Project Initiation Document: GLAM-wiki toolset project

Author: David Haskiya

Date: 5 May 2012

Status: 1.0 (Reviewed by the Steering Group)

Summary: This document describes the high-level business logic and implementation of a **joint Wikimedia and Europeana project**. The purpose of the project is to **develop a scalable and maintainable system for mass uploading (open) content from galleries, libraries, archives and museums (henceforth GLAMs) to Wikimedia Commons** and to **create GLAM-specific requirements for usage statistics.** A comprehensive infrastructure does not currently exist, making it difficult for willing GLAMs to open up their content via Wikipedia. Opening up access to Cultural Heritage is a strategic aim of both Wikimedia and Europeana. Europeana has experience in setting standards and has access to a network of GLAM partners and is therefore well positioned to facilitate such an infrastructure in close cooperation with Wikimedia.

The project is a **partnership** between the Wikimedia Foundation, Wikimedia Netherlands, Wikimedia UK, Wikimedia France and the Europeana Foundation. Further Wikimedia Chapters and other funding organisations will be invited to support the project. Formally and financially the project will take the form of **grants** from the Wikimedia chapters to the Europeana Foundation for the purpose of running this project.

The estimated budget is **a maximum of 400 000 euros over 24 months incl. VAT**, with a project **start on June 1 2012**. Funding for the first 12 months of the project is secured.

The scope of this document is to provide enough information for the project stakeholders to decide to initialize the project or not. It represents the current understanding by Europeana of the project scope and is to be understood as a pre-cursor to a project plan but not the project plan itself.

Background

The Wikimedia Foundation and Wikipedia community are seeking collaboration with GLAMs for both media and editorial content. Some major hurdles for GLAMs who want to share their collections with Wikimedia Commons are:

1. The lack of user friendly tools for (mass) uploads of content to Wikimedia Commons
2. The lack of support for typical GLAM-metadata standards in Wikimedia Commons templates
3. The lack of user friendly tools to export metadata and data back to GLAMs when it has been improved by the Wikipedia Community
4. The lack of easily accessible and fit for purpose usage statistics for GLAM-content once uploaded to Wikimedia Commons

For all of the hurdles listed there are a number of existing tools and solutions. They overlap in purpose, are scripts without user interfaces, are often specific to uploading the content of specific institutions and only their creators usually can use them. Statistics tools exist, but are unstable and do not necessarily meet the requirements of the GLAMs.

This makes GLAMs entirely dependent on individual Wikipedians (volunteers) to create upload scripts and perform uploads. This has resulted in a long waiting list of GLAMs who want to share content on Wikimedia Commons, but who can't. To scale Wikimedia's GLAM-outreach, uploading content to Wikimedia Commons needs to be simple enough for GLAM collection managers to do themselves.

The Wikimedia community contacted Europeana to suggest a partnership, together with a group of Wikimedia Chapters, to develop a user-friendly and extensible system for uploading GLAM-content (data and metadata) to Wikimedia Commons, to re-export content from Wikimedia Commons back to the providers and to create clear GLAM-specific requirements for improved usage statistics.

A meeting was held at Wikimania 2011. The purpose was to agree on the rough scope of a possible project and how it could be financed. An outcome of the meeting was the decision that Europeana would create a Project Initiation Document (PID). Based on the PID the stakeholders would make a decision to launch the project

Project overview

Business goals

The Wikimedia community is dedicated to encouraging the growth, development and distribution of free, multilingual content, and to providing the full content of these wiki-based projects to the public free of charge. It supports active participation in Wikimedia Projects through independent national Wikimedia Chapters. The community is seeking collaboration with GLAMs concerning both media content and editorial content.

The primary goal of Europeana is to increase access to Europe's cultural and scientific heritage online. A partnership with the Wikimedia community with the purpose of improving and simplifying the way GLAMs share content, information and knowledge with the Wikimedia projects will facilitate increased access to Europe's digitised heritage, position Europeana as a facilitator of innovation in the heritage

sector, and make it possible also for Europeana to distribute content to the Wikimedia Projects. Europeana also wishes to have stronger brand awareness for its facilitation work among GLAMs.

Project deliverables

This project will follow the Agile principle of "Working software over (technical and project) documentation". Project artefacts will be kept at a minimum.

The team will use SCRUM as its development methodology. SCRUM is used in all Europeana development projects and the team is committed to continual improvement.

Deliverable 1: A GLAM Upload system

Content uploading from GLAMs to Wikimedia is currently a manual, tedious and difficult process. To assist and encourage GLAMs to share content with Wikimedia Commons an upload system will be developed that will:
- Support the major metadata standards and their mappings to Wikimedia Commons templates
- Allow user friendly creation of mappings between GLAM metadata templates to Wikimedia Commons templates
- Allow user friendly creation of partnership templates for proper attribution
- Allow GLAMs uploading content to Wikimedia Commons to preview them prior to final upload and publication
- Have a plug-in architecture allowing third-party developers to create provider API and format specific plug-ins or plug-ins for specific processes e.g. public domain checking or file transformation
- Allow for batch export and download of specific GLAM-collections from Wikimedia Commons with metadata output in standard GLAM metadata formats
- Be able to monitor updates and changes in the uploaded content and metadata and download the changes for re-integration at source
- Have the necessary hooks to communicate with the MediaWiki backend for usage statistics/data analytics
- The system web pages will contain clear attribution to Europeana
- The project aims to conform to the best practices and standards for development of MediaWiki as defined by the Wikimedia community

Deliverable 2: Report on requirements for usage and re-usage statistics for GLAM content

Existing tools for GLAM metrics and statistics don't fulfil all the needs for metrics from GLAMs. As part of this project a full report will be written on GLAM needs of usage statistics and the desired tool support for those needs. The report will be based on:

- Curators and collection managers will interviewed to form the basis of business requirements
- Wikipedians who have liaised with GLAMs will be interviewed to form the basis of business requirements
- The Wikimedia Foundation Data Analytics team will be interviewed to get a technical understanding of the affordances offered by the statistics back-end
- Existing tools for statistics will be evaluated
- A mock-up prototype of a GLAM usage statistics dashboard will be developed

Project artefacts/Milestone documents

- Project plan, Month1
- Interim review, Month 3
- Product Design and Development plan, Month 3
- Community outreach plan, Month 3
- Major project review, Month 6
- GLAM-metrics and statistics requirements report, Month 9
- Major review, Month 9
  - At this milestone a decision to continue with the project's second year will be taken by the Steering Group
- First year/interim report, Month 12
- Interim review, Month 15
- Interim review, Month 18
- Interim review, Month 21
- Project report, Month 24

Out of scope

Tasks or products that are out of scope for this project are:

- A working GLAM-usage statistics report system
  - This will be developed by the Wikimedia Foundation Data Analytics teams
- Leading the final integration of the GLAM upload system into the Wikimedia Foundations technological infrastructure
  - This will be performed by the Wikimedia Foundation, but can be assisted by the project in the implementation phase

Project scope may be changed by the Steering Group.

Work packages

The project will comprise three work packages focusing on the three strands of the project: management, development and community outreach.

WP1: Project management

The main purpose of this work package is project management and business requirements analysis.

Tasks and deliverables:

- Reporting project progress and deviations to the Steering Group
- Risk management
- Acting as the Product Owner by proxy on behalf of the Steering Group
- Creating and maintaining the Product backlog
- Coordinating acceptance testing
- Ensuring the integrity of the Europeana and Wikimedia brands
- Creating the report on the requirements for usage and re-usage statistics for GLAM content

Work package leader: the Project Manager.

WP2: Development

The main purpose of this work package is to develop the GLAMwiki system.

Work package tasks include:

- Implementing and adapting a Scrum-based development methodology
- Configuring the development and test environments
  - Using Wikimedia Labs as the development framework
- Setting code standards and ensuring code quality
  - In accordance to existing best practices and technical guidelines
- Defining the system architecture
  - With special focus on a plug-in architecture
- Liaising with community developers who wish to contribute to the development

Work package leader: The Lead Developer

WP3: Community outreach

The purpose of this work package is to ensure maximum uptake of the developed toolset among GLAMs and to ensure continued support of the project by the Wikimedia Commons community.

Work package tasks include:

- Planning and setting up the forms and media of the community outreach

  - Including a plan for representation at relevant events

- Liaising with GLAMs who are contributing content to Wikimedia Commons or are interested in contributing content to Wikimedia Commons

  - Recruiting pilot GLAMs and testers of the system

- Liaising with the GLAM-wiki community

  - Recruiting project ambassadors

- Outreach to the Wikimedia Commons community

  - Build a network of volunteers within the Commons community

Work package leader: Europeana's Business Developer for GLAM-relations

Project phases

The overall project is 24 months beginning June 1st 2012.

Project start

June 1st, 2012

The project will start at the Berlin hackathon.

Planning and prototyping phase

3 months, 3-4 sprints

The official project kick-off is tentatively planned to take place at a European MediaWiki hackathon in April.

Development, phase I

7 months, 9-10 sprints

The first development phase will result in a beta version of the tool that is functional and stable enough for actual use by a number of closed beta GLAMs or as a completely public beta.

Development, phase II

9 months, 10-12 sprints,

The second development phase will incrementally improve the beta bringing it to full production step by step. If scope permits specific "nice to have" features and plug-ins can be developed.

If a decision is taken at the end of 2012 to not continue with the project beyond March 2013 this phase will not be initiated and instead the project will enter the Implementation phase.

Implementation phase

Last 3 months of the project, 4-5 sprints

While testing (both technical and user testing) will be integrated into the regular development sprints the implementation phase will comprise intensive testing in the form of a public beta period and technical quality controls to make the system ready for full production in a first version.

End of project

Project organisation

Steering Group

The project is governed by a Steering Group with representatives from the project partners and main stakeholders.

Steering Group main tasks:

- Provide high-level guidance on business priorities and budget to the Project manager
- Set scope limits for the project
- Accept or reject milestone documents/reports, thus initiating the project itself and its progression from quarter to quarter
- Attend Sprint demos

Steering Group members:

- Wikimedia Netherlands: Maarten Brinkerink
- Wikimedia UK: Ashley: Ashley Van Haeften
- Wikimedia France: Christophe Henner
- Wikimedia Foundation: Diederik van Liere
- Europeana Foundation: Jan Molendijk
- GLAM-wiki community representative: Maarten Dammers
- GLAM community representative: To be decided

The Steering Group will elect a chair at the first meeting.

Project team

The team will be cross-functional and work as a unit. However, the different developers will have specific areas of focus and will need to have the necessary experience and complementary skill-sets. Work efforts are given in full-time equivalents (FTEs).

Lead developer
This developer is an experienced PHP-developer used to develop to industrial standards, perform configuration management and code quality control. He or she should ideally have experience in developing for the MediaWiki platform.

Senior Developer and GLAM metrics analyst
This developer will focus on functionality specifications, functionality control and definition of GLAM-specific usage metrics and reporting. He will have experience in developing for the MediaWiki platform, knowledge of open licensing standards and have experience working within the GLAM-sector. This developer will be a part-time contractor.

Intern developer
During the project lifetime intern developers will be invited to contribute to the project.

Information Architect and Senior Developer
The primary responsibility of the Information Architect will be to model the information flow and implement the various mapping stages of uploading and downloading metadata. This developer has played a similar role in Europeana's own data ingestion flow. This developer is part of the Europeana core Development team.

Metadata specialist
The metadata specialist will provide expertise on standard metadata formats used within the GLAM-sector and on best practices for metadata mapping. This specialist is part of the Europeana Operations team.

Project manager
The Project manager ensures that the project progresses according to plan, scope and budget set down in the project plan and by the Steering Group. Further, the Project manager monitors and mitigates risks and communicates them to the Steering Group. The Project manager is part of the Europeana Business Development team.

Community manager
The Community manager plans and coordinates project outreach to the GLAM-community and Wikipedia community. The Community manager is part of the Europeana Business Development team.

Budget

A maximum of 400 000 euros over 24 months.

Budget

Role	Hourly rate	# MM	Total hours	Total amount	VAT	TOTAL
Project Manager	49,00	2,40	312	15.288	2.905	18.193
Senior Developer	36,00	4,80	624	22.464	4.268	26.732
Front-end developer	49,00	2,40	312	15.288	2.905	18.193
Metadata specialist	33,00	2,40	312	10.296	1.956	12.252
Community manager	45,00	1,20	156	7.020	1.334	8.354
Senior Developer/GLAM Analyst	60,00	4,80	806	48.384	9.193	57.577
Lead developer	70,00	19,20	2.496	174.720	33.197	207.917
Total personnel costs		37,20	5.018	293.460	55.757	349.217
Travel expenses				10.000	-	10.000
Marketing materials				4.500	855	5.355
Overhead coverage 10%				29.300	5.567	34.867
Total budget				**337.260**	**62.179**	**399.439**

Please note that the intern/Junior Developer will not be paid out of the project budget and so does not appear in the budget.

Payment schedule

As per the Grant Agreements. To be maintained through a master schedule on behalf of the Steering Group.

Appendix: Project Risks Register

The project risk register will be maintained by the Project manager throughout the project.

Risk description	Probability	Impact	Mitigation
Corporate culture differences between Wikimedia and Europeana.	High	Medium	Assume good faith. Open and clear communication. Devise Win-Win scenarios.
Unclear scope leads to conflicting opinions on whether definition of done has been reached or not in the Steering Group	Medium	High	Clarify scope between the sign-off of the PID and the creation of the Project plan. Begin the project with a 3-4 sprint phase that produce requirements, data flow charts, interaction wireframes, conceptual designs etc. in order to properly define the scope. Assume good faith. Open and clear communication.
Cannibalisation of Europeana. All European GLAMs upload to Wikimedia Commons but don't contribute to Europeana.	Medium	High	Europeana plug-in to the GLAM-upload tool as part of the project. Europeana metadata schemas a supported in-data format.
GLAM fear of loss of attribution leading to reticence to contribute content.	High	High	Clearly branded templates with links to original source in both Wikimedia Commons and Wikipedia article displays as part of the customisation a GLAM-officer can do.
GLAM fear of loss of control over content rights.	High	High	Agreed upon content usage guidelines. Authorisation for batch upload via provider specific plug-ins.
Europeana lost in the more high profile Wikipedia brand.	Medium	Medium	Dual branding of the tool and external communications. Official strategic partnership between Wikimedia and Europeana.
News of a Europeana – Wikimedia partnership disturbing Europeana's	Medium	Medium	Clearly communicate that this is not a Europeana network specific project. Clearly communicate that Europeana partner content

negotiations with partners concerning licensing all Europeana metadata as CC0.			contribution to Wikimedia is optional.