eChallenges e-2011 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2011
ISBN: 978-1-905824-27-4

Europeana API – Example of Use in Polish Digital Libraries

Adam DUDCZAK, Agnieszka LEWANDOWSKA, Marcin WERLA
Poznań Supercomputing and Networking Center
ul. Noskowskiego 12/14, 61-704 Poznań, Poland
E-mail: {maneo, jagna, mwerla}@man.poznan.pl

Abstract: Europeana is a pan-European metadata aggregator of information about cultural heritage objects published on-line by various European memory institutions. Since February 2011 Europeana publicly exposes API, which main purpose is to provide easier access to the Europeana data. During the pilot stage of the API deployment, cooperation between Poznań Supercomputing and Networking Center and Europeana Foundation was established to create prototype API-based applications for two PSNC's services. This cooperation was later extended to co-organize Polish edition of Hack4Europe event in June 2011– one of the four hackathons aimed to show the potential of open cultural heritage data. The objective of this paper is to present a case study of the adoption of the Europeana API in a country scale digital libraries infrastructure in Poland and also to describe the results of the Polish edition of the Europeana API hackathon.

1. Introduction

In September 2005 the European Commission published a communicate titled „i2010: Digital libraries", which emphasized the need to ensure wide on-line access to digitised materials coming from European libraries, archives and museums, reflecting the rich history of Europe and its cultural and language diversity. Access to such materials should make easier for the citizens to "appreciate own cultural heritage, and achievements of other European countries, and use it for education, professional or recreational activities" [1].

Effective and wide implementation of this idea requires both the significant amount of scientific research and the development of practical solutions, including necessary pilot deployments. To make this possible, several consecutive funding programmes were announced by the Commision, including eContent, eContent*Plus* and CIP ICT PSP. These programmes allowed to digitise and put on-line large amounts of cultural heritage material and to build new services based on such data. The flagship initiative in this domain is Europeana, pan-European metadata aggregator and portal developed in a series of projects co-financed under mentioned EC programmes. The Europeana portal is available since November 2008 at http://europeana.eu/.

Creation of such international service required close cooperation between cultural heritage institutions on a scale unprecedented earlier. One of the concepts developed for this cooperation was the hierarchical aggregation model which assumed that there will be a number of national or regional metadata aggregators established across Europe. Such aggregators, working directly with local and regional memory institutions, would be a natural intermediary between thousands of possible data providers and Europeana, making this large scale data transfer much easier to establish and sustain. This model was deployed into practice in the EuroepanaLocal project.

In case of Poland, the international activities shortly described above were preceded by intensive development of the national infrastructure of digital libraries [2]. The first Polish

regional digital library, the Digital Library of Wielkopolska (DLW, http://www.wbc.poznan.pl) was launched in October 2002. Now it is the largest Polish digital library with more than 135 000 objects from over a dozen of institutions from the Wielkopolska region. Close cooperation of Poznań Supercomputing and Networking Center (PSNC) and Polish scientific and public libraries initiated by the launch of this digital library resulted in the last ten years in the creation of innovative platform of distributed digital libraries [3]. Majority of these libraries are based on the dLibra software developed by PSNC (http://dlibra.psnc.pl/). In June 2011 in Poland there were over 70 digital libraries with more than 670 thousands of objects available on-line.

The rapid development of digital libraries in Poland led to the need of virtual integration of these libraries into a coherent system. In June 2007 PSNC launched the PIONIER Network Digital Libraries Federation (DLF, http://fbc.pionier.net.pl/), a network service designed to be a national metadata aggregator and a catalyst in cooperation between its data sources. In the December 2009, under the frame of EuropeanaLocal project, the connection between the Federation and Europeana was established. Poland was the first country with national metadata aggregator connected to Europeana.

The development of the DLF was run in parallel to the development of Europeana and the cooperation between these two initiatives was not limited only to periodic data transfer. In 2010 Europeana started development of the API, a tool facilitating the automated access and use of data aggregated by Europeana. After the Europeana Open Culture Conference in October the same year, PSNC started cooperation with Europeana Foundation on a prototype use of the Europeana API in some of PSNC's services. It was decided to develop two widgets based on the Europeana API: one for the Digital Libraries Federation and the other one for the Digital Library of Wielkopolska. The next chapters of this paper describe the process of design and implementation of these widgets. The paper ends with an analysis of the impact achieved by the widgets production deployment, the description of the Hack4Europe! event and with a summary.

2. Analysis and Design of the Europeana API Applications

The purpose of the development of widgets using Europeana API was to provide easier access to European cultural heritage artefacts for users of Polish digital libraries, without forcing the users to change their usual workflow. It was assumed that search in aggregated metadata is the main DLF functionality for the majority of end users. The search result element on the DLF search results page contains only few elements of the harvested metadata and redirects the user to full information in the source digital library (e.g. in DLW). The functionality left to the source digital library is to display the full metadata record and to give the access to the content of the digital object. The final idea was to enrich the information presented to DLF and/or DLW users with links to additional objects available via Europeana, which can be practically done by putting widgets based on Europeana API on the DLF search results page and DLW full metadata record page.

Further analysis was focused on technical aspects. Europeana API is an Open Search protocol interface. To get some results, an input query is needed. It was decided that for the DLF widget the input will be the query submitted to the DLF by the user, and for the DLW the query will be built from selected elements of a particular metadata record displayed to the user.

As the metadata from DLF is visible in Europeana, another issue was the fact that DLF database is updated each night and DLF to Europeana data transfer in practice takes place every three months. As a result, DLF is a more up-to-date source of information for Polish digital libraries metadata search. On the other hand, Europeana of course contains a lot more information than the Federation. The final decision was to join the data from Europeana and the DLF at runtime: when preparing the final set of information to be shown

to the user, the results from Europeana should not include data from DLF, as this data should be taken directly from DLF. Another challenge was related to cross-language searching. It was decided that the subject element from the DLW metadata records will be translated with Google translate to English, Spanish, German and French before it is sent by the widget as a query to the Europeana API.

3. Implementation of API-based Widgets

The final result of the technical discussions was the architecture presented on the Figure 1 below.

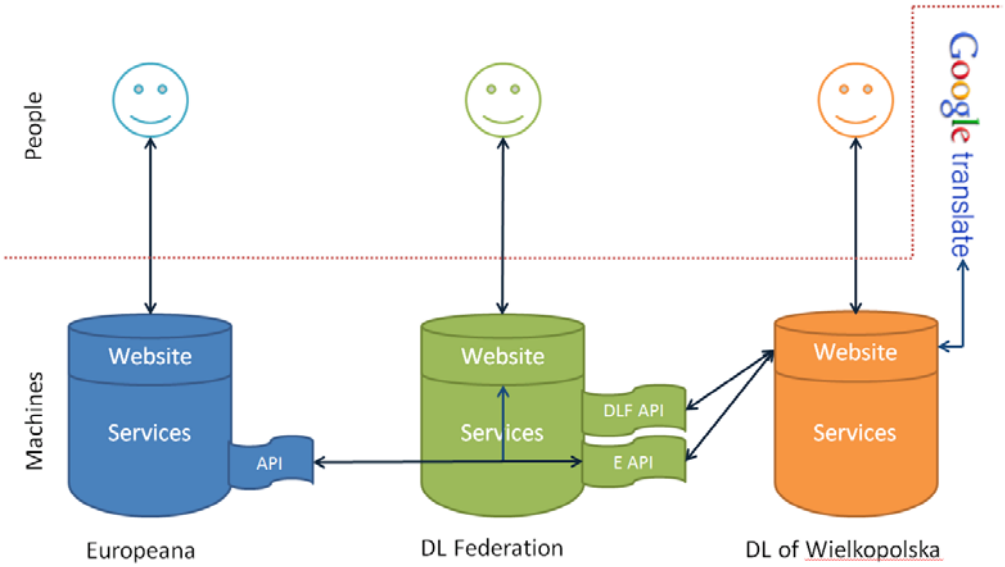

Figure 1. Final Architecture of Europeana API Usage by the Digital Libraries Federation and the Digital Library of Wielkopolska.

The proposed solution integrates the functionalities provided by three services – Europeana, Digital Libraries Federation and Digital Library of Wielkopolska:

- Europeana exposes Open Search API [4] giving the possibility to perform advanced search in all aggregated metadata and get the results in convenient machine-readable format.
- Digital Libraries Federation uses Europeana API [5] to provide the search results from Europeana together with the search results from the Federation. Those results are presented together in the Digital Libraries Federation website, as shown see on Figure 2. Moreover, the Federation exposes two Open Search APIs for external services. One of those is the Federation's API and the second one is a proxy to Europeana API dedicated for use by Polish digital libraries. The main reason for which the Europeana API has a proxy in the DLF is the ease of development and use of the widget prepared by the Federation for Polish digital libraries, related to the need of a key to use the Europeana API and to the cross-site scripting issues.
- The widget embedded in websites presenting metadata records of digital objects published by the Digital Library of Wielkopolska extracts parts of the metadata, translates them on the client side with the Google Translate service and sends the translated metadata together with the OAI identifier of the digital object to both Open Search APIs exposed by the Federation. After responses are processed by the widget, the search results are presented as a part of the website with the digital object metadata. The example of such results in shown in the left column on Figure 3.

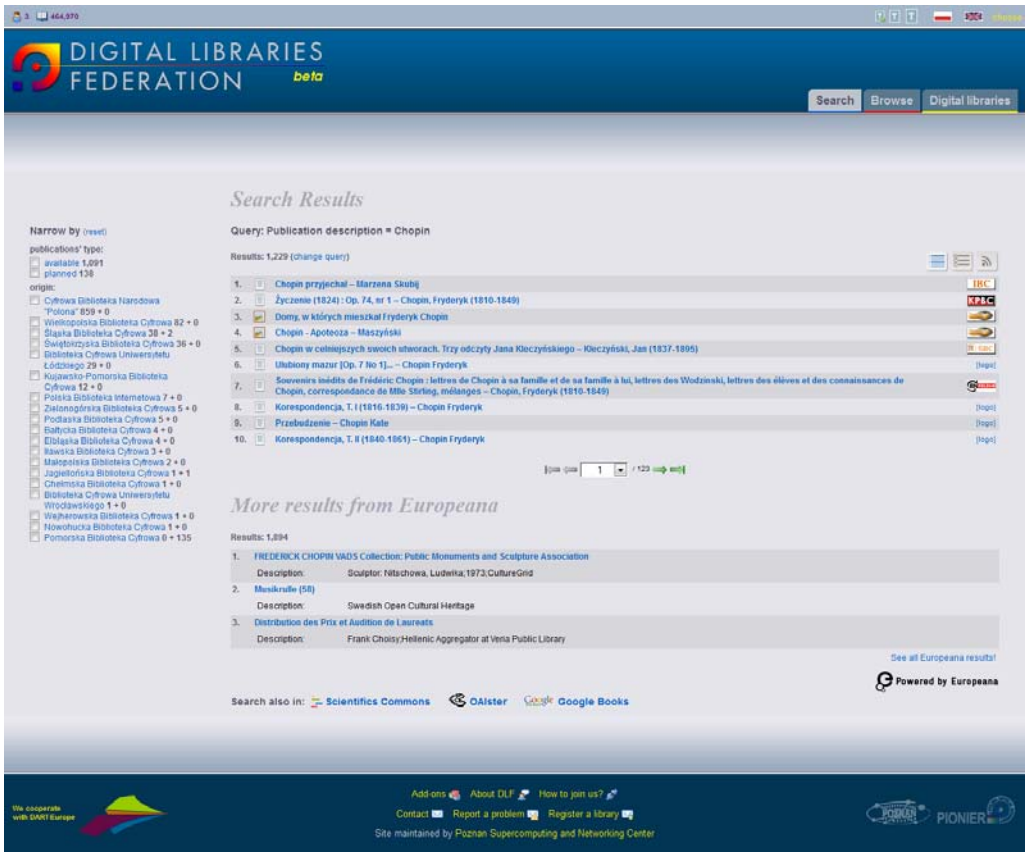

Figure 2. DLF's Search Results Page with Results from Europeana.

Figure 3. Search Results from Europeana and DLF on the Digital Libraries of Wielkopolska Site

The design and implementation of both widgets and other necessary code took about 10 person-days of a skilled programmer. The API-based widgets were first deployed in the test environment and consulted with Europeana Team which was also responsible for providing technical information about the API. Then on the 22nd of December 2010 widgets were deployed in the production environment.

4. Usage and Impact of Widgets

Successful deployment of widgets based on the Europeana API in services which are directly or indirectly contributing data to Europeana should effect in a win-win situation. In this situation contributing services propose new functionality to the users attracting them to Europeana, where they can discover new interesting resources. On the other hand the increasing pool of Europeana users effects in increased traffic in services which are providing data, as Europeana links each metadata record to the related digital object presented on the website of the object provider. This user flow is presented on Figure 4.

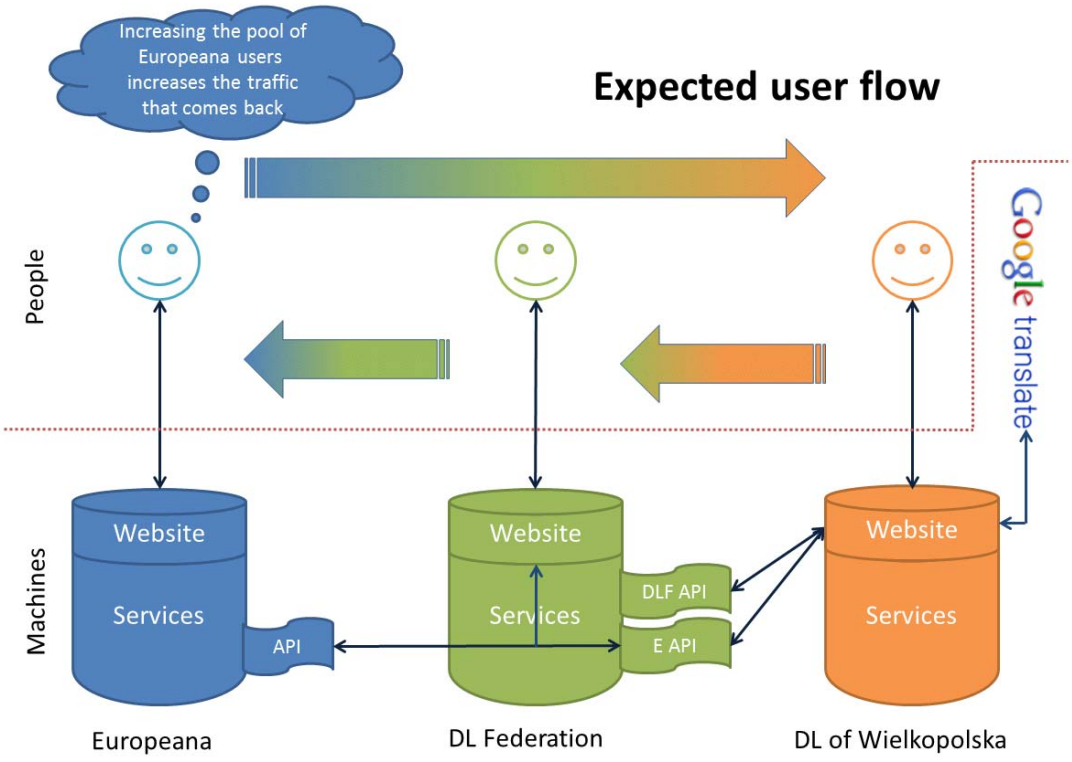

Figure 4. Expected User Flow between Europeana and Services Utilising Europeana API

To find out whether the widgets are really useful for end users, a complex traffic analysis in all involved services should be performed, including deep analysis of HTTP access logs. Unfortunately such logs for Europeana are not publicly accessible. To at least estimate the extent of widgets usage it was decided to use Google Analytics data, including data from Europeana portal thanks to the courtesy of the Europeana Foundation.

Both DLF and DLW get about 70,000 visits each month. Figure 5 contains a comparison of the percentage share of three types of such visits for DLF (blue bars) and DLW (red bars). The middle pair of bars (marked as 100%) represents visits during which a visitor displayed a page with the widgets installed. Those pages (DLF: search results page; DLW: metadata record page) are so crucial to the functionality provided by the service that it was assumed that any visit skipping those pages probably have been somehow accidental. The pair of bars on the left shows the number of all visits. Around one third of all visits are not reaching the crucial functionality of the website. This is for sure something that could

be improved. But coming back to Europeana API widgets – the last pair of bars on Figure 5 shows the percentage of users who reached the page with widget and decided to click on the digital object's link provided by Europeana via the API. As you can see, 7.5% of the Federation users went to Europeana and 0.67% of DLW did the same.

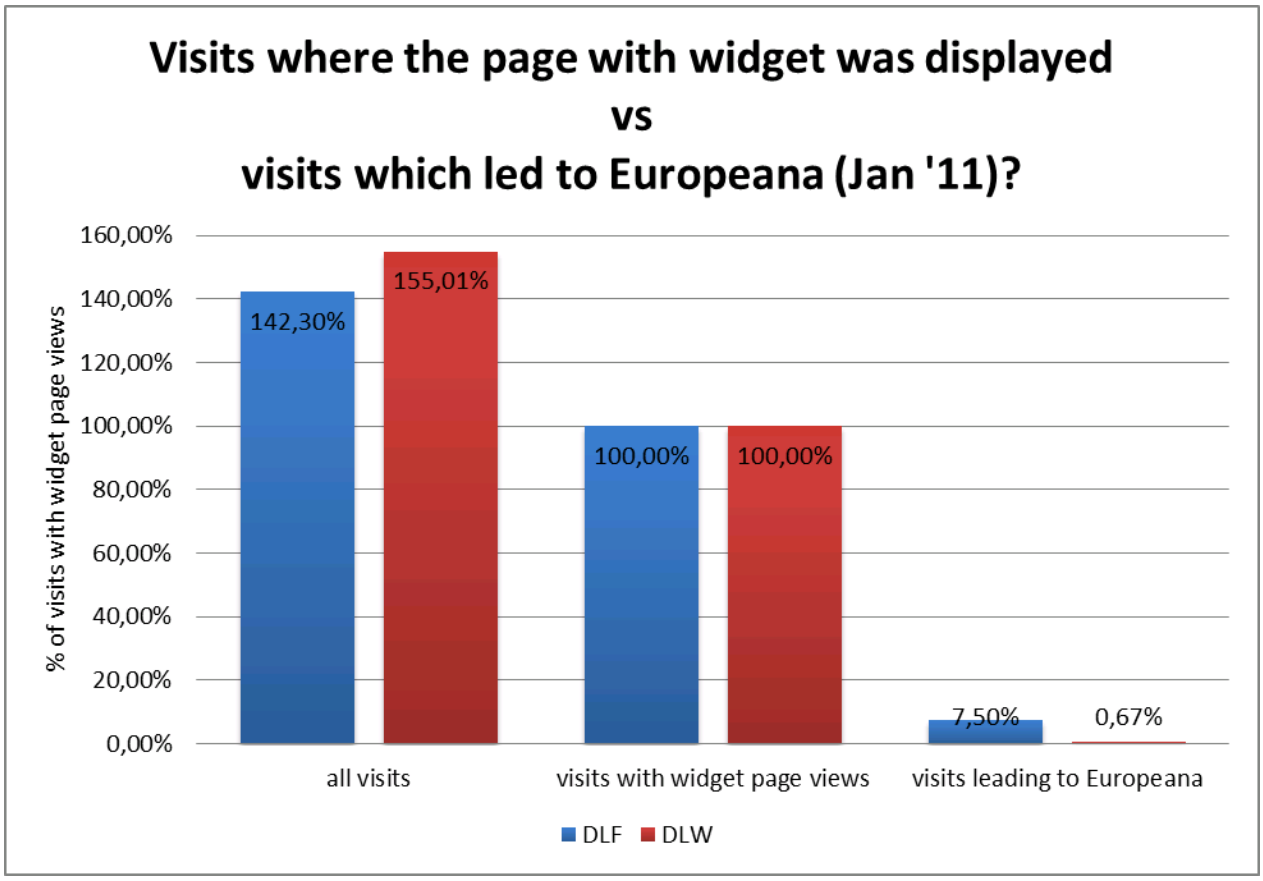

Figure 5. Comparison of User Visits in DLF and DLW and Utilisation of Implemented Widgets (Data for January 2011).

At this stage it is hard to estimate whether this is satisfactory, but treating these additional links as some kind of targeted advertisement placed on a cultural heritage website, the results may be seen as quite good.

The above data describes the user traffic from right to left on the Figure 4. Another statistic should be used to analyse the traffic coming from left to right of the figure. What can be easily calculated is the comparison of traffic sources for the Digital Library of Wielkopolska (the largest Polish digital library). Traffic sources are ways in which users reached the service. Such comparison for January 2011 is presented in Table 1.

Traffic source	Visits	%
(direct)	27 936	38,34%
google	14 925	20,48%
fbc.pionier.net.pl	8 286	11,37%
europeana.eu	3 143	4,31%
wtg-gniazdo.org	2 391	3,28%
pl.wikipedia.org	2 152	2,95%
genealodzy.pl	1 589	2,18%

Table 1. Comparison of User Visits in DLF and DLW (data for January 2011).

The table shows all traffic sources for the DLW in January 2011, which generated at least 1% of overall January traffic. Two first results are quite obvious – direct access (for example a bookmark in the web browser) and access from Google search results. But position 3 and 4 are very interesting. The 3rd place is the national aggregator – the Federation, and the 4th is the European aggregator – Europeana. The last three positions are taken by two genealogical services and Polish Wikipedia. These statistics show that the model of multilevel aggregation described in the Europeana Content Strategy is very good at attracting users to the participating digital libraries and it is worth to be a part of it.

Successful deployments of prototype applications based on the Europeana API was one of the important steps before a start of the wider promotion of the API. The public launch of the API was announced at the end of February 2011 and in the beginning of April the first Europeana API Hackathon was organised in the Netherlands. Two months later, in the beginning of June, a series of further hackathons under a common name Hack4Europe! was held. Details of the Hack4Europe! events are described in the next section of this paper.

5. Hack4Europe!

During the second week of June Europeana foundation have organized a series of four hackthons under the common name Hack4Europe! Events were organized by local partners in Poznań, London, Barcelona and Stockholm. Polish hackathon was organized by Poznań Supercomputing and Networking Center and The Kórnik Library of the Polish Academy of Sciences.

All hackathons aimed at the development of innovative applications created on top of data about 19 million of cultural heritage objects collected by Europeana and exposed via the API. Developers competed in four categories: application with greatest commercial potential, application with greatest potential for social inclusion, most innovative application and audience award (this one was voted by developers).

As a result Hack4Europe! Events gathered in total 85 developers, who have prepared 48 prototypes. The most common development themes included applications designed for mobile devices, applications using the potential of social networks, solutions allowing users to curate content, integrating Europeana content into various games, connecting cultural heritage data with Wikipedia and finally various visualizations showing how various objects are related.

Polish edition of Hack4Europe! was held on 7 - 8th of June with 18 participants from various cities all over the Poland. During two very intensive days they managed to create 8 prototypes. Three of them were awarded by the jury:

- "Art4Europe" was the winner in the greatest commercial potential category. This application allows its users to identify given art work using picture taken by the camera of a mobile phone. It identifies given object, presents the description of the object, apart from this it can also translate it to any European language and read this description aloud. Users might be also informed about the possibility of buying reproductions or books about given art work.
- In "Europeana Field Game" (most innovative application) the user can virtually carry and pin elements to a location and see elements pinned to a location by other users. The game encourages geotagging by introducing quests to accomplish and interaction with other users.
- "Hackmemory", which won the social inclusion award and the audience is a simple educational application for kids and adults based on well know memo game. Players have to find two exactly matching pictures. After finding each pair user can read about the content of the picture. User can create his/her own quiz and simply share it with friends on using various social media. The content of the puzzle comes from Europeana and it is filtered by the creator of the quiz (i.e. teacher).

Rest of the prototypes were tools allowing for integration of Europeana API with Google Maps and with MediaWiki. It is also worth to mention that awards for best projects were funded by Speed Up Group sponsors of Poznań Hack4Europe!

Polish winners took apart in second round to compete with applications awarded during hackathons in Barcelona, Stockholm and London. The Art4Europe project won once again the most commercial potential award, but this time on the level of Europe. Apart from this project, also three other applications were awarded:

- Casual Creator (London hackathon) application which facilitates using pictures of the cultural heritage objects in teaching.
- Time Mash (Stockholm hackathon) fully functional geo-location aware search of Europeana for mobile phones.
- Timebook (Barcelona hackathon) integrates content from Europeana and DBpedia and presents it in an easy to use, Facebook-like format with posts for famous quotes, friends status for influential persons and photos of paintings etc.

Hack4Europe! awards ceremony was organized on 16th of June 2011 during the Digital Agenda Assembly in Brussels. Winners received prizes from European Commission Vice-President Neelie Kroes.

6. Summary

This paper presented the PSNC's experience with the adoption of Europeana API for creation of applications in the digital libraries domain. The API offers very interesting possibilities and it is easy to use, as it is based on the well-known Open Search standard. The basic analysis of the use of widgets designed, implemented and deployed by PSNC shows that users appreciate this functionality and that Europeana portal can be a valuable source of incoming traffic for its contributors.

The series of hackathons organized by the Europeana Foundation shows that the aggregated open cultural heritage data is very valuable. The interest of over 80 developers from all around Europe and the prototypes they created are proofs that the data have potential for business use, can be used for social inclusion or can be a driver for innovation.

References

[1] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - i2010 : digital libraries. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0465:FIN:EN:HTML

[2] C. Mazurek, M. Mielnicki, M. Werla, Selective harvesting of regional digital libraries and national metadata aggregators. In: Proceedings of 9th ACM/IEEE-CS Joint Conference on Digital Libraries. Austin, TX, USA, 15-19 June 2009 r. pp. 429-430. ISBN 978-1-60558-322-8. http://portal.acm.org/citation.cfm?id=1555497

[3] A. Lewandowska, C. Mazurek, M. Werla, Enrichment of European Digital Resources by Federating Regional Digital Libraries in Poland. In: Research and Advanced Technology for Digital Libraries, 12th European Conference, ECDL 2008, Aarhus, Denmark, 14-19 September, 2008, Proceedings. LNCS vol. 5173, pp. 256–259. http://www.springerlink.com/content/p75044w33g1x4610/

[4] Open Search Protocol Specification ver 1.1 http://www.opensearch.org/Specifications/OpenSearch/1.1

[5] Europeana API Services Website http://version1.europeana.eu/web/api/