

Roundtripping: Returning Commons community metadata additions and corrections to source

Version 1.0, David Haskiya & Albin Larsson, after initial review and feedback from the Wikimedia Foundation. CC-BY - the Swedish National Heritage Board.

1. Project direction

1.1. Background

Across the world galleries, libraries, archives and museums (GLAMs) have chosen to share media files from their collections on Wikimedia Commons. Among them are many Swedish GLAMs¹.

Once included in Wikipedia articles these media files are viewed, played and downloaded by more users than ever even visit most GLAMs websites^{2 3}. Those same users, Wiki community members, are not only free, but invited and encouraged to add their knowledge to the metadata and descriptions of those files on Wikimedia Commons. This can take many forms: corrected spelling mistakes, translations into other languages, categorisations, suggested corrections to errors, and new information are among them.

With very very few exceptions, those additions remain on Wikimedia Commons and are not returned to the original source collections database of the GLAM. This means that quality improvements remain only on the Wikimedia platforms and are not available in the source collections databases themselves. The project posits that this is a significant opportunity loss for the GLAMs participating in GLAMwiki collaboration and who's core mission typically includes researching and improving their own collections.

This project is partially funded by Europeana Common Culture.

¹ <https://ambrosiani.github.io/Swedish-GLAMwiki-stats/>

²

<http://www.catchingtherain.com/portfolio/use-and-impact-of-cultural-heritage-images-on-wikimedia-commons-and-wikipedia/>

³ <https://www.metmuseum.org/blogs/now-at-the-met/2018/open-access-at-the-met-year-one>

1.2. Purpose and impact

The purpose of this project is to research, design and prototype technical solutions that would make it much easier and less work intensive for GLAM collections managers to review, copy and add to source additions to the metadata of the media files they have shared on Commons. We believe this will make it:

- A. even more attractive for GLAMs to share media on Commons in the first place - they know it will result not only in increased usage but increased data quality for their collection
- B. make it even more attractive for Wikivolunteers to add their knowledge to them once uploaded - they know their contributions will be valued
- C. in the long-term improve the quality of GLAM-collections who choose to engage with the Wiki-community
- D. contribute to an increased appreciation of the value of GLAM-collections and GLAMs

1.3. Goals

- Research the desirability and requirements of GLAM-collections managers in regards to retrieving metadata added to their files post-upload on Wikimedia Commons
- Develop and test a prototype tool that supports GLAM-collections managers in identifying, reviewing and retrieving added or changed metadata to media files.
- Report on lessons learned and recommend future actions (or inaction).

1.4. Expected results

A "by doing and prototyping" - *proven insight into the desirability (for GLAMs) and feasibility (for the GLAMwiki developer community in general and the National Heritage Board in particular) of developing tool support for reviewing and retrieving Commons community created additions and corrections to GLAM-provided metadata.*

Part of the feasibility analysis will entail an approximate cost estimation of developing and maintaining a production ready tool.

2. Constraints

2.1. Extent and constraints

- The project itself will not develop any software that is tailor-made for any specific collection management system unless such a system is open source.
- The project will not develop a tool for viewing or download statistics for media files uploaded to Commons by GLAMs. This is a practical constraint to keep the project scope focused.

- The policy of the National Heritage Board sets a constraint on how software development companies, design companies and individual freelancers can be contracted.
- The project will not broach the subject of metadata licensing.

2.2. Time constraints

The project runs November 5 2018 - June 20 2019.

2.3. Budgetary constraints

We expect to be able to dedicate a minimum of 200 000 SEK and upwards to 400 000 SEK to buy consultancy services - primarily software development and UX design.

2.4. Prioritisations

We are aiming for a prototype that we can learn from - a **Minimal Learnable Product**. We are not aiming for a production ready tool.

All participants intend to contribute to an open process, open source, and accessible results.

2.5. Organisation

2.5.1 Project owner

David Haskiya, Swedish National Heritage Board, Head of Unit

2.5.2 Team

Albin Larsson, Swedish National Heritage Board, Lead Developer and Product Owner

Susanna Ånäs, Project Administrator (contractor)

Maarten Zeinstra, UX Researcher (contractor)

Paweł Marynowski, Software Developer (contractor)

Petter Johannisson, Swedish National Heritage Board, Interaction Designer, may join the project in 2019.

2.5.3 Participating organisations

Swedish National Heritage Board (Riksantikvarieämbetet)

The Swedish Performing Arts Agency (Musikverket)

The Nordic Museum (Nordiska museet)

Wikimedia Sverige

Wikimedia Foundation

The German Digital Library may join the project in 2019.

2.5.4 Reference persons and experts to consult

A mix of active GLAM-wikipedians and GLAM professionals who work in practice with collections management and development. Some members of the reference group may also be suitable test users. This group should be no more than 5 persons.

2.6. Dependencies and related work

There's a technical dependency with Wikimedia Foundation's work to implement structured data on Wikimedia Commons through Wikibase. In order not to be outdated from the start the project will require access to a fairly mature test version of the Wikimedia Commons Wikibase-APIs.

The FindingGLAMs project run by Wikimedia Sverige plans to do upload case studies that tackle challenging content or processes. Roundtripping may be one of those case studies.

3. Security

No security concerns have been identified.

4. Hand-over

The results of the project will in effect be handed over to the GLAMwiki community. Their attitude and willingness will decide whether to develop a production ready tool or not.

All project results - reports, source code, presentations, etc. - will be publicly available under open licencing.