you and offers an easy way to take rephotographs of the same views.

In order to make the rephotography more convenient, you can change the transparency of the historic image on your camera stream by sliding along the
More value and potential of Commons' media files beyond Wikipedia
Who already uses Wikidata in GLAM projects?
How and why?
Workshop time!
List your affiliate's upcoming GLAM + Commons projects

1. Do this **individually**
2. In this spreadsheet:
   
   **https://is.gd/glamcommons**

3. 10 minutes maximum
What to do:

only add your info in the red and blue columns (for now)
Project highlights and issues

1. Do this **in pairs**
2. In the same spreadsheet: [https://is.gd/glamcommons](https://is.gd/glamcommons)
3. 10 minutes maximum

**What to do:**

Write info about your neighbor's project in the green area.
3 x 15 minutes...

1. **Opportunities** of SDC
2. New working **processes** with SDC: what do you need?
3. **Ideas** for (new projects) with SDC
Changes, needs, ideas!

1. 4 or 5 larger groups
   a. Similar interests
   b. Mix of beginners and more experienced people
   c. Assign a note-taker and a reporter

2. 15 minutes per question + break - max. 60 min total
   a. Alex, Ben, Liam, Pam, Sandra keep time and structure

Notes here:
https://etherpad.wikimedia.org/p/sdc-pilots
Process

1. **Discuss** the questions **together** (take good notes!)
2. Remaining ideas, issues? **Post-it notes**

Arthur Fry, inventor of the Post-it Note, with one on his forehead bearing a picture of a lightbulb. Photo by Tinkeringbell, Public Domain
https://commons.wikimedia.org/wiki/File:Fry-lightbulb-on-forehead1.jpg
Reporters!

1 point per 'block' that stood out (3 in total)

Town crier on the island of Terschelling, the Netherlands. Nationaal Archief / Spaarnestad Photo; photographer unknown. No known copyright restrictions, from Flickr the Commons. https://commons.wikimedia.org/wiki/File:Dorpsomroeper_-_Town_crier_(3280639091).jpg
Opportunities (15 mins)

- Does SDC help reach new audiences? How?
- Does SDC help you to work with new partners, on new collaborations? How?
- Which other opportunities does SDC bring?

Think of examples from your own projects.

1. Discuss the questions together (take good notes!)
2. Remaining ideas, issues? Post-it notes
### Process and needs (15 mins)

- **People** involved in your programs (Wikimedians, affiliate members/staff, GLAM staff…)
  - Are they interested in structured data?
  - What do they need to get 'ready' for such projects?
- **Your working process** for GLAM-Wiki projects
  - How will this change?
  - Do you need help? In which areas?

1. **Discuss** the questions **together** (take good notes!)
2. Remaining ideas, issues? **Post-it notes**
Grab a snack or a drink!
Ideas brainstorm (15 mins)

- Which of your own planned projects seem most ready for SDC?
- Which other (new) projects can you imagine with SDC?
  - By your own affiliate
  - Together with other affiliates

1. Discuss the questions together (take good notes!)
2. Remaining ideas, issues? Post-it notes
Town crier on the island of Terschelling, the Netherlands. Nationaal Archief / Spaarnestad Photo; photographer unknown. No known copyright restrictions, from Flickr the Commons. https://commons.wikimedia.org/wiki/File:Dorpsomroeper_-_Town-crier_(3280639091).jpg
Back to your own planned projects...
New insights on SDC applications

1. Do this **individually**
2. In the same spreadsheet: [https://is.gd/glamcommons](https://is.gd/glamcommons)
3. 10 minutes maximum
What to do:

write new thoughts about your project in the white area

write your own conclusion for follow-up in the yellow column!
Conclusion
Who thinks they already have a quite clear idea for an implementable pilot project?

Let's follow up! □
What are your main takeaways?
Thank you!