

# Design, implementation and evaluation of a user generated content service for Europeana

N. Aloia<sup>1</sup>, C. Concordia<sup>1</sup>, A. M. van Gerwen<sup>2</sup>, P. Hansen<sup>3</sup>, M. Kuwahara<sup>4</sup>, A. T. Ly<sup>5</sup>, C. Meghini<sup>1</sup>, N. Spyratos<sup>5</sup>, T. Sugibuchi<sup>5</sup>, Y. Tanaka<sup>4</sup>, J. Yang<sup>5</sup>, N. Zeni<sup>1</sup>

<sup>1</sup> Istituto della Scienza e delle Tecnologie della Informazione, National Research Council, Pisa, Italy

<sup>2</sup> EDL Office, The Hague, The Netherlands

<sup>3</sup> Swedish Institute of Computer Science, Sweden

<sup>4</sup> MEME Media Lab., University of Hokkaido, Sapporo, Japan

<sup>5</sup> Laboratoire de Recherche in Informatique, University of Paris South, Orsay, France

## Abstract

*The paper presents an overview of the user generated content service that the ASSETS Best Practice Network is designing, implementing and evaluating with the user for Europeana, the European digital library. The service will allow Europeana users to contribute to the contents of the digital library in several different ways, such as uploading simple media objects along with their descriptions, annotating existing objects, or enriching existing descriptions. The user and the system requirements are outlined first, and used to derive the basic principles underlying the service. A conceptual model of the entities required for the realization of the service and a general sketch of the system architecture are also given, and used to illustrate the basic workflow of some important operations. The planning of the user evaluation is finally presented, aimed at validating the service before making it available to the final users.*

## 1 Introduction

In the 2011-2015 Strategic Plan, Europeana [1] announces User Engagement to be one of the strategic tracks by which the organization will deliver value. By the term ‘Engage’ Europeana refers to cultivating new ways for end user to participate in their cultural heritage. The Europeana network comprises communities of archivists, curators and librarians who show a growing interest in exploring new methods of access and dialogue. Europeana intends to enhance the user experience and offer services that allow users to interact and participate.

User-generated-Content (UGC) is one aspect of this renewed way of participating. Information about cultural heritage exists outside the heritage institutions; artifacts, written sources and memories of individuals complement collections held in institutions. UGC services are designed to provide users with means to support and interpret content. They will be involved in storytelling, curating of virtual exhibitions, reviews and even the creation of new collections. Greater participation will increase users’ interest and loyalty. Europeana is therefore devoting increasing resources to initiatives that bring out the value of the contribution those users can make.

In response to these needs, the ASSETS [2] Consortium has included the support of user-generated content amongst the services it is going to develop for Europeana. ASSETS is a two-year Best Practice Network co-funded by the CIP PSP Programme to improve the accessibility and usability of Europeana. The ASSETS Consortium comprises 24 partners, including institutions from ten different European countries and Japan, which are active in the field of cultural heritage and digital libraries.

Rather than focusing on a specific set of UGC applications, ASSETS is developing a general purpose, back end component that aims at supporting any UGC service Europeana will want to offer to its users. To this end, the ASSETS back end component implements an Application Programming Interface (API) for creating, storing and manipulating UGC Units of Work, and for submitting these Units of Work to Europeana, in the form of Europeana Submission Information Packages (SIPs). Final users will interact with their Units of Work through client interfaces, which will hide the unnecessary technical details and complexities of the back end to them, providing them with the level of

representation that is most suitable for the specific UGC task at hand. Indeed, it is expected that every UGC task will be supported by a different final user interface. But this will have no impact on Europeana, since every different front end will talk to Europeana through the same API.

The API will relieve future UGC applications from implementing any server side functionality; they will have to code only the client side, connecting the user world to the API. At the same time, the API will move away from Europeana the technical interoperability problems that would arise upon integrating into its database the possibly different objects coming from future UGC applications. The service will rely on the Europeana Data Model (being developed by the Europeana version 1.0 project [3]) in order to tackle the more serious semantic interoperability problems.

The definition of the conceptual model underlying the UGC API is the most difficult challenge that the ASSETS UGC team is facing. The model has to strike the optimal balance between simplicity, so to be quickly learned and easily coded against by the future UGC service developers, and generality, so to satisfy the needs of any possible future UGC service. This conceptual model has been defined during the first year of the ASSETS project, based on an analysis of the different types of requirements that are in place. The model has been subsequently used to define the UGC API, which is now being implemented to be demonstrated at the first year project review, due in May/June 2011. On top of this API, a few, initial UGC tasks will be implemented through specific front ends. These tasks will allow Europeana users to contribute to the contents of the digital library in several different ways, such as uploading new objects along with simple descriptions, annotating existing objects, or enriching existing descriptions.

The UGC service will be evaluated by the task developers and by the final users within the project lifetime. The former will evaluate the adequacy of the API, while the latter will evaluate the adequacy of the different front ends. The outcome of evaluation will lead to a refinement of both the back end component API and implementation, and of the front ends UGC applications.

The paper presents an overview of the UGC service. The user and the system requirements are outlined first, and used to derive the basic principles underlying the service. The conceptual model required for the definition of the API and a general sketch of the system architecture are also given, and used to illustrate the basic workflow of some important operations. The planning of the user evaluation is finally presented, aimed at validating the service before making it available to the final users.

## **2 Requirements**

### **2.1 User Requirements**

User requirements can take different forms:

- First of all, the submission of objects to a repository. Through a user interface (typically a website), users upload digital objects of various formats (images, audio and video). These must be objects that they actually hold rights over (for instance, by creating them). A minimal set of metadata will have to be provided in order to support the interpretation, discovery and management of the object. The user requires to be free of choosing which metadata format to use, but the system must propose a default one.
- Secondly, metadata enrichment. Users contribute factual metadata to an object, such as location, date, names, or tags. The object can be created by the user, but also by another user; the object may also existing content in Europeana.
- Thirdly, annotations: users are contributing their views, comments. opinions to an object
- And finally, there is contextualization: through storytelling or creating virtual exhibitions and galleries and possibly adding narratives to them, users are combining existing objects into a new context (without changing the objects and metadata itself).

Before publishing user generated content, moderation may be added as an intermediate step in the process. Authorized users review the UGC and decide to accept and publish it. In most cases, this includes a feedback loop to the user who originally contributed the data.

## 2.2 User Interfaces

Europeana audiences include academic researchers with a high level of language and computer skills but also people who are hardly familiar with foreign languages or using the internet. Typical target groups are secondary school students and their teachers as well as cultural explorers and travelers. While possessing intermediate to good knowledge of foreign languages and online search, these groups generally expect services to be easy and intuitive. At the same time, they want to understand what happens with their contribution and who keeps control over their content. Secondary school students are generally known for focusing on images rather than digesting long texts.

User Interfaces should therefore preferably be simple, straightforward and visual. Additionally, clear information must be provided about rights regarding the content.

## 2.3 System Requirements

The back end component has to comply with the Europeana architecture, which is based on an Open Source policy. Europeana has also defined a set of guidelines regarding the coding, the testing and the deployment of the components that make up its architecture. These guidelines are publicly accessible on the wiki of the EuropeanaLabs [4] and are being followed by the ASSETS UGC development team.

One important requirement concerns the management of multimedia content. As outlined in the previous section, Europeana expects users to contribute media files of various formats including text, images, videos and the like through its UGC service. However, Europeana was not initially meant to deal with (i.e., collect, store and make accessible) media files. With UGC, this has to change, of course. Moreover, the change has to happen gradually, by evolving the current architecture in a stepwise manner, so as not to compromise the operation of the existing components. In order to meet this requirement, ASSETS will provide a simple media object repository, and will initially deploy it on a separate server, the ASSETS server, to be later integrated into the Europeana architecture. More details on the media object repository are given in the architecture Section.

## 3 The conceptual model of the UGC service

In order to meet the user and system requirements, the ASSETS team designed a UGC service based on the concepts outlined below and presented in Figure 3.1 as a UML class diagram. In the diagram, boxes represent classes and arcs represent associations. Every association is bi-directional, but only the name in one direction is given for readability purposes. The name given refers to the association going from the class closest to the name to the other class. The reader can derive the name of the association in the opposite direction by following any notational convention.

From the UGC server point of view, at any point in time there exists a set of users of the UGC service. Each user is in fact a role, identified by an id and a password, behind which a whole community may actually operate. Each user has its own Workspace (WS) on the UGC server. A WS is simply a container of the objects that the associated user needs to perform UGC tasks.

The creation of a single UGC object may take a long time and span several sessions of work. In between one of these sessions and the next, the partial results achieved so far have to be persisted, in order not to be lost and to be resumed at the beginning of the next session. The concept of “partial” UGC is captured by the notion of Unit of Work (UoW). The UoWs of a user are maintained in the user’s WS.

A single UoW contains objects, identified by URIs, and their accompanying descriptions. The objects in a UoW can be of two kinds:

- Existing Europeana objects, that the user has included in the UoW in order to link them to new objects (see below) as values of some property, or in order to enrich them with new descriptions. Existing Europeana objects can be retrieved for inclusion in a UoW via a query issued to Europeana.

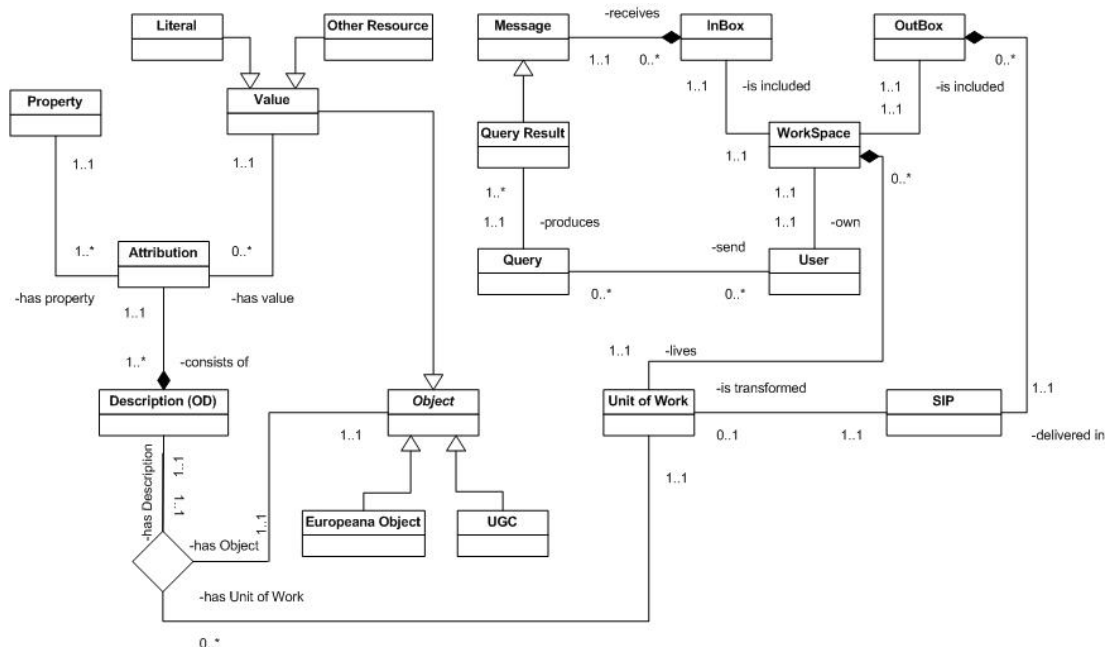


Figure 3.1 UML class diagram of the UGC service conceptual model

- Newly created objects, which are called UGC objects. These objects are original contributions to Europeana, and can be of three kinds:
  - digital objects having an associated media file with the content;
  - digital objects for which no media file is available;
  - non-digital objects.

Every object in a UoW has an associated description. A description represents a metadata record of the object and is modelled as a set of attributions, each attribution consisting of a property and a value. Different attributions can have the same property with a different value. A value can be itself an object, or a literal or another resource, external to the digital library.

When a UoW is ready to be submitted to Europeana, the user can do so by using an operation that transforms the UoW into a well-formed Submission Information Package (SIP) and places a message signalling the existence of the SIP into the Outbox. Each user WS is endowed with an Outbox. Europeana retrieves messages from Outboxes in order to harvest the corresponding SIPs.

As already mentioned, users can issue queries to Europeana in order to retrieve objects. Each query returns a result, in the form of a message stored in a special area of the user WS called the Inbox. For each retrieved object, the result of a query contains a subset of metadata of the ESE schema of the object, a URI to the full set of metadata and a URI to the digital object on the content provider's web site, if any.

Each user WS is endowed with an Inbox. Messages in the Inbox are of two kinds: query results and notifications that communicate the result of submissions. These latter notifications may be positive and report the successful ingestion of a SIP, or may be negative and report the reasons why a certain SIP could not be ingested. In the latter case, the rejected SIP can be retrieved and re-transformed into a UoW so to allow the user to perform the necessary repairing actions.

It is important to notice that these concepts define a general-purpose schema, whose machinery need not be used by every UGC application. For instance, a simple UGC task that takes place in a single session, such as an image upload, may be implemented by directly building the corresponding SIP, so by-passing the UoW stage. On the other hand, another UGC application may decide to publish a finished UoW to a community of users in order to perform a socially oriented form of mediation before submitting the UoW to Europeana. These decisions will be taken by the client side of the applications, relying on appropriate shortcuts offered by the UGC API.

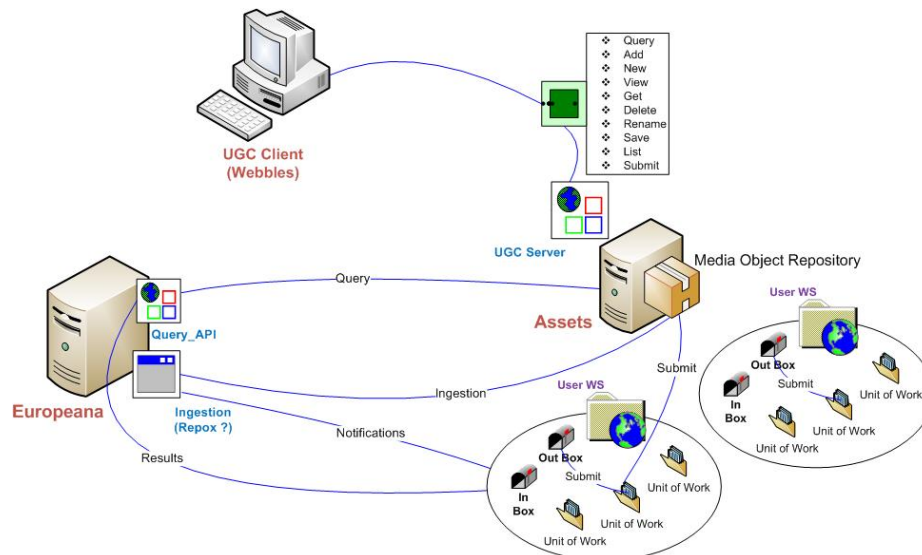


Figure 3.2 UGC service architecture

## 4 Architecture

For the purposes of developing, testing and evaluating with users the UGC functionality, the UGC server will be deployed on the ASSETS Server. After successful evaluation, the Server will be moved into the Europeana production server and its functionality will be made available to the Europeana users.

Figure 3.3 shows the architecture of the UGC service. The three main components of this architecture are:

**Europeana Server.** This is the server implementing the Europeana search functionality. It provides an API to search for content in the digital library, using optional filters on metadata fields. Europeana query API implements the Open Search directives (<http://www.opensearch.org/Home>). The Europeana server also provides an application for harvesting of data (e.g. Repox, is a GUI based tool for managing these activities).

**ASSETS UGC Server.** In the context of UGC, the ASSETS UGC server provides functionality to communicate with Europeana and with the UGC client. The communication with the Europeana server is by invoking the Query API module (namely OpenSearch API). The ASSETS UGC server manages the workspaces of the users, and provides an API to manipulate the UoWs in the workspace. The UGC server provides API as REST Web services and is independent from any specific UGC client. The User WS contains an Inbox, an Outbox and the set of Units of Work that the user is currently playing with. In addition, the UGC Server maintains the ASSETS Media Object Repository (AMOR), logically partitioned into two sub-repositories: the repository storing the SIPs not yet harvested by Europeana, and the repository storing the SIPs already harvested by Europeana but not yet ingested by it. On the former repository, AMOR implements OAI-PMH functionality to allow Europeana to harvest SIPs.

**UGC Client.** Is a browser-based GUI supporting the user in a specific content generation tasks. The UGC Client interacts with the ASSETS Server via REST web services provided by the UGC Server module.

## 5 Evaluation

Digital Libraries, Archives and Museums represent the main target for Europeana and through them it will be possible to gain a better understanding of user needs and requirements that will contribute to the

design of future services for the users in general as well as for specific groups of users and within specific environments such as the mobile usage and social media applications.

Within the Europeana project, there will be two types of evaluations carried out: a technical and a user-centred evaluation. The technical evaluation will be focused on verifying that the functionalities delivered by ASSETS fulfil efficient and scalable requirements. These evaluations will be conducted closely related to EuropeanaLabs. The present paper does not describe these evaluations. The user-centred evaluations aim at verifying that the services delivered by ASSETS fulfil the expectations of the end-users. The EDL Foundation together with ASSETS will conduct the evaluation. More specifically, the goal for the evaluation will be a) to utilize a user-oriented approach, and b) to focus on usability aspects of the services proposed and their end-users.

## **5.1 User-oriented evaluation**

One of the main issues within ASSETS will be to highlight the user's perspective. This will involve a specific attention to the users needs and requirements for the design of new functionalities in order to be incorporated into the Europeana system. A relationship with the Europeana Connect project (<http://www.europeanaconnect.eu/>) will be established.

The evaluation will focus on the usability of services for the end-users, including professionals (such as curators and scholars) as well as the non-expert user. Both kinds of users will be provided by Europeana and by the content providers that are members of the ASSETS Consortium. The outcome of these evaluations will be fed back to the developers of individual functionalities. The evaluation task will serve two objectives as stated in the project plan: to support a user centred deployment and to define exploitation channels.

## **5.2 Methodology**

We intend to perform an iterative evaluation procedure. In practice this means that we will perform at least two evaluation tasks during the lifetime of the project. The first evaluation task will take place after the presentation of the first release of the ASSETS user interface (June 2011), and the second evaluation task will be performed at the end of the project involving the final version of the ASSETS user Interface (March 2012). Between these two main evaluation stages, we may conduct further and smaller evaluation on specific functionalities. This means that the components, functionalities and user interfaces will be evaluated several times.

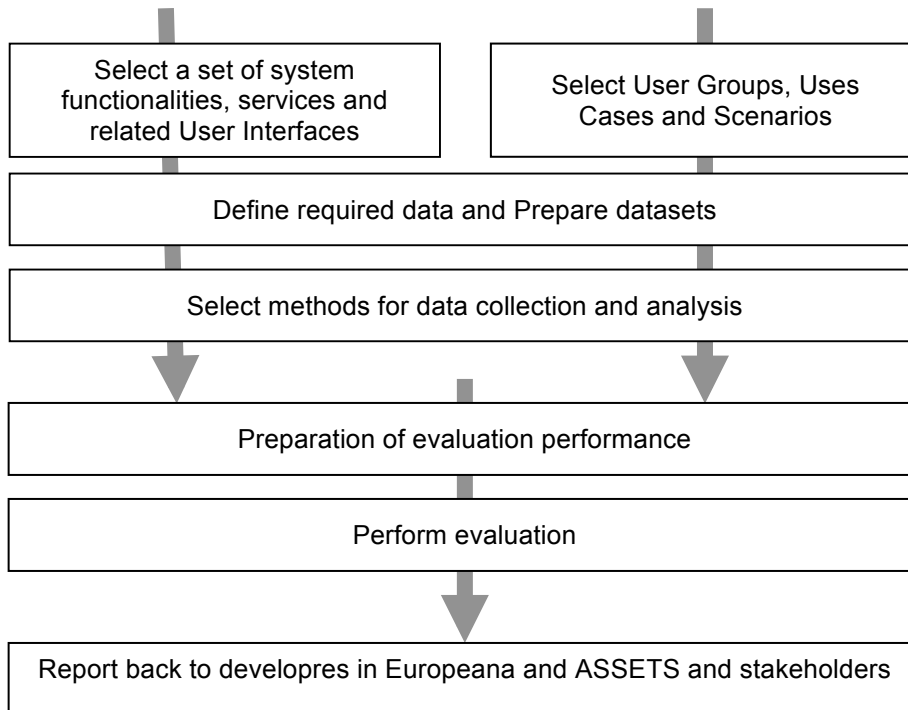
The methodology for the user-centred design and user evaluation is based on both laboratory involvement and experimentation and field studies. First, we will perform the evaluation in a laboratory environment under more controlled forms. Secondly, we will perform the same evaluation on site together with the users, using a slightly different protocol. This evaluation will be more uncontrolled but may be performed under more realistic situations, and thus, may reveal interesting behaviour. Possible data collection methods may be used: focus groups and on-site observations in order to get qualitative data for analysis and questionnaire and log-statistics for quantitative data to analyse.

The evaluations will include laboratory-based interactive sessions with users with different age, background, experiences, information needs, search strategies, language skills and cultural backgrounds. These factors, among others, will affect how single users will interact and utilize the user generated content service offered by Europeana. This in turn will affect how the user interfaces and the functionalities behind the user interface will be iterated. We will include users from the data providers within the ASSETS Consortium as well as using the Europeana panel of end-users if needed.

A set of broad use cases has been defined in the requirement collection phase, and within the evaluation framework, we will utilize a set of these use cases across user groups. The evaluation activities can be summarized in the steps. Diagrammatically illustrated below.

In the first user evaluation task, the following functionalities will be targeted:

- Image uploading
- Object annotation
- Metadata Enrichment



The outcome of the evaluation will inform refinement of these specific content generation tasks and the development of additional tasks requiring interaction with the Europeana system. The result of the evaluation will contribute to a set of iterations of specific functionalities, qualitative changes in user interfaces supporting better interactions and ultimately contribute to extended usage of the service.

## 6 Conclusions and Outlook

The main concepts and architectural features of the user-generated content service have been illustrated. The service is being implemented by the ASSETS Best Practice Network and will be evaluated within the lifetime of the project. The basic principles of the evaluation methodology have been described.

The UGC service developed by ASSETS is based on a general-purpose back end, which is meant to relieve Europeana from dealing with the specificities of the possibly very many UGC tasks that may be offered to users. At the same time, the back end relieves developers of UGC tasks from implementing the server side of their applications.

## 7 Acknowledgements

This work has been partially supported by the PSP Best Practice Network ASSETS (CIP-ICT PSP-2009-3, Grant Agreement n. 250527).

## References

- [1] [www.europeana.eu](http://www.europeana.eu)
- [2] [www.assets4europeana.eu](http://www.assets4europeana.eu)
- [3] [version1.europeana.eu/web/europeana-project/](http://version1.europeana.eu/web/europeana-project/)
- [4] [europeanalabs.eu/wiki/](http://europeanalabs.eu/wiki/)