IFLA WLIC 2017 WROCŁAW

Submitted on: 07.06.2017

Cloud technologies as a solution for digital collections management in small libraries

Marcin Werla
Digital Libraries and Knowledge Platforms Department, Poznań Supercomputing and Networking Center, Poznań, Poland.
E-mail address: mwerla@man.poznan.pl

Abstract:

Aim of this paper is to present experiences related to development and maintenance of LoCloud Collections platform (https://locloudhosting.net), including the identified needs of small cultural institutions and features of the described platform. This platform was developed during EC-funded project and is running as a completely free production-quality service since 2015. It is currently being used by 40 institutions from 11 European countries. This service allows quick and easy creation of your own digital library and publishing of your data on-line and to Europeana. This service is available in so called Software-as-a-Service model. It means that using LoCloud Collections does not require to own IT infrastructure and staff. The service is free and will remain free in the nearest future, as it is funded from public sources.

Keywords: digital libraries, cloud services, cultural heritage, on-line access, small institutions

Introduction

Setting up on-line services which meet requirements of libraries interested in wide, on-line access to their collections and interoperability with services like Europeana, usually requires investments in proper IT infrastructure and staff or possession of such assets. This is possible mostly for big institutions, having proper funding. Smaller cultural institutions, including local libraries, often on their own cost or with the support of small grants, are digitizing their collections, but they cannot afford to ensure constant on-line access to these resources. The result is that unique local collections remain hidden and the only chance to discover it is to personally visit such a local institution.

Thanks to the widespread of cloud technologies, it is now possible to propose professional, but also very economical solution allowing also the smallest institutions to exist with their collections on the internet. The IT infrastructure no longer has to be owned and major upfront investments are not needed. Virtually anyone can easily buy access to a server in one of public clouds offered by many companies worldwide. The charge for using such server is based on the actual consumption of its resources so if someone needs just a small digital place to publish its business card it can be done really cheap, with a cost comparable to buying a cup of coffee once per month.

The still existing challenge in the context of small heritage institutions is the availability of IT personnel, that will be able to set up and maintain cloud-based server to serve the needs of such institution. High demand for IT professionals on the labor market results in high salaries in this sector and low availability of potential employees. For small institution it practically not possible (and probably not needed) to have a dedicated full time IT administrator or software developer. There are companies utilizing this fact and offering remote administration services, but in most cases they are not really interested in getting into the details of cultural heritage software systems nor into specifics of such enterprises like collaboration with Europeana. The final result of such situation is that often the on-line visibility of small heritage institutions is limited to a few static pages with general description, contact details and photos of selected items from the collection of this institution.

Works undertaken in EU-funded LoCloud project were aimed to deliver an on-line service which would utilize the potential offered by cloud technologies to provide small cultural institutions professional, but also very economical solution allowing also the smallest institutions to exist with their collections on the internet, without the need of having any IT expert on site. In the following sections this paper describes how users requirements for the service were defined and how the final service, LoCloud Collections available since 2015 at https://locloudhosting.net/ was able to address them.

Digital library for small institution means what?

Initial users requirements for LoCloud Collections system were collected during three workshops organized as a part of the project in August and September 2013 in Copenhagen, York and Madrid. During these workshops some initial ideas regarding LoCloud Collections service were presented[1] and discussed with project partners. Following potential user groups were identified:

- Small and medium memory institutions with lack of IT support and lack of resources in general
- Institutions with a working repository, looking for a way to cut costs
- Hobbyists and professionals with no knowledge about digital library standards and Europeana
- Large supporting organizations (NRENs, domain aggregators) interested in supporting small organizations in given domain/country
- End-users which are interested in local heritage and require intuitive user interface

Another outcome of these discussions was also a list of general requirements which are summarized below:

- It should be very easy to use, at least for basic features, no IT staff required to start using the service.
- It should support a way to deposit and retrieve at least 4 basic objects types: photographs, textual documents (multiple pages), audio content and video content.
- It should have multilingual interface to increase accessibility of collections
- It should be more affordable than self-hosted digital library
- It should support for migration from major digital libraries systems would be appreciated

[1] http://locloud.eu/Media/Files/Presentation-Files/The-LoCloud-lightweight-digital-library-and-alternative-content-sources

- It should be able to handle significant traffic should be scalable in terms of storage and traffic.
- Look & Feel of the archive's website should be customizable
- It should allow to restrict access to given resources
- It should support batch upload as the user would like to move from his own portfolio to the new LoCloud-based repository
- It should supportin creation of high quality metadata, incl.:
    - geolocation,
    - controlled vocabularies.
- It should be provided with documentation and (optional) support in local language.
- The content stored in LoCloud Collections can be downloaded when the library decides to go somewhere else.
- Batch upload of items should allow to enter metadata values which are the same for all uploaded objects e.g. newspapers.

The final list of key high level requirements/features for LoCloud Collections was defined in the following way:

1. **Easy and fast start** – after registering in the service and filling in a simple form, the user should be able to start making his collections available on-line in just few minutes.
2. **Professional, yet easy to use** – the service should offer a digital collections management system which will be easy to use (no need for a long training nor a lot of manuals reading), but will also have more advanced features for more demanding users.
3. **Support for many different collections** – the user should be able to put on-line many different collections in one account.
4. **Support for many data formats** – the user should be able to publish photos, scans, text documents, as well as audio and video materials.
5. **The possibility to adjust end-user interface** – the user should be able to choose one of predefined graphical templates, and then adjust it for your needs.
6. **The possibility to adjust metadata schema –** the user should be able to adjust the metadata schema for his needs, incl. defining vocabularies for certain metadata elements. The system should offer support providing geolocation in metadata.
7. **Batch upload of data** – users of LoCloud collections should be able to make a batch uploads of data – both metadata and content, to facilitate uptake of this service and migration from other local.
8. **Access to documentation and support** – on-line support for LoCloud Collections users should be provided.
9. **Fees dependent on the use of resources** – assuming a business model which will require fees, the level of fees for using the LoCloud Collections service should be defined by the chosen pricing plan which corresponds to the amount of system resources consumed by the user.
10. **No need to make large financial investments** – using LoCloud Collections should not require users to invest in hardware or software before starting to make their collections available on-line.
11. **Multilingualism** – taking into account that LoCloud project is aimed to very small heritage institutions, most probably the presented collections will have very local or at most regional scope. In such context it is very important to be able to provide the user interface in the local language as well. LoCloud Collections should be fully prepared for having multilingual user interface.

12. **Compatibility with Europeana** – using LoCloud Collections should assure that there will be no technical issues in providing information about your collections to Europeana.
13. **Access to advanced statistics about usage of collections** – anyone who makes collections available on-line is interested in information about how much they are used. LoCloud Collections should offer access to dedicated analytical service, in which detailed information about the visitors traffic is automatically collected (with respect to users privacy).
14. **Scalability** – The user should be able to change the amount of available system resources at any time, adjusting the storage space which is available for actual needs, so that he can start with the smallest pricing plan and gradually increase the available storage space, as his digital collections will grow.
15. **Constant access** – the system should be able to guarantee that collections will be available on-line almost all the time (aiming at ~99.9% of service availability).
16. **Increased data safety** – putting digital collections to LoCloud Collections should give user one more backup copy, which he can use in case of problems with data stored locally, therefore the data in LoCloud Collections should be safe.
17. **Visibility in search engines** – websites based on LoCloud Collections should be ready for automated crawling and indexing by robots, so that people can find collections on Google and other popular search engines.
18. **Openness** – LoCloud Collections should be based on open source code, so that the users are sure that if sometime in the future they would like to stop using LoCloud Collections, they can download their data dump and use the same software to make their data available on-line somewhere else. Beside each collection in LoCloud Collections should make the data available via open API and OAI-PMH protocol.

The above requirements were used for analysis of available technologies, tools and frameworks which could be used to develop LoCloud Collections system, and then for the implementation of the service.

LoCloud Collections service implementation

After the initial step of requirements gathering, the service design and development stage was started. In technical context, the following open source systems were taken into account as the basis for LoCloud Collections implementation: DSpace (http://dspace.org/), E-Prints (http://www.eprints.org/), Omeka (http://omeka.org/). Beside open-source licensing, another aspects taken into account were:

- How big and stable is the community of users of the system?
- How actively the system is developed – when last two releases were made?
- How lightweight is the technology stack of the system? Is it easy to move it to cloud environment?
- What is the main purpose for which the system is being developed?
- Is that system already widely offered as a hosted service?
- How well the system supports the key requirements defined in the analysis phase?

All three systems are well-known open source solutions with stable users and developers community, although it seemed that E-Prints software was least actively developed. Each three systems have different technology stacks: DSpace has the most heavy one (which is problematic for SaaS where many instances will be set up on shared resources), E-Prints and Omeka have similarly light stacks, but PHP hosting is in general more popular among cheap

web hosting options. This is also the basis of many popular CMS systems. Only Omeka was initially developed for cultural heritage sector, but DSpace is becoming more and more general so it could be also used for such kind of collections. Omeka is provided as hosted service by its creators but also by many other companies in "one-click install mode" . There are also companies who offer DSpace and E-Prints as hosted service, but availability is far away from "one-click" and pricing is very high when compared to native Omeka.net offering (http://omeka.net). Finally the functional analysis of all three systems shown that each of them fulfils in general the requirements collected during content providers workshops, but in each case some customizations will be needed anyway.

The final decision after this review was to use Omeka, because it seems to be the most suitable choice for a small cultural institutions in terms of functional scope, technical "lightness", ease of setup and costs of hosting. After making this decision, main development of LoCloud Collections was divided in two threads: adjusting Omeka to suit the needs of LoCloud Collections users and building hosting environment which would allow quick and easy setup of new Omeka instances for its users and management of such instances.

The first implementation of the LoCloud Collections service is available on-line at https://locloudhosting.net/. This website, publicly available since January 2015, allows anyone to register and to create and manage his own digital library. Figure 1 shows example of such digital library available at https://ptpn.locloudhosting.net/, made by Poznańskie Towarzystwo Przyjaciół Nauk, small cultural heritage institution from Poland.

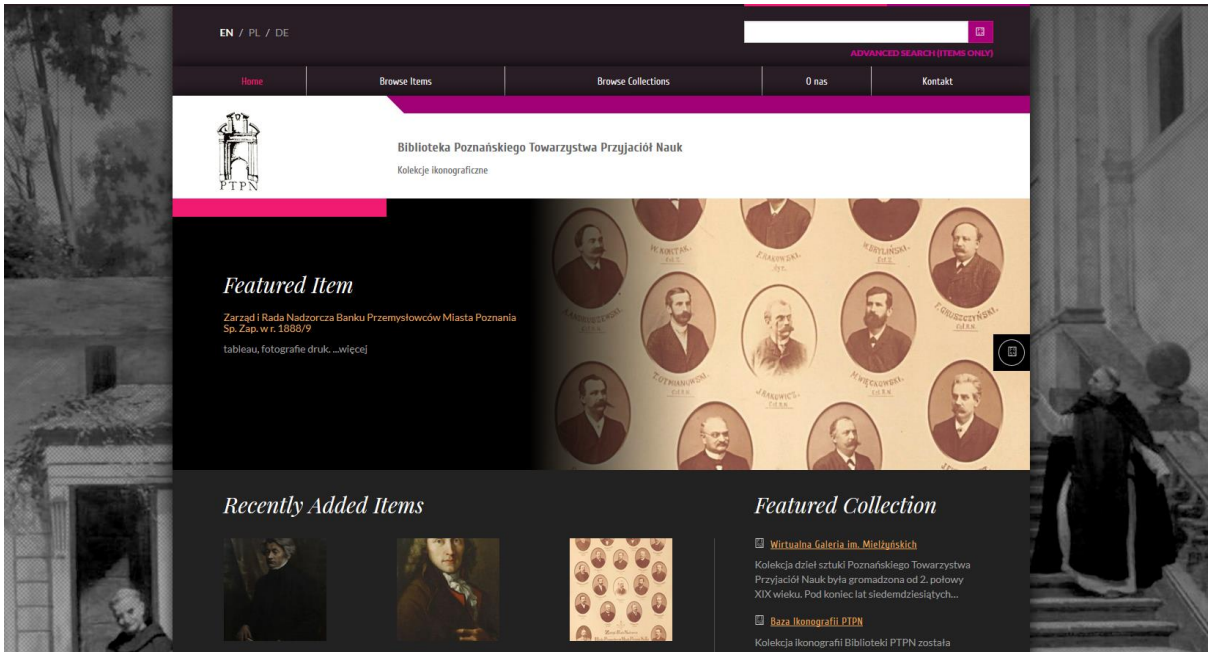

Figure 1. Example digital library in LoCloud Collections (https://ptpn.locloudhosting.net/).

Functional overview of the service, including screenshots, is presented in on line documentation - "LoCloud Collections: Basic User Guide"[2]. Final form of the most important functional aspects of the service is discussed in the next section of this paper.

[2] http://support.locloud.eu/tiki-index.php?page=LoCloud+Collections%3A+Basic+User+Guide

LoCloud Collections as a service useful for small memory institutions

When users enter the main page of LoCloud Collections they needs to do just a few steps to have their first digital library up and running. They need to fill in the registration form, activate the account by click on activation link in e-mail, log in to the service and fill in form with basic details of the digital library that is being created. The entire procedure takes up to 5 minutes. After that user have to wait around 2 minutes, while their new digital library is being created and then they gets access to the administrative panel of the digital library, when they can start customizing it and uploading data to be published.

As it was written above, digital libraries which are created within LoCloud Collections are based on Omeka software. This software is intuitive and easy to use and user can start putting content on-line in just few minutes of looking around in the system. A lot of more advanced functionality is provided to Omeka by additional plugins. To facilitate use of such functionality by beginners, in LoCloud Collections each Omeka by default gets over a dozen of plugins preinstalled and preconfigured. These plugins were chosen by PSNC staff while developing LoCloud Collections and tested if the work properly. The fact that these plugins are preconfigured allows users to benefit from their functionality without going into the details at the beginning. Later on when they get more experienced, they can check how each plugin is configured and adapt its settings. They can also request to have more plugins installed if needed.

Omeka supports having many different collections in one installation. So if users of LoCloud Collections service would like to organize their items into several collections, they have such possibility. If for example the nature of items is so diverse that users will not want to have a single Omeka installation with these items divided into collections, LoCloud Collections allows to have several Omeka installations in a single user account. Thereby LoCloud Collections adds additional layer on top of Omeka, which makes possible to use completely separated on-line collections systems. Omeka in general does not limit types of data formats which the user can upload. Everything which is uploaded to Omeka is in the same form returned to end users browsing public Omeka interface. Therefore it is just recommended to use widely supported data formats (e.g. JPEG, PDF, MP3, Ogg Theora Vorbis). Limitations can be introduced on a level of Omeka administrative interface for security reasons, for example to block the possibility of hosting executable files. Omeka system allows its administrators to choose one of predefined visual themes and then configure the properties of that theme to customize it. While preparing Omeka for LoCloud Collections, three themes which come with Omeka were chosen to be included and additionally four more were designed. There were two reasons for designing new themes. The original themes were rather basic and could not match tastes of collections curators looking for more sophisticated surrounding for their collections. Additionally original themes were not designed according to responsive web design principles, so using them to present collections on mobile devices was not giving satisfying results.

The possibility to adjust metadata schema was implemented on the basis of basic Omeka features which allow to extend Dublin Core Metadata Element Set with custom metadata elements grouped in item types. So Omeka administrator can define several item types like photo, song, map etc. and add new custom fields for each of that item types. The editor will have the possibility to describe each item with chosen DCMES elements and additionally with elements specific to chosen item type. Beside, a number of plugins was added to Omeka to make the metadata-related features more advanced, including support for geolocation, hiding

some metadata elements from the public and predefining vocabularies of allowed values for selected metadata elements.

Users of LoCloud Collections does not have to make any significant financial investments because this system is offered in Software-as-a-Service model. It means that it is running on cloud environment hosted by PSNC and users are accessing it (and utilizing PSNC's infrastructure) via network connection. To make the uptake of the service even more easier and less risky from the financial point of view, it was decided that an introductory free storage plan will be available, so that anyone can try LoCloud Collections with 500 MB of space available. It should be perfectly enough to test the service. It can be also enough for the production use of the system for really small institutions. Larger plans are also available, up to 50 GB, and for now they are free as well.

In order to provide advanced statistics about usage of collections hosted in LoCloud Collections open source web analytics software package called Piwik (http://piwik.org/) was integrated with LoCloud Collections environment. A dedicated installation of Piwik was setup and is available at https://locloudhosting.net/piwik/. Each time new user is registered in LoCloud Collections, he is automatically registered (with the same username and password) in Piwik. Each time new digital library is added/removed from LoCloud Collections it is also added/removed from Piwik monitoring mechanisms. The monitoring of single Omeka instances is implemented with custom Piwik plugin developed by PSNC. This plugin automatically sends all traffic information from Omeka to Piwik, excluding traffic generated by administrators logged in to Omeka admin panel. From the point of view of LoCloud Collections user who manages a digital library, in the detailed view of this library, in LoCloud Collections user panel, he can see basic traffic statistics including:

- chart showing daily number of visits during last 30 days,
- detailed traffic statistics (over 10 parameters) of visits during last week,
- geographical distribution of visitors during last week.

Additionally, below these statistics a link to multilingual Piwik installation is provided where the user can get access to a lot of other traffic information, including:

- real time view of visitors and their actions,
- statistics about time spent on website,
- popularity of items hosted in Omeka (including chart of item views over time),
- traffic sources,
- internal search queries.

In LoCloud Collections service single digital libraries are based on open source Omeka software. Each Omeka is preconfigured to publicly expose OAI-PMH protocol, allowing automated and open access to stored metadata and data – for the purpose of Europeana as well as for other aggregators and services. Beside Omeka offers REST API which gives even more possibilities than OAI-PMH interface and can be used for building applications on top of Omeka or to integrate Omeka with other tools.

Summary

All of the initial high level requirements defined for LoCloud Collections service were implemented in the service described in this paper. Some of these requirements are fully developed, while few are still in basic form and can be improved further in the future. The end-user feedback collected during the service operation period was positive, and the service was

described as user friendly and intuitive. Most common remarks collected were related to the visual side of the service – more user interface themes and translation were requested.

Another issue that was pointed out was the fact that in some countries it's not recommended (or even forbidden) to public institutions to store their data in cloud systems located abroad. This can be serious limitation for the adoption of the service in some countries, but hopefully the fact that the LoCloud Collections cloud is located in a public sector data center, inside the European Union, will be acceptable for majority of potential users.

The LoCloud Collections service works in fully operational mode since January 2015 and is actively maintained and developed by PSNC. In March 2017 another user research was done, which shows that the service is interesting for cultural institutions and is seen as something that can help small organizations to share their collections on-line.

Acknowledgments

The work described in this paper was initially co-funded by the European Commission as a part of LoCloud project, financed from CIP ICT-PSP programme. Further development of the service in 2016/2017 was co-funded by the EC, as a part of Europeana Digital Services Infrastructure 2 project, under the CEF programme.