Project Acronym: Europeana v2
Grant Agreement number: 270902
Project Title: Europeana Version 2

D6.2: Release Report

Revision	[1.1]
Date of submission	30 September 2013
Author(s)	Pavel Kats - The Europeana Foundation
Dissemination Level	[Public]

Project co-funded by the European Commission within the ICT Policy Support Programme

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1.0	30 September 2013	Pavel Kats	The Europeana Foundation	Author
1.1	30 September 2013	David Haskiya	The Europeana Foundation	Remarks

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Introduction

The main theme of Europeana development during the reporting period (October 2012 – September 2013) is deployment into production and stabilization of version 2.0 of Europeana software. This release, which was partially developed during the previous reporting period, was about a major overhaul of all the components of the Europeana software stack – portal, API, dataset, and ingestion system – to become compliant with the Europeana Data Model (EDM).

The deployment was done gradually during the second quarter of 2013 and was followed by an intense phase of bug fixing and performance stabilization. The amount of new features introduced during this period was limited but many structural improvements were introduced into the code during the stabilization phase.

In parallel, some improvements have been introduced into the development process itself. They were necessary to sustain further development of Europeana software in the future, when increasing user base, growing dataset, and new amendments to the ingestion system will require a more stable and flexible development process.

Europeana Portal/API Changes

Europeana Portal and API were stabilized to support searching and rendering EDM data objects under growing amount of traffic to the portal website. Profiling tools were employed to prepare detailed breakdown of CPU and memory consumption during generation of web pages. Based on this analysis the code generating web pages was optimized for a single display and also to support parallel requests.

To better support search engine bots and to prevent degradation of performance of the website for human users as a result of massive traffic generated by these bots, a dedicated web server instance was allocated to handle the bots' traffic.

Many improvements were done in the software code itself. These improvements were required to mitigate technical debt, accumulated during intensive development phases under time pressure. This debt manifested mainly in lack of unit- and integration testing, lack of documentation, and inflexible structuring of some of the core components.

All in all significant efforts were invested in optimization and refactoring of the portal software so that it would be both more scalable and more flexible to easily allow software extensions by the main team and by external contributors.

Ingestion Improvements

Europeana ingestion infrastructure – the Unified Ingestion Manager (UIM) – required considerable stabilization work because it runs under high performance load. Everyday ingestion activities run constantly by the ingestion team, sometimes involving massive datasets of more than a million data objects, consume substantial amount of memory and CPU cycles. Additionally, since very often more than one ingestion process is run in parallel, there are special requirements for high parallelism between the UIM internal components. The system was stabilized by identifying and eliminating performance bottlenecks, sections of code allowing for insufficient parallelism and memory optimization.

The version 2.0 of the UIM contains three external components –(1) a CRM system which is a customized version of the open-source SugarCRM project, (2) a metadata mapping tool MINT developed by the National Technical University of Athens, and (3) data aggregator REPOX developed by the Instituto Superior Técnico in Lisbon. The interaction between these components and core UIM services often involves transfer of data and hence becomes prone to performance issues when large volumes are involved. Stabilization of these interfaces often required investigation by both Europeana and the team responsible for the product. This work resulted either in performance or functional improvements on one side or altering the interface protocol.

Additional work was required to stabilize the enrichment process, which is one of the core components of the system. Here required changes were revealed gradually, as

more datasets were being ingested, and new variations of metadata structure and values were encountered. The changes required both fixes in the enrichment component itself and the configuration of the underlying indexing database (Apache SOLR).

Process Improvements

The Europeana development team has started introducing significant structural changes to the development, testing, and release processes. These changes are necessary to support key components of the new strategic vision of the organisation: a shift from an end-user products supplier to a platform supplier, a shift from a team-based development to network-based development and a growing involvement of open-source communities. To support all this the team must further enforce agile approaches to development by introducing continuous mode of delivery of products. Continuous delivery will allow both more frequent releases and easier integration of external contributions.

To facilitate continuous delivery many chains of the delivery process have been automated. Building and deploying the software for the testing and acceptance environments is done automatically by a continuous integration open-source server. Further automation of testing of both portal and API is being in the works.

The source code management was moved from an internally maintained SubVersion system to the popular GitHub [repository](). This move both transfers the responsibility to manage the code from the team to a specialized external party and makes the code more visible, analyzable and better accessible to external contributors.

A centralized issue management system, [Assembla](), was successfully introduced to all Europeana development projects. The system allows better management of the development process due to higher control of development assignments, more transparency into resource allocation and better reporting abilities.

To sum up, considerable efforts and resource are being invested into improvement of development processes and practices in the team. The results of these efforts will have a delayed effect which will support changes in Europeana business strategy in the future.