

Project Plan: Wikimedia Commons Data Roundtripping

Version 1.0, Albin Larsson, RAÄ-2018-3594

1. The project's focus and scope	3
1.1. Background	3
1.2. Purpose	3
1.3. Outcome targets	4
1.4. Project Goals	4
1.5. Deliveries	4
1.6. Project sponsor	4
1.7. Project basis	5
1.8. Success factors	5
1.9. Scope and constraints	5
1.10 Dependencies and related work	5
2. Execution	6
2.1. Description of execution	6
2.2. Methodology	6
2.3. Decision points	6
2.4. Milestones	6
2.5. Timeline	6
2.6. Project audits	7
2.7. Project documentation	7
2.8. Meeting plan	7
2.9. Communication plan	7
2.10 Security	7
3 Resources	7
3.1 Economy	7
3.2 Resources	8
3.3 Administrative Data	8
4 Organization and stakeholders	8
4.1 Team	8
4.2 Stakeholders and partners	8
5 Risk management	9
5.1 Risk analysis and risk management	9
6 Handover and definition of done	9
6.1 Handover of results	9
6.2 Management	9
6.3 Definition of done	9
7 Evaluation	10
References	10
8.1 Steering and supporting documents	10
8.2 Glossary	10

1. The project's focus and scope

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, Wikimedia 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 opportunity loss for the GLAMs participating in GLAMwiki collaboration and who's core mission typically includes researching and improving their own collections.

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

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. Outcome targets

The long term effect that this project aims to achieve is a museum sector which is more conscious and knowledgeable about good practices for making their digital collections available on the web for communities and users to enrich. By receiving the result of this project (reports/prototype) they get enhanced knowledge about crowdsourcing and the open ecosystems on the web.

1.4. Project 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).

1.5. Deliveries

- Initial research report that reports on the interviews and questionnaires
- Midtime report that reflects the project progress and its challenges
- Final report that discusses and reports the results

1.6. Project sponsor

David Haskiya, Swedish National Heritage Board, Head of Unit

1.7. Project basis

Since the government proposition 2016/17:116 the Swedish National Heritage Board is responsible for offering technical and methodology support to museums⁴.

The Wikimedia Foundations work towards structured data on Wikimedia Commons enables the required technical needs for this project⁵.

This project is partly funded by the Europeana Common Culture project.

1.8. Success factors

Communication both internal to handle the geographically divided team as well as with external partners to handle dependencies to other work.

Close corporation with museums and museum personnel.

1.9. Scope 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.

⁴

<https://www.regeringen.se/4933fd/contentassets/127b80d33b084194a415d72b85721874/161711600web.pdf>

⁵ https://commons.wikimedia.org/wiki/Commons:Structured_data

1.10 Dependencies and related work

There's a dependency with Wikimedia Foundation's work to implement structured data on Wikimedia Commons through Wikibase.

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.

Wikimedia DC is currently executing a project related to crowdsourcing depicts statements for Wikimedia Commons/Wikidata. This could potentially be a resource if we decide to test the concept around campaigns.

2. Execution

2.1. Description of execution

See section 2.5.

2.2. Methodology

The project consists of two main activities, an initial research part and one prototyping/testing part. The initial research will be done through a questionnaire and interviews. The prototyping/testing part will be done through user testing and common design methodology practises (based on LEAN and Design Sprints).

2.3. Decision points

These decisions points are always carried through by the project sponsor.

- Accepting the project plan and timeline
- Accepting the midterm report
- Accepting the final report

2.4. Milestones

- The project plan and timeline is accepted, November 2018
- The midterm report is accepted, Last of February 2019
- The initial research report is done, Last of February 2019
- The final report accepted, June 20 2019

2.5. Timeline

See [Timeline - Roundtripping](#).

2.6. Project audits

The goal of the midterm report is to audit the progress of the project and potential process issues or blockers.

See also 2.3

2.7. Project documentation

The project documentation will be registered under the registration number RAÄ-2018-3594. Core documents are also published under CC-BY 4.0 on Wikimedia Commons.

2.8. Meeting plan

The team intends to meet during multiple 2-3 day meetings where the aim is to test the prototype and interview relevant museum personnel together.

2.9. Communication plan

Ongoing communication is done through the This Month in GLAM Newsletter, the Special Story section can be used when needed otherwise the Sweden Report section can be used. The Project Brief, Project Plan, Midterm Report, and Final Report should be published on Wikimedia Commons and shared through channels of the teams choose. Relevant events to present this work and/or its results should be tracked in a separate document.

2.10 Security

No reasons to classify any documents has been identified. Personal data will be handled according to the Swedish implementation of GDPR.

3 Resources

3.1 Economy

All numeric values are in Swedish krona (SEK).

Budget subject	Sum	Comment
Contractors	300 000	
Trips internal	30 000	Trips for RAÄ- employees
Trips external	60 000	Trips for non-RAÄ employees (contractors etc)

The budget may be expanded, if proven necessary, by the Project Sponsor.

3.2 Resources

See 4.1 and 1.8.

3.3 Administrative Data

Administrative Data	
Activity	Metodutveckling, digital förmedling
Cost code (Kostnadsställe)	640
Project registration number	RAÄ-2018-3594
Date	2018-11-05 - 2019-06-20

4 Organization and stakeholders

4.1 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.

4.2 Stakeholders and partners

The Wikimedia Foundation

Wikimedia Sverige

5 Risk management

5.1 Risk analysis and risk management

Risk	Probability	Impact	Management
Geographically divided team leads to inefficient cooperation	4	5	Common Slack instance and the ability to collaborate in real time using Google Drive
High technical dependency on the Wikimedia foundation blocks development	4	4	Ongoing communication with the WMF and initial prototyping against Wikidata.
Geographically divided team makes it hard to have joint meetings	4	3	Early Doodles and the ability to have meetings without all team members attending.

6 Handover and definition of done

6.1 Handover of results

The results are accepted by the project sponsor.

6.2 Management

The project result will be shared as defined by the Communication plan. Reports and major documents will be released under CC-BY on Wikimedia Commons. Source code will be made available under MIT on Github. No maintenance of documents or software is planned. If the team finds that the Swedish National Heritage Board's Github is not suitable for hosting the source code they can transfer it to another host, during or after the project.

6.3 Definition of done

- The deliverables is accepted by the project commissioner
- All major documents, code, and contracts are registered under RAÄ-2018-3594
- All final source code is published on Github
- All final documentation is available on Wikimedia Commons

7 Evaluation

The team and stakeholders will do a retrospective and the outcomes will be documented and shared internally. If there is a clear value in sharing the process and the things learned from the process it will be done in a public forum.

References

8.1 Steering and supporting documents

[Project Brief: Wikimedia Commons Data Roundtripping](#)

8.2 Glossary

GLAM - Galleries, Libraries, Archives, and Museums

GLAMwiki - GLAMs (as defined above) work with the Wikimedia movement and its platforms

Commons - the Wikimedia Commons platform

WMF - Wikimedia Foundation