Deliverable

Project Acronym:	**LoCloud**
Grant Agreement number:	**325099**
Project Title:	**Local content in a Europeana cloud**

D3.7 Report on services developed for local cultural institutions

Revision: Version 1.0

Authors:

AIT Angewandte Informationstechnik Forschungsgesellschaft mbH	**Odo Benda, Astrid Höller, Gerda Koch, Walter Koch**
ATHENA RC Athena-Research and Innovation Centre in Information, Communication and Knowledge Technologies	**Dimitris Gavrilis, Costis Dallas, Dimitra-Nefeli Makri**
AVINET Asplan Viak Internet AS	**Stein Runar Bergheim**
IPCHS Institute for the Protection of Cultural Heritage of Slovenia	**Franc J.Zakrajsek**
UPV/EHU Universidad del País Vasco	**Aitor Soroa, Arantxa Otegi**
VUKF Vilniaus Universitetas	**Rimvydas Laužikas, Ingrida Vosyliūtė**

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	**Public**	**x**
C	**Confidential, only for members of the consortium and the Commission Services**	

Revision History

Revision	Date	Author	Organisation	Description
Draft	18.11.2014	Task leader T3.2-T3.5	AIT, EHU, IPCHS, AVINET, VUKF	Contributions T3.2, T3.3, T3.4, T3.5
Draft	19.11.2014	A.Höller	AIT	Integration of contributions from T3.2, T3.3 and T3.5
Draft	25.11.2014	G. Koch	AIT	Introduction, References, Revision
Draft	25.11.2014	K. Fernie,	2CA	Review
Version 0.1	27.11.2014	Task leader T3.6, G. Koch	Athena, AIT	T3.6 contribution, executive summary
Version 0.1a	28.11.2014	H. Wright	UoY ADS	Review
Version 1	1.12.2014	G. Koch	AIT	Adding T3.5 license information, internal review comments

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This deliverable contains original unpublished work except where clearly indicated otherwise.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.

Contents

Figures and Tables

1 Executive Summary

The LoCloud project is dedicated to providing services for cultural heritage institutions. These services support the easy ingestion of cultural heritage content into Europeana[1]. In addition, the services help content providers capture metadata and enrich the data in order to increase the potential semantic linkage of the data within the World Wide Web. It is one of the major aims of LoCloud to reduce technical, semantic and skills barriers for providing content, and therefore cloud technology is used and cloud-based services have been implemented for facilitating the aggregation of digital content from small and medium cultural institutions.

This deliverable provides a summary status report of the microservices that have been developed for the LoCloud project. Microservices are small, lightweight software services that each perform a set of narrowly defined functions. The various services can be arranged in independently deployable groups and can communicate with each other via defined interfaces. The distributed architecture makes it easier to deploy new versions of services, and to scale the development. In the LoCloud project the development effort for the various services was organized around multiple development teams realizing five main groups of microservices. Among these services are geolocation enrichment tools, metadata enrichment and vocabulary services, historic place names and Wikimedia crowdsourcing services. These microservices can be used and combined as needed by the content providing institutions. All microservices have been implemented on virtual machines in a cloud testlab (using the OpenNebula cloud computing platform) and all services are also integrated in the LoCloud aggregation platform MORe[2].

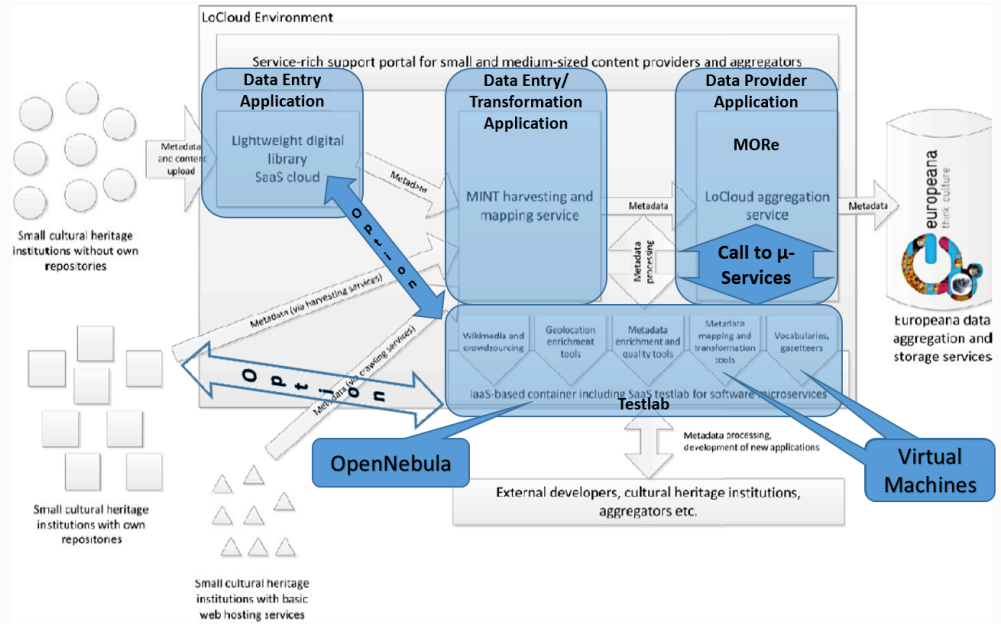


Figure 1: Options for using the microservices during the metadata ingestion workflow

[1] http://www.europeana.eu 27th November, 2014
[2] http://more.locloud.eu Metadata & Object Repository, 27th November, 2014

Figure 1 shows some of the various possibilities for implementing and using the LoCloud microservices. A generic enrichment service was designed that integrates the microservices into the MORe aggregator. Within the platform, users can choose which services should be called to enrich the metadata after transformation, or they can choose to capture data from selected Wikimedia resources. Besides using the services in the MORe aggregation facility, it is possible to implement individual microservices in a local cataloguing application through the available API's of the services. Some of the LoCloud microservices have even been deployed as dedicated online tools. Among these are the geolocation, vocabulary and historic place names applications.

The LoCloud cloud-based SaaS[3] services serve the cultural heritage community in many ways. The enrichment services help to improve data quality for the content providers and Europeana. The data capturing service allows re-use of already existing online cultural content and allow its proper ingestion into Europeana. The scalability and various options for deploying and using the microservices help tailor the services offered for cultural heritage institutions, and to meet their needs and capacities.

[3] Software as a Service, http://en.wikipedia.org/wiki/Software_as_a_service 27th November, 2014

2 Introduction

2.1 Purpose of this document

This consolidated report provides an overview on the current status of the microservices available through LoCloud for public and professional consumption. Five different types of microservices have been developed and each microservice is described in detail in the following deliverables:

D3.2 Geolocation Enrichment Tools
D3.3 Metadata Enrichment
D3.4 Vocabulary services
D3.5 Historic place names service
D3.6 Wikimedia and crowdsourcing

The deliverables may be downloaded from the LoCloud project website at: http://www.locloud.eu/Resources/Deliverables.

A first version of each service was developed and the implementation up to the end of August 2014 is described in the deliverables listed above. From September to November 2014, LoCloud content providers and technical partners tested the microservices. Test results and recommendations have been gathered and the services were updated accordingly.

All updates of the services have been documented in the relevant user and technical documentation available from the central LoCloud Support Centre (https://support.locloud.eu/tiki-index.php) and are referenced in the relevant deliverables.

2.2 Summary

The LoCloud microservices are a suite of independently deployable services dedicated to the capture and enrichment of metadata associated with cultural heritage data sets.

There are three major ways of using the various services:

1) The services invoked during the LoCloud aggregation process either on demand via the MORe system (http://store.locloud.eu) or automatically.
2) The services integrated in the content providers' local cataloguing environment via web services.
3) The services used via online tools that are hosted in a cloud-based environment.

The following table provides an overview of the various **usage possibilities** per microservice:

	LoCloud aggregation workflow (MORe)	**Integration in local cataloguing systems**	**Online Tool**
Geolocation Enrichment Tools	Geolocation API LoGeo	Geolocation API LoGeo web service	Geocoding application using LoGeo
Metadata Enrichment	Vocabulary matching Background linking		
Vocabulary services	Vocabulary services used in Enrichment Plans and Subject collections	Vocabulary web services	Experimental application
Historic place names service	Part of the metadata enrichment via the vocabulary services	Historic place names web services	Historic Place Names Online Tool
Wikimedia and crowdsourcing	Wikimedia application		

Table 1: Usage of the microservices

Within MORe, the LoCloud aggregation environment, the content providers can select which microservice should be applied to the transformed metadata. The content providers create an "Enrichment Plan" and add the desired services to the plan. The sample below shows an "EDM enrichment Plan" that invokes five LoCloud microservices.

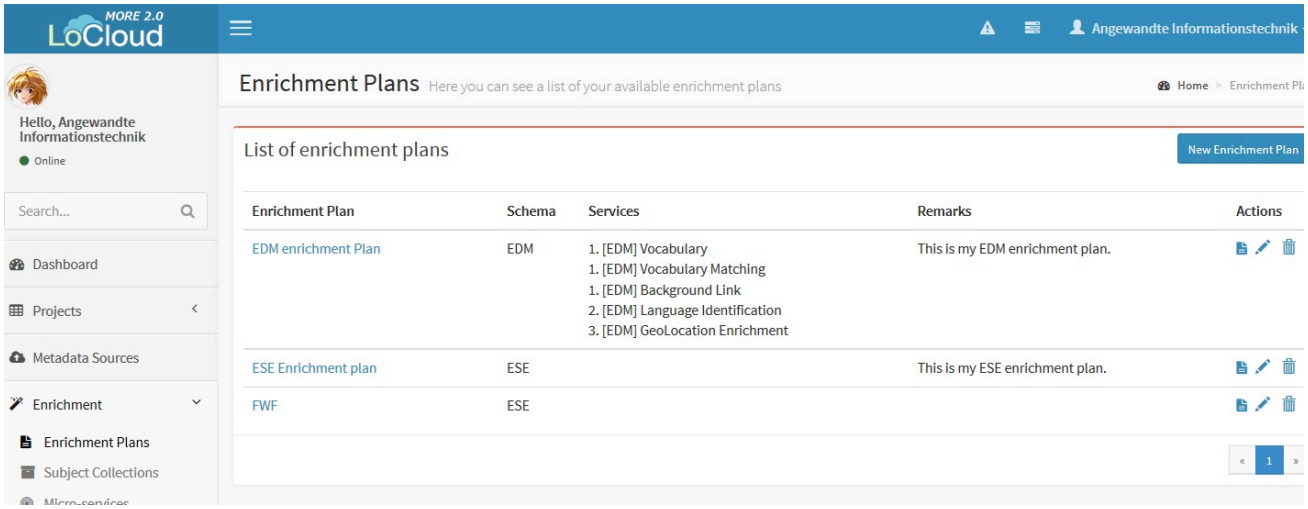


Figure 2: Enrichment services for EDM

The next table shows a summary feature overview of the various services provided:

	Service Feature	Service Feature	License Information
Geolocation Enrichment Tools	**Geolocation API (LoGeo)** The service provides name to coordinate resolution services based on a wide range of geographical name sources, including mainstream web sources such as Google and GeoNames – but also names collected from several of the national geographical names registries in the home-countries of LoCloud partners	**Geocoding application (using LoGeo)** The application allows institutions to add geographical locations to any existing content in a controlled crowd-sourcing environment – and then load the augmented data back into their original production databases.	GPL
Metadata Enrichment	**Vocabulary matching** The vocabulary matching microservice automatically assigns relevant concepts and terms from selected vocabularies to cultural heritage items.	**Background linking** The background link microservice (for English and Spanish) links cultural heritage items to DBpedia elements.	Apache License 2.0
Vocabulary services	**Vocabulary provision** The microservices provides an extendable set of vocabularies that can be called and used in the LoCloud aggregation process or in local cataloguing environments	**Vocabulary administration creation** This is an experimental cloud-based application that can be used to easily create or extend existing vocabularies in a collaborative way and integrate them in the LoCloud aggregation process.	GPL
Historic place names service	**Historic place names** **Place verification and proposal**: This microservice helps verify the availability of a historic place name in the LoCloud HPN Thesaurus; enhance existing historic place data by interconnecting historical place names with contemporary place names, its linguistic variations and geographical coordinates; propose new historic place names for the HPN Thesaurus	**Historic place names** **Integration of data:** With the tool the user can analyse and enrich local historical geo-data with existing HPN Thesaurus; export selected LoCloud HPN Thesaurus data sets and integrate them into the local information system; supplement LoCloud HPN Thesaurus with local data	GPL

	Service Feature	Service Feature	License Information
Wikimedia and crowdsourcing	**Wikimedia application** The service supports the harvesting of content from Wikimedia installations and the automatic update of enriched content back to Wikimedia.		Based on Wikimedia Commons Licensing

Table 2: Services Overview

Chapters 2 to 6 in this document provide an overview on the current status of each microservice's development. The benefits and usage possibilities of the services are described, and information is provided on how to access the user and technical documentation. In addition, we also describe some of the lessons learned during the development and testing phases. The LoCloud project has just passed its midpoint, and usage of the services from now on will provide additional feedback.

From the start, the LoCloud microservices have generated considerable interest in the content provider network and the Europeana community. Ideas for further development and sustainability have been included in the conclusion of each chapter.

3 Geolocation enrichment tools

The LoCloud geolocation enrichment tools include two tools. The first is a geolocation API (LoGeo) that provides name to coordinate resolution services, based on a wide range of geographical names sources including mainstream web sources such as Google and GeoNames – but also names collected from several of the national geographical names registries in the home-countries of LoCloud partners.

The second tool, the "Geocoding application" is an end-user oriented application that allows institutions to add geographical locations to any existing content in a controlled crowd-sourcing environment – and then load the augmented data back into their original production databases. The GeoCoding application includes LoGeo geolocation API search capabilities.

In most of the sections below, the two tools are described under separate sub-sections to allow for greater clarity for the reviewers who might wish to relate this deliverable back to D3.2. However, with respect to the conclusion and outlook section at the end, the future prospects of the two applications are linked.

3.1 Development overview

The development of the geolocation enrichment tools is described in detail in D3.2 "Geolocation Enrichment Tools". This section describes how the tools have continued to evolve.

3.1.1 LoGeo API

The first version of the API, LoGeo_1_1, was published on the 27th of August 2014, with a minor update published on the 15th of November 2014, LoGeo_1_2. The LoGeo API and console are available at the following links:

- LoCloud API: http://locloudgeo.eculturelab.eu/LoGeo_1_2/loGeo.aspx
- LoCloud API console: http://locloudgeo.eculturelab.eu/Console_LoGeo_1_2_m/

The improvements of the LoGeo API since the first release are:

- The database of place names has been updated. The 258,418 places within the United Kingdom (not including Northern Ireland) have been added to previously used national gazetters of settlements (including also small setlements) and other geographical places.
- The input place name can be more flexibile, e.g. Paris, Île-de-France, France. A hierarchy of place names can be submitted. If the place name on the first level is not found, then a search is performed on the second level.
- The efficiency of the LoGeo API console map was improved (speed of map display, numbering of scores etc.)

3.1.2 Geocoding application

Asplan Viak Internet AS (AVINET) developed the Geocoding application. From its first release, it was made available as a public repository on the social coding platform GitHub. The code was released under the terms of the GPL license, and may be accessed from the repository: https://github.com/runarbe/LoCloud.Geocoding.

A demo-site where the latest version of the Geocoding application is running may at all times may be accesed at http://locloud.avinet.no/demo.

Following the release, the application has been subject to testing by partners and included as part of the regional training workshops of the LoCloud project. It has been most extensively tested in Norway by the National Archive (Riksarkivet).

3.2 Impact and benefits for the community

Both the geolocation API and the geocoding application make it easier to add spatial locations to content. With the growth of the Europeana information space, it becomes increasingly important to be able to find ways of facilitating access, in addition to a continued emphasis on the aggregation of content.

The "WHERE-dimension" of metadata is an important way of facilitating access to deep content that otherwise would not be found, unless explicitly searched for. However important, most cultural heritage content includes location only by reference, i.e. geographical names instead of coordinates. While such references may be read and interpreted correctly by individuals who have expertise in the subject matter, or are contextually aware, the same may not be said for computer applications that can only distinguish between text patterns.

In order to resolve location references into unambiguous spatial coordinates, it is necessary to use tools like LoGeo and the geocoding application. Both of these tools are simple and intuitive and significantly leverage the user-threshold compared to existing tools and services that are technically sophisticated but very complicated to use.

3.2.1 LoGeo API

- the LoGeo API is designed for the recognition of the place name from the cultural heritage metadata
- the LoGeo API can be simply integrated and used in any other LoCloud microservices and applications.

3.2.2 Geocoding application

The take on metadata enrichment in the context of Europeana has so far mainly been on up-stream auto-enrichment in aggregators' repositories – or augmenting metadata ingested into Europeana with implicit values where metadata may be missing. This approach has facilitated the implementation of faceted search and similar exploration mechanisms. However, with respect to geolocation enrichment, this approach has a clear limitation in that the name-to-coordinate

resolution is not quality assured – and in some cases not even persistent between two ingestions of the same underlying content.

It is obvious that those who are best able to evaluate a geocoding match are not robots or generic experts on an aggregator or Europeana level; they are those who have subject-matter expertise and who are aware of the spatial context of the collection that is being geocoded.

There are justifiable objections to doing metadata enrichment on local level; it is expensive, time-consuming and the institutions have neither the time, nor the money, nor the tools required.

- With the geocoding application, LoCloud offers an easy entry-level tool.
- The tool is released under a GPL license and can be run using very modest hardware from mainstream hosting providers with costs in the range of 5-10€ per month.
- The tool offers a platform for engaging users to aid in geocoding content. This addresses the third barrier to local geocoding.

It is also typical of upstream auto-enrichment that content resources are attached to coordinates with very coarse spatial resolution, i.e. an object sits in a museum in Berlin, and hence it is linked to the spatial coordinates of Berlin. This is suitable for exploration of content on a map of Europe – but makes significantly less sense when explored on a mobile device while walking around IN Berlin.

The geocoding application permits geocoding of any metadata record with a point, a line or a polygon – and it promotes the capture of very high spatial resolution coordinates through global high-res imagery sources and basemaps. Resolution is important, as the majority of Internet access now comes from mobile, location aware devices.

3.3 Usage of microservice within LoCloud

The two geolocation-enrichment tools have different target user groups. The LoGeo API is a web service consumed by other applications in order to provide name-to-coordinate resolution and does not implement any stand-alone user interface (except for testing purposes). The geocoding application, on the other hand, is targeted at the public as end-users. As such, it would be appropriate to say that the LoGeo API is targeting professional use through application developers whereas the geocoding application is a public end-user service.

3.3.1 Public consumption

The testing of the Geocoding application has so far mainly been by LoCloud partners. In one instance it is already being used for production purposes, this is by the National Archive of Norway, who uses the application as part of their workflow to convert the Norwegian census of 1910 to a 5-star Linked Open Data resource.

The records of the 1910 census have been converted fom their original analogue form into digital records through a massive data-punching effort. The records contain references to street addresses for urban areas and land parcels for rural areas.

First, the data has been run through an auto-enrichment process based on a complex set of semantic rules and match prioritizations. Thereafter, data that yielded no or poor matches have been loaded into the geocoding application to be augmented by interested private individuals who reside in and have extensive knowledge of the areas in question. These individuals use the application to link unmatched records to geographical locations and the data is then downloaded and updated in the original data source.

The geocoding application has been an instrumental part of this work, and as is evident from section 2.4.1, has consequently been improved in terms of robustness and scalability.

3.3.2 Professional consumption

Whereas the geocoding application requires user interaction, the LoGeo API is a microservice that can be consumed by other services and applications and be used for auto-enrichment-workflows - or to provide geographical names search in electronic forms or web applications.

The LoGeo API could be directly used within the other microservices for small and medium institutions (WP3) such as the geocoding application, the metadata enrichment service (see Chapter 4 below), and the historic place names service (see Chapter 6 below).

The API is integrated both in the data ingestion workflow from content-providers to Europeana in the MORe repository – and as a search provider in the geocoding application.

3.4 Final status

3.4.1 LoGeo API

LoGeo API was invoked 481 times from the 1.9.2014 until 16.11.2014. During this time, no significant bugs, problems and demands for the changes of the API were identified.

3.4.2 Geocoding application

There are at present four running instances of the geocoding application. Three used for testing and production purposes on real-world data, and one for demonstrating the capabilities of the latest version of the application at all times.

The table below shows (1) the number of datasets uploaded to each of the instances, and (2) the number of registered users.

	Instance 1	Instance 2	Instance 3	Instance 4 (demo)
(1) Datasets	38	94	87	32
(2) Users	14	13	13	39

Table 3: Data statistics from the geocoding application

As is evident from the table above, the application has been extensively tested with real-world data, alebit by a limited number of users. However, through the testing following the first release, AVINET has collected important feedback from the pilot users.

While developing a system, it is common to work with a controlled sample of well-known data. However, only when the application is used with unknown data sources, will more subtle bugs and issues surface.

Issues that were identified and addressed during the testing include:

- User management (rights to add and delete users have been restricted to specific roles). An initial bug resulted in users with a basic user level being able to elevate themselves or other users to administrators.
- Self-registration results in users being added to a group that may contribute to existing geocoding projects but that may not create new ones and administer users unless explicitly authorized/elevated by a system administrator. This restriction will be lifted once the service goes public.
- A bug in the upload module caused import of larger CSV files to fail without a notice.
- A feature to log the zoom level of the map along with the recorded coordinate was added in order to detect the quality of the metadata.
- A facility to record lines and polygons as well as points was included to handle locations of records that were not possible to accurately represent by means of a single XY coordinate pair.
- Default search databases were extended to include the Geonames API for names and Wikipedia entries as well as the LoGeo API.
- Facility to add WMS layers to maps was added to support other basemap providers than the defaults shipped with the application.
- Facility to manage access to datasets added in order to permit many datasets being geocoded simultaneously by different user groups in the same instance of the geocoding application.

The improvements address both workflow-related bottlenecks that must be overcome in order to successfully deploy the application in a professional environment, as well as addressing any technical issues.

3.5 Lessons Learned

This section summarizes the lessons learned from challenges that had to be overcome throughout the implementation of the geolocation enrichment tools.

3.5.1 LoGeo API

- Geographic location is one of the most important attributes of every cultural heritage item. It can describe provenance, the current institution, the location of the event or other related events.
- Formalized location attributes (e.g. geocode or geographical coordinates), even if they are using a finer level of spatial detail; for example county or town, are indispensable for enhancing the power of searching, raising the usability for users and enabling the visualization of the content in extensive international portals as is Europeana for example.
- Besides the technical tools for geo enrichment, the conceptualisation of the geolocation of a certain digital object is also very important (e.g., what are the geocodes of a book?).

3.5.2 Geocoding application

The balance between “feature-richness” and “simplicity” is difficult to achieve when trying to bring together the professional domain of collection management with that of geographical information sciences.

On the one hand, we would like to make it as simple as possible to add geolocations to existing content – on the other we would like to give interested users a powerful tool that will enable them to create high-quality data that can be re-used in high-resolution geo-aware applications.

Geocoding as a concept is not unique to the LoCloud project. Geocoding is a technique that has been around since businesses first tried to convert archive records linked to street addresses into point data for service planning purposes in the mid-late 1980s. However, it is a technique that has not been accessible to non-GIS professionals.

Tools and services that exist are all targeted at either programmers or people with above-average knowledge about software technologies. Together with the LoGeo API, the suite of tools offered by LoCloud bridges the MLA community with that of GIS, offering a simple user-friendly solution for geocoding, crowdsourcing and quality assurance of geolocations for any form of record based data.

3.6 Available documentation

The latest documentation for the Geolocation enrichment tools is available in the Wiki at this link: http://support.locloud.eu/Geolocation%20Enrichment%20Tools

3.7 Conclusion

Both geolocation enrichment tools are efficient catalysts for geo-enrichment of digital metadata within the domain of cultural heritage.

The API and the geocoding application will both be promoted for use within other Europeana projects when they are enriching their metadata.

The LoGeo API may easily be integrated in the cultural heritage management systems/repositories, used by museums, libraries, archives, and other cultural institutions. The code developed within the LoCloud project is under a GPL licence, but the integrated 3rd party components can only be used with respect to the licences of their providers (eg. Google, OpenLayers, Geonames).

Open access to the geocoding application is secured through making the code publicly available under a permissive GPL license. Open access is however not the same as sustainability. AVINET will explore the potential of releasing the geocoding application under a dual license, permitting the establishment of a commercial service based on the tool in the longer term. This would be an important step towards long-term sustainability of the platform.

4 Metadata Enrichment

Metadata enrichment services automatically link Cultural Heritage (CH) items with information such as links to DBpedia pages or vocabulary concepts.

Figure 3: Example of a vocabulary match and background link for a Europeana item

Metadata enrichment is a useful task that assists users by providing contextual information about a particular item or by offering choices to help categorizing CH items into relevant vocabularies. For example, Figure 3 shows an example of a vocabulary match and background link for a particular Europeana CH item. The bottom of the figure shows the CH item describing ``The Major Oak'' as displayed by the Europeana portal[4]. The item has been enriched and linked to two different elements: the vocabulary concept ``Oak'' as defined by the Library of Congress Subject Headings vocabulary (the left side of the figure), and a Wikipedia page describing the same tree (the right side of the figure).

The LoCloud project provides a set of metadata enrichment microservices integrated into the aggregation workflow. Cultural providers may use these enrichment microservices and automatically annotate CH items with the relevant information. The microservices are throughly explained in Deliverable D3.3 "Metadata Enrichment". In this deliverable we report on the impact and benefits of the microservices as used in the LoCloud project, and how users and technicians can use the services outside Locloud.

[4] http://www.europeana.eu/portal/record/2022323/C052AA1727D9C258801CF676473953A0861A47C0.html

Generic enrichment service in MORe

In order to hide the complexity of the enrichment microservices from the users, a generic enrichment service was designed to integrate these services into the MORe aggregator in the best possible way. Another major advantage to having a generic enrichment service is that it can hide the complexity of the aggregation tasks from the microservice providers. This means that microservice providers do not need to build metadata schema specific services, or worry about aggregation specific taks etc.

In the LoCloud aggregation environment, MORe, the content providers can select which micro service should be applied to the transformed metadata. This is done during the enrichment step.

The first step is to create a new Enrichment Plan (**Enrichments** -> **Enrichment Plan** -> **New Enrichment Plan**).

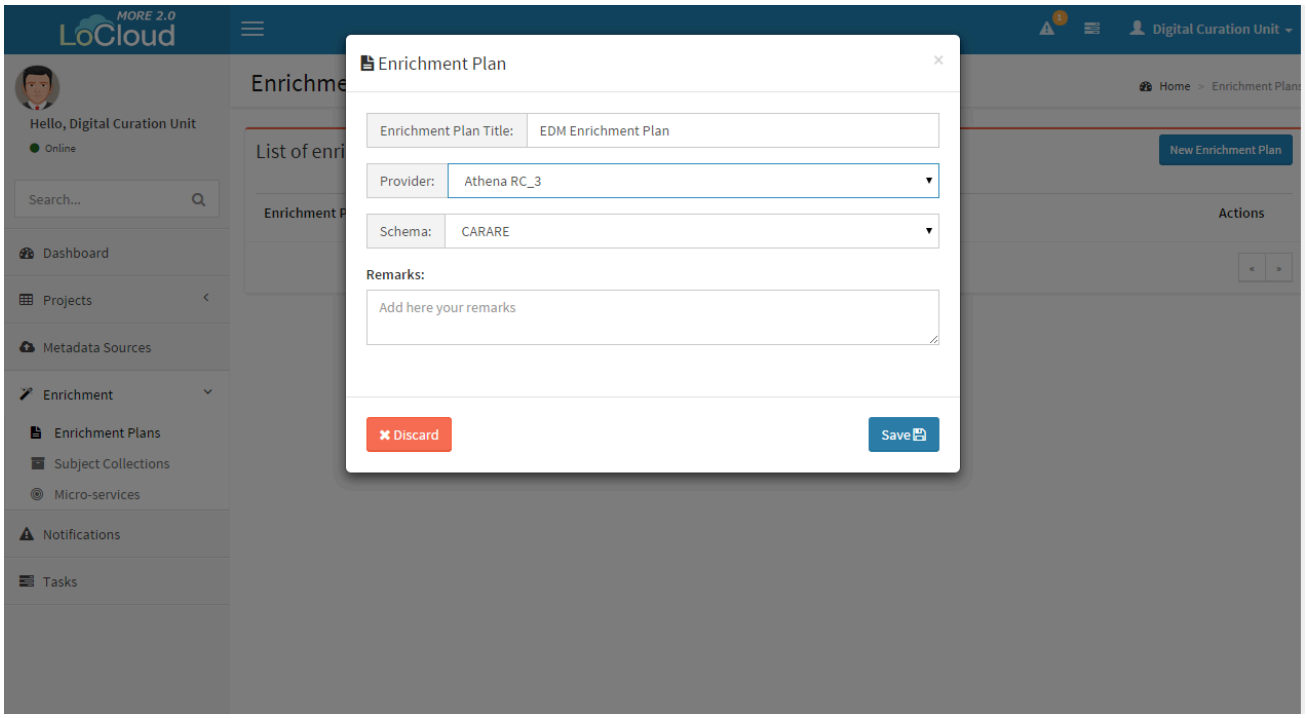


Figure 4: New Enrichment Plan

Now a service (**Add Service**) has to be added to the Enrichment Plan. All the microservices are available to be used in EDM schema.

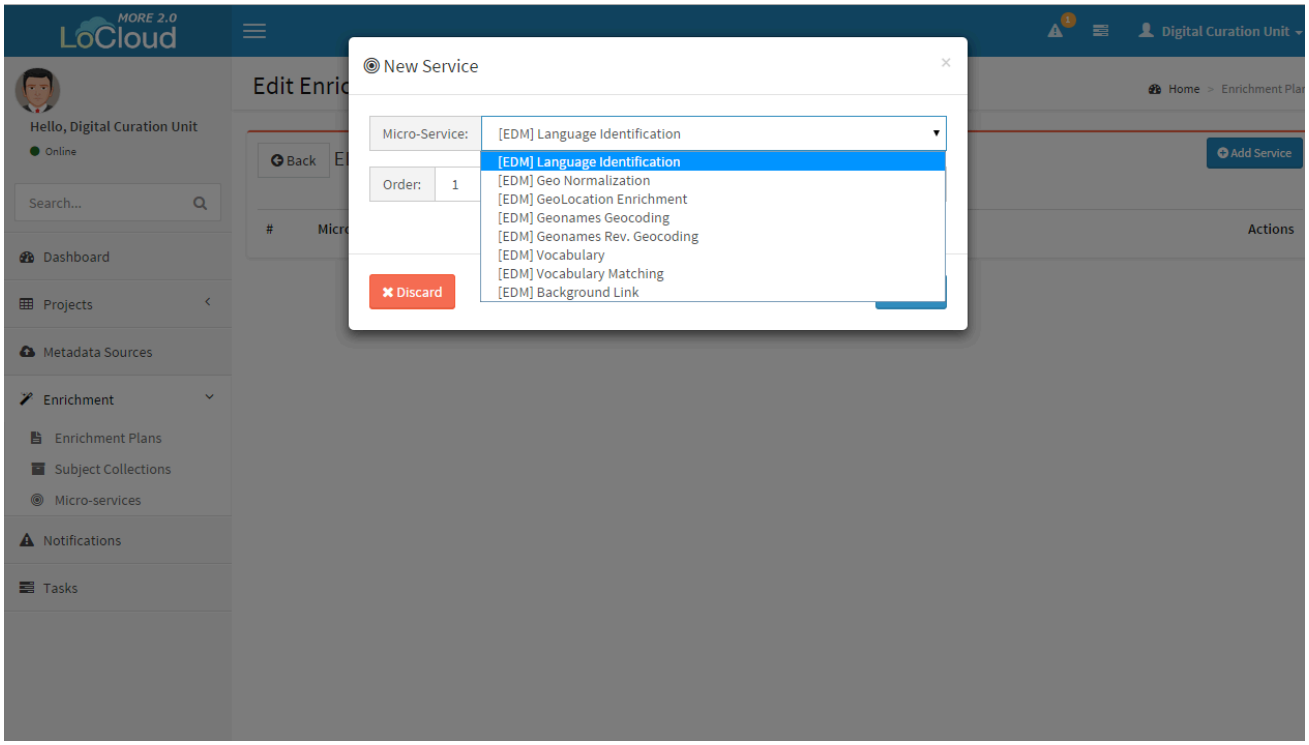


Figure 5: Add service

For some microservices, parameters need to be added. More specifically:

- *Geo Normalization* service, which normalizes in an appropriate way, the coordinates for the class edm:Place, takes as parameters the delimeter (, /, -, _) used to the records and the invert status (x instead of y, y instead of x).
- *Vocabulary service*, which allows users to create collections of thesauri terms and insert them automatically into all items of their aggregated packages, takes as parameter the Subject Collection.

For using the Vocabulary Service, the user first needs to create a new Subject collection (**Enrichments** -> **Subject Collection** -> **New Collection**) and has to add concepts from the various vocabularies to this collection (**Add Concept**).

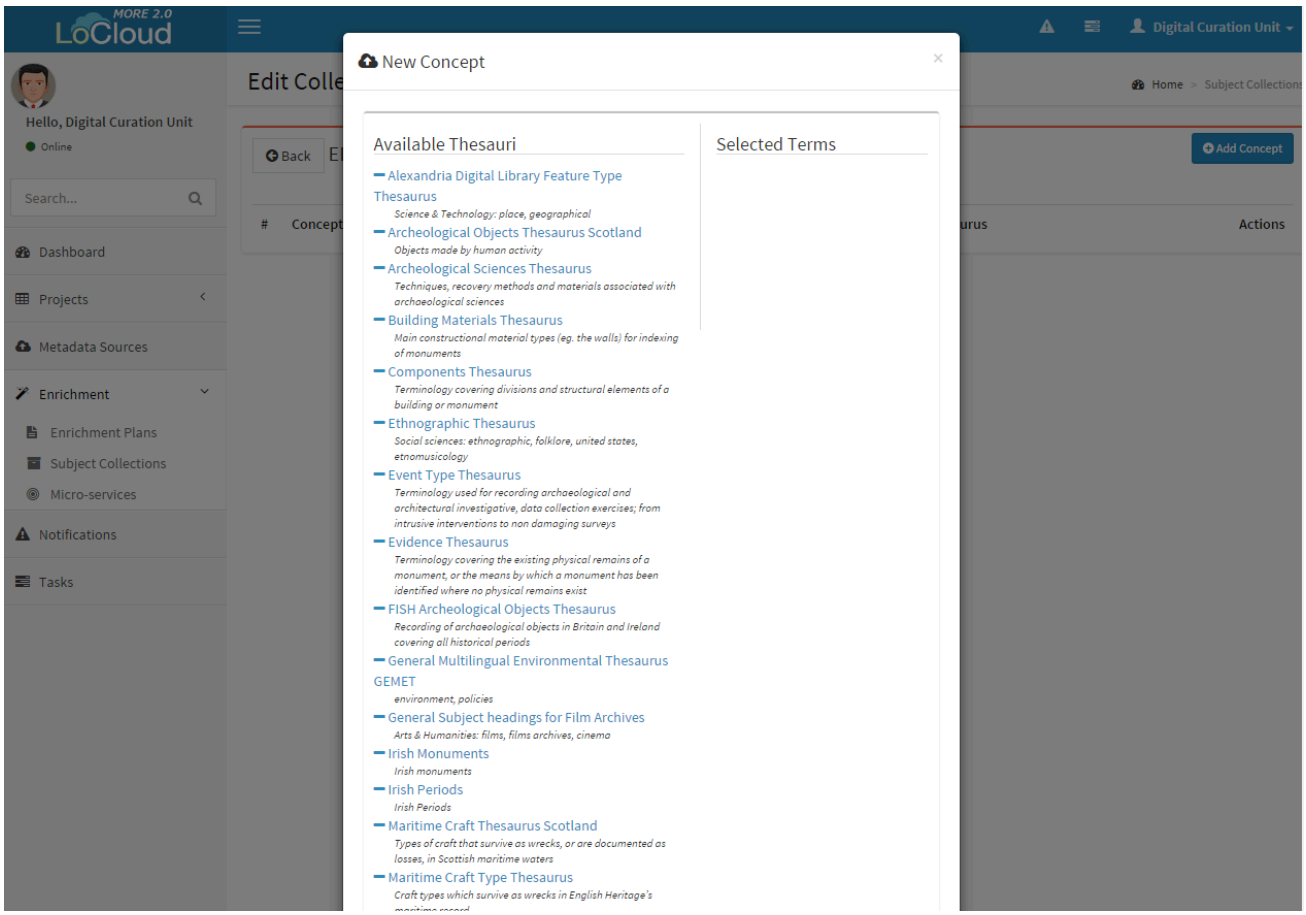


Figure 6: Add concepts to a subject collection

The concepts derive from all the vocabularies within the Vocabulary Service. The user can browse through the various vocabularies and select the most appropriate concepts that match to his dataset.

In general, all the microservices are integrated into the LoCloud aggregation environment through their APIs. The result of each microservice after enrichment phase is:

- The **GeoLocation service** adds coordinates and/or place names in the EDM record. More specifically, in the class **edm:Place**, the elements <wgs84_pos:lat>, <wgs84_pos:long> and <skos:prefLabel> are added.
- The **Vocabulary service** adds the selected subject to the records, through the element **dc:subject** in the class edm:providedCHO.
- The **Vocabulary matching service** matches the subject of the item with the appropriate vocabulary concept of the list of vocabularies. This information is not depicted in the EDM record yet but the user can see it in the “Enrichment Details”.

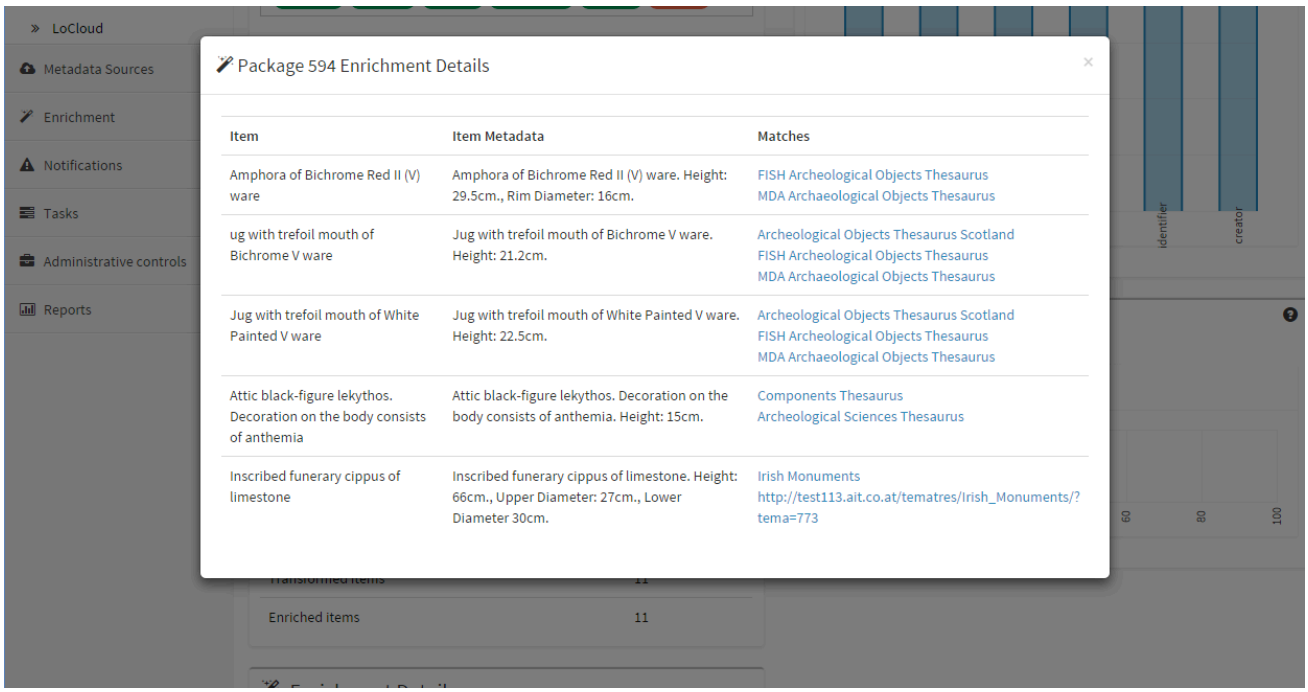


Figure 7: Vocabulary matching

- The **Background link service** matches the subject of the item with a subject of DBPedia. This information is not depicted in the EDM record yet but the user can see it in the "Enrichment Details".

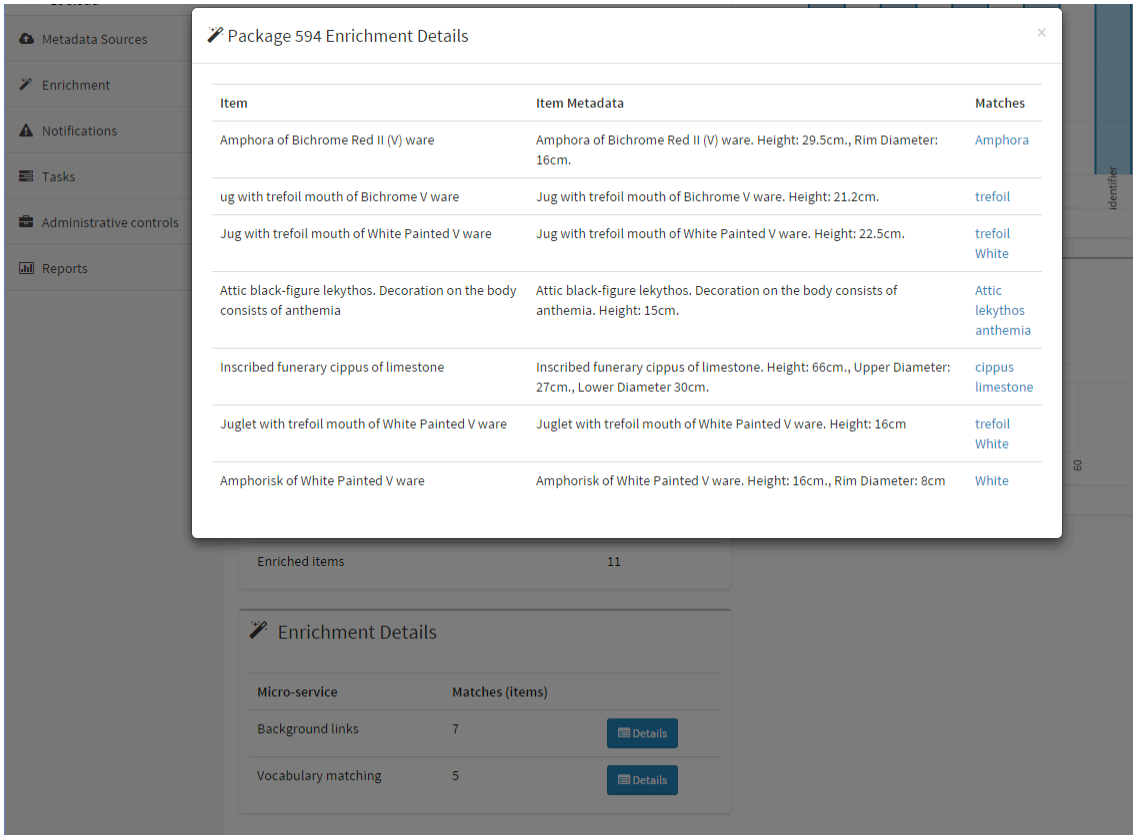


Figure 8: Background link

4.1 Development overview

The enrichment services comprise the background link microservice (for English and Spanish), which links CH items to DBpedia elements, and the vocabulary matching microservice, which maps CH items with relevant vocabulary concepts. All microservices are implemented as REST services and are deployed into virtual machines (VM). We have implemented the following three microservices for metadata enrichment:

- A background link microservice for English.
- A background link microservice for Spanish.
- A multi-lingual vocabulary matching microservice.

As explained in the previous section, the metadata enrichment microservices are used automatically in the various enrichment workflows through the LoCloud Generic Enrichment Service. In particular, the generic service sends textual descriptions to both the background link and vocabulary matching microservices (e.g. a title, description, etc) and integrates the returned information into an enriched version of EDM.

The background link problem is closely related to the so-called Named Entity Disambiguation (NED) task. Nowadays there exist many systems for performing NED, and therefore the microservice has been implemented using such a NED tool. Deliverable D3.3 provides a study of several state-of-the art systems on NED tools and datasets, and presents an evaluation of the chosen tool. The task of linking CH items with relevant vocabulary concepts has also gained much attention during recent years. However, to our knowledge there exists no tool or standard datasets to test them. We thus designed and implemented the vocabulary matching microservice from scratch.

4.1.1 Background link

The background link microservice is built upon DBpedia Spotlight package, version 0.6, which is publicly available[5]. Two instances of DBpedia Spotlight were used, one with the models trained for English and one with the models trained for Spanish, for the English and Spanish background link microservices, respectively [6].

Each DBpedia Spotlight instance is installed on a different virtual machine. The addresses of the virtual machines are the following:

- Background link English: http://test183.ait.co.at/rest/bglink
- Background link Spanish: http://lc013.ait.co.at/rest/bglink
- Vocabulary matching: http://test183.ait.co.at/rest/vmatch

[5] https://github.com/DBpedia-spotlight/DBpedia-spotlight/wiki/Statistical-implementation
[6] DBpedia Spotlight models are available here: http://spotlight.sztaki.hu/downloads/

A PHP wrapper was implemented to acts as the endpoint of the REST service. The wrapper listens to a particular port on the VM, reads the input as described in the API documentation, runs DBpedia Spotlight, and transforms the output to the required format, which is then returned.

4.1.2 Vocabulary matching

The aim of the vocabulary matching microservice is to automatically assign relevant concepts and terms from selected vocabularies to CH items as provided by cultural institutions. The service receives metadata information as input, and returns the items enriched with links to one or more relevant concepts from a list of vocabularies.

The vocabulary matching microservice requires a set of vocabularies to match the CH items against. Many schemes have been proposed to describe and manage cultural heritage data. Those schemes are usually found in the form of classification schemes, subject heading lists, etc, and many times they are focused on a particular field, institution and even collections. Within the LoCloud project, the vocabulary microservice (see chapter 5), a collaborative platform to explore the potential of crowdsourcing as a way of developing multilingual, semantic thesauri for local heritage content.

The vocabulary matching microservice consists of two modules. One module retrieves the vocabularies from the vocabulary service and creates an internal database with the concepts and lexicalizations in several languages. This module is executed on a regular basis so that the local database is synchronized and up to date with the vocabularies from the vocabulary service, whose concepts and terms are updated in a continuous fashion. Once the vocabularies are retrieved and stored in the internal database, a second module annotates the CH items with appropriate concepts and terms as found in the vocabularies.

Both modules have been designed and implemented for LoCloud. In addition, another PHP wrapper was built which actually implements the REST service, much like in the background link microservice.

4.2 Impact and benefits for the community

Current efforts for the digitisation of Cultural Heritage are providing users with access to vast amounts of material. Sometimes, though, this quantity comes at the cost of a restricted amount of metadata, with many items having very short descriptions and a lack of rich contextual information. For instance, the “Mona Lisa” item as displayed in Europeana[7] does not provide rich information associated with the item. Wikipedia, in contrast, offers in-depth descriptions and links to related articles for many CH items. Linking CH items to relevant Wikipedia or DBpedia articles allows the user to obtain a large amount of information related to the particular CH item. Linked Open Data (LOD) repositories are heavily inter-related, and thus the background linking can be seen as the first step to relate the CH items with many reference KBs of various types.

7 http://www.europeana.eu/portal/record/08501/BFEDA8C5F46930228355F50FA9F5298D982DD6B8.html?start=3&query=La+Joconde+mona+lisa&startPage=1&rows=24

Cultural Heritage institutions categorize their content by linking CH items with one or more relevant concepts from a controlled vocabulary. With the advent of information technology and the desire to make available CH resources to the general public, there is an increasing need to facilitate interoperability across these different contexts. There are many situations in which users may benefit from having the items organized into subject categories for browsing and exploration. For example, when users do not have clearly defined information needs, when attempting complex search tasks or when they want to gain an overview over a collection. In such cases the provision of only a simple search box is insufficient. This is particularly relevant to digital libraries where rich user/information interaction is common and requires alternative methods to support users. The provision for browsing functionalities through thesaurus-based search enhancements have all been shown to improve the search experience.

4.3 Usage of microservice within LoCloud

The metadata enrichment microservices are used in the various enrichment workflows automatically through the LoCloud Generic Enrichment Service, which allows the orchestration of various REST microservices into complex enrichment workflows.

4.3.1 Professional consumption

The code for both microservices is accessible under the following addresses:

https://github.com/ixa-ehu/locloud_bglink
https://github.com/ixa-ehu/locloud_vmatch

Both repositories contain detailed installation instructions that describe how to install and deploy the microservices.

As explained in the API documentation, the microservices are agnostic with regard to the particular schema used for describing the metadata information associated with the CH item. Instead, the microservices accept textual snippets, and return the most relevant vocabulary concepts and DBpedia links, respectively. Within LoCloud, the generic enrichment service is responsible for selecting the proper metadata information from the CH items (title, description, subject, etc), sending it to the microservices as textual snippets, and encoding the enrichment information received by the microservices into EDM fields. Users wanting to implement these services are thus not enforced to use any particular metadata schema for enriching CH items.

4.4 Final status and lessons learned

Both microservices have been successfully installed into virtual machines and deployed on the LoCloud cloud testlab at AIT. We have throughfully tested the services and measured the response times for several requests. Our conclusions in this regard are the following:

- Both microservices perform the enrichment quickly. In general, enriching small CH items (up to 100 words) is achieved in less than 5 seconds. The background link service is able to disambiguate 17 mentions per second, whereas the vocabulary matching performs better with circa 40 matchings per second.

- The vocabulary matching service needs some time for initialization and database loading. Currently the vocabularies retrieved from the server are not very big, with circa 40,000 concepts lexicalized in 55,000 different ways in four languages (en, fr, de, es). The time needed by the matching service is proportional to the vocabulary size. Therefore, should the vocabularies grow significantly, faster database initialization methods should be devised and implemented.

- A stress test was also performed by sending a large batch of text for enriching. The enrichment services required grow linearly with the size of the input. In general, both microservices were able to answer in time, and there were no significant failures.

4.5 Available documentation

The documentation of both microservices, including full API specification, is avaliable under this address:

http://support.locloud.eu/LoCloud%20Enrichment%20Microservice

4.6 Conclusion

We have successfully implemented the metadata enrichment microservices for the LoCLoud project. The microservices include a multilingual vocabulary matching and a background link microservice for Spanish and English.

The vocabulary matching microservices automatically use the concepts and terms developed in the vocabulary server, as part of 3.7 of the LoCloud project. Thus, automatic vocabulary enrichment of CH items using virtually any vocabulary and taxonomy can be performed. Cultural providers can curate the terms and concepts of interest in a collaborative way using the vocabulary server, or use the server to upload existing and well-established vocabularies in the CH domain. The vocabulary microservice will automatically load these concepts into the database and perform automatic matching of CH items using them.

Both microservices are used in a coordinated way by the generic enrichment module as part of the MORe platform. The generic enrichment service allows integration with the REST microservices into complex enrichment workflows. The user can create a workflow by selecting and combining the microservices he/she wants.

The complete code for both microservices is publicly available under Apache License 2.0 and can be used by third party partners, who may therefore implement the same functionality within different projects and scenarios. The code for both microservices is accessible at the following addresses: https://github.com/ixa-ehu/locloud_bglink and https://github.com/ixa-ehu/locloud_vmatch. The background link service is based on DBpedia, a free software project for performing Named Entity Disambiguation. The analysis described in Deliverable D3.3 compares several alternatives and concludes that DBpedia is the best tool for the task at hand. The vocabulary matching service is implemented from scratch.

5 Vocabularies and languages

Task 3.4 established vocabulary services and an experimental application for enabling local cultural institutions to collaborate in the development of multilingual vocabularies for local history and archaeology and allow the retrieval of terms to be integrated during semantic enrichment in the LoCloud aggregation process. This application is made available for the local users as a cloud-based service.

The vocabulary services are used:

a) in the various enrichment workflows automatically through the generic enrichment service (Vocabulary provision)

b) through the aggregator user interface (Vocabulary provision)

c) in local cataloguing systems via web services (Vocabulary provision)

d) with a cloud based online tool (Vocabulary creation and import)

5.1 Development overview

At http://test113.ait.co.at/tematres/unesco/index.php a test installation with a multilingual thesaurus was set up for the LoCloud staged testing phase in spring 2014. Furthermore 30 vocabularies have been established with TemaTres (for example UNESCO, UKAT or PICO Thesaurus). These vocabularies are used by the enrichment services developed in LoCloud (see chapter 4 above). Different extensions have been added to the tool TemaTres, for example an importer for multilingual vocabularies in SKOS and plain text format. For simplifying the import process of multilingual vocabularies a new identifier for concepts was introduced. Furthermore it is now possible to add geographical coordinates via the "Notes Editor" and switching between the vocabularies (see also 5.5). Administration functionality was added to allow users to add a new vocabulary or delete an existing vocabulary.

5.2 Impact and benefits for the community

The tool supports the creation and updating of (multilingual) thesauri for local history and archeology as well as the retrieval of terms to be integrated during input/semantic enrichment in the LoCloud aggregation process.

5.3 Usage of microservice within LoCloud

5.3.1 Public consumption

The generic enrichment service of LoCloud orchestrates the various REST microservices into complex enrichment workflows that can be managed by the users individually. The user can determine a workflow by selecting and combining the microservices. More specifically for the vocabulary service the user is able to select terms to attach to a harvested package (subject collections) in the LoCloud aggregation platform MORe. The user can browse the various thesauri, searching and navigating through the concepts, and can finally select the concepts that shall be

included in the enrichment. A more detailed description on how the vocabulary matching and concept selection is done in the MORe environment can be found in chapter 4, Metadata enrichment.

5.3.2 Professional consumption

Professional users, for example content providing organisations (like archives, museums, etc.) that have their own catagloguing systems and want to create and use (existing) vocabularies have two options to use the microservice:

a) **Invoking LoCloud vocabularies via web services:**
A complete technical documentation regarding the APIs can be found at http://support.locloud.eu/tiki-index.php?page=Vocabulary+API+technical+documentation.
The vocabulary web services of the tool can be used for the integration of the vocabularies into remote local systems. And vice versa it is possible to integrate existing concepts from remote vocabularies via the web services into the cloud based vocabulary tool. Therefore the webservice calls *"import"* and *"linkTerm"* were added to the TemaTres tool.

b) **Creating and contributing own thesauri:**
The process of creating an own thesaurus with TemaTres is described under http://support.locloud.eu/Vocabulary%20application%20user%20documentation. The vocabularies that are created by the users may be integrated in the LoCloud vocabularies microservice and can be used from other parties as well. As the tool is available in a cloud-based collaborative environment, the various users may also cooperate in the development of certain vocabularies. A standard format for importing and contributing existing vocabularies is SKOS[8] but it is also possible to import text files (tabulated and tagged text) into the tool.

5.4 Final status

The testing process consisted of two testing phases. For the testing the UNESCO thesaurus under http://test113.ait.co.at/tematres/unesco/index.php has been used. It can be accessed with log in credentials that are available on demand.

The goal of the first test phase (July/August 2014) was the editing of a monolingual vocabulary with the LoCloud experimental vocabulary application and verifying the tool's functionality. In the second test phase (October 2014) a multilingual vocabulary was created through the import function and was then edited. The new feature of adding a geographic note was validated.

For each test phase a testing result template was sent to the testing participants to receive feedback for further improvement of the tool. The feedback concerned primarily minor bugs of the TemaTres tool (e.g. missing "synonyms", no warning when creating homonyms or excluding

[8] http://www.w3.org/2004/02/skos/ 20th November, 2014

candidate terms) and the difficulty of switching between the different LoCloud vocabularies. Moreover, some issues when importing or exporting a vocabulary have been mentioned.

Multilingual thesauri are challenging and there exist different approaches with varying benefits and challenges to manage them. A rather simple approach is to have a model with a central vocabulary and then add terms in different languages to make it multilingual. But this is more a translation of an existing vocabulary into other languages. Another, more open, approach is to have a federated model and regard any term as its own concept and connect concepts from different languages by an equivalence relation. LoCloud has chosen the latter approach in the implementation. Besides a few technical bugs found by the testers, one issue was the complexity introduced in the user interface. We are addressing this in the lession learned below.

5.5 Lessons Learned

After the testing the vocabularies are now available with the new namespace vocabulary.locloud.eu. This web address (http://vocabulary.locloud.eu) is also the new access point to view the LoCloud vocabularies and enter the experimental application.

Testing showed as well the need to streamline certain administration processes in the tool. Due to the structure of creating independent multilingual vocabularies that are connected via concept identifiers, the process of creating a multilingual vocabulary could be rather time-consuming when it is done from scratch. To accelerate the process for the users several new features have been implemented in the baseline TemaTres software: The user interface is now being extended to allow the creation of a term in another language. This should combine the currently necessary steps of creating the foreign term and connecting it with current language into a single operation.

Quick guides for creating quickly simple vocabularies, for creating a multilingual vocabulary and for import and export have been created.

Concerning switching between vocabularies a new feature has been added to the tool (see Figure 9). When logged in the tab *"My vocabularies"* appears in the top menu. A click on it displays a list of all LoCloud vocabularies for this user and the menu allows switching among these various vocabularies.

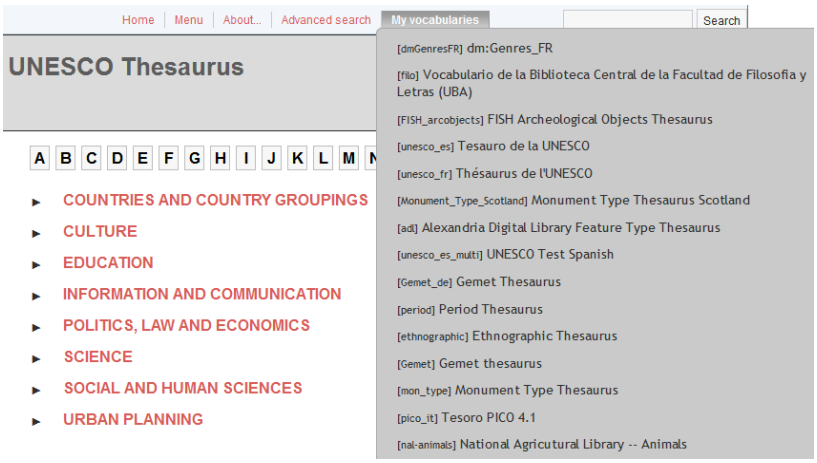


Figure 9: Switching between vocabularies

5.6 Available documentation

There are a number of sources to get additional information and documentation on the tool:

- Vocabulary Entry page: http://vocabulary.locloud.eu/
- LoCloud Support Center: http://support.locloud.eu/tiki-index.php
- Downloading technical documentation: http://support.locloud.eu/Vocabulary%20API%20technical%20documentation
- Downloading user documentation: http://support.locloud.eu/Vocabulary%20application%20user%20documentation
- FAQ Section: https://support.locloud.eu/qna/

Furthermore, there is a course on a "Mediathread"[9] platform (http://mediathread.ait.co.at) available. Log-in credentials are available on request.

5.7 Conclusion

The vocabulary repository is going to grow continuously due to the ongoing contribution of user created or extended vocabularies. In addition, new standard vocabularies available in SKOS format shall be imported continuously. In future we will need to establish an organizational procedure for integrating the new vocabularies to the LoCloud metadata enrichment process. Increasing usage of the experimental application might also lead to the addition or upgrade of features.

5.7.1 Terms of use

The experimental vocabulary application (eva) is freely available under GNU General Public License version 2.0 (GPLv2). This is the same licence as for the standard TemaTres tool. Once the LoCloud developments have been finalized and fine-tuning has finished it is envisaged to contribute them to the TemaTres repository (http://sourceforge.net/projects/tematres/).

[9] http://ccnmtl.columbia.edu/our_services/tools/mediathread/ 6th November, 2014

6 Historic place names

6.1 Development overview

Historic place names (HPN) microservice has been developed in collaboration between Vilnius University Faculty of Communication, VšĮ "Atviro kodo sprendimai", Digital Curation Unit of the Institute for the Management of Information Systems of Athena Research and Innovation Centre in Information Communication & Knowledge Technologies, Angewandte Informationstechnik Forschungsgesellschaft mbH (AIT) and Javni Zavod Republike Slovenije za Varstvo Kulturne Dediscine (ICHPS).

The development of HPN microservice was based on Agile Unified Process methodology and consisted of 7 development stages. Needs analysis and initiation were performed during the preparation of LoCloud project proposal (1). Requirements analysis and concept development was carried out as part of "LoCloud 1.4.1 Content evaluation and planning survey". Based on this investigation a concept of HPN microservice was developed during the LoCloud project face-to-face meetings (2). A work plan was prepared during LoCloud plenary meeting and Skype sessions (3). VUFC and VšĮ "Atviro kodo sprendimai" led the design (4) and development process (5). The HPN microservice was peer reviewed and tested by LoCLoud partners (2Culture Associates, AIT, Provincie Limburg) and the representative of Europeana (6). Implementation of microservice was realized via Internet by custom Django application named "united_geonames" (7).

6.2 Impact and benefits for the community

In digital heritage systems Historic Place Names serve as a link between current and past place names. Historic Place Names ensure interoperability of the current world and digitised information about past places, interoperability of "real" and "digital", internal interoperability within the information system and external systems leading to more efficient communication of digital data in general.

However, by making available current and former place names we can acknowledge different perceptions of historical spaces, increase participation by different cultural groups and help create a European narrative. In this sense HPN microservice could be used to aggregate, preserve and improve the interoperability and semantics between historical geo-information and contemporary geo-data, and also be a source of historical geo-information that could be accessed in the descriptive information of cultural heritage objects.

The service could also serve as a browsing assistant within Europeana, LoCloud and CARARE databases. HPN microservices also could be used as a research tool that provides information and contextual knowledge about historical place names and other places.

The potential of the HPN microservice could be realized by using following functions: crowdsourcing and enrichment of providers' and aggregators' historical geodata, HPN geodata visualization and HPN interoperability provision.

The benefits of HPN microservice for the community should be understood as better data interoperability between system(s) helping local cultural institutions and research organizations to become more effective in managing data.

6.3 Usage of microservice within LoCloud

6.3.1 Public consumption

HPN microservice is accessible online via: http://hpn.aksprendimai.lt/unitedgeo/. It is public and free to use. The service in orientated towards digtital content providers and aggregators and people who are interested in history, geography and historic geography.

The functionalities of the microservice vary depending on the type of user (registered and non-registered users). Registration is open to general public, however its functionalities are orientated towards professional users and require specific knowledge of the field or technical skills. Non-registered users constitute any public consumption of the microservice. The process involves end-users and HPN microservice administrators. Non-registered users can perform the following tasks:

- To verify the availability of historical place name at LoCloud HPN Thesaurus.
- To enhance existing historical place data by interconnecting historical place names with contemporary place names, its linguistic variations and geographical coordinates.
- To propose new historic place names for the HPN Thesaurus.

The microservice administrator must approve all enhancements and propositions made by end-users.

6.3.2 Professional consumption

Professional users of the HPN microservice could be digital heritage database administrators, content providers and aggregators, memory or research organizations and especially budget-limited cultural institutions that cannot afford to invest in development of high quality computer infrastructure. There are a few options for registered users with a professional interest in using HPN data:

- To analyse and enrich local historical geo-data in the database with existing HPN Thesaurus information about interconnected contemporary place names, different historical and linguistic place name forms and geographical coordinates.
- To export selected LoCloud HPN Thesaurus data sets in computer understandable format and integrate it into the local information system.
- To supplement the LoCloud HPN Thesaurus with local data exported in any computer understandable format by sending it to HPN microservice administrator.

6.4 Final status

Final status of the HPN microservice is considered to be version V1.2 following two previous versions (V.1.0 Initial release and V1.1 Additional search method). V1.2 includes microservice improvements that were made after testing and comments received from peer-reviewers. The current version of the microservice includes three types of changes:

- More friendly user interface (e. g. description of fields, “Register/Login” and “Logout” buttons, greeting text).
- Improved usability (e. g. “New search” button, obligatory field “Remark” made mandatory, case-insensitive search).
- Additional information (e. g. Original source field, geographical administration (Country, Subregion, Region), Synonim ID).

6.5 Lessons Learned

The main challenge in the development of HPN microservice was the existing difference in information systems, as well as different approaches to geodata and its management applied by the creators of these systems.

A CIDOC-CRM object orientated approach was applied to the development of HPN microservice, which considers HPN to be a place appellation referring to several places having in mind that its application could change over time. The place as an object is determined as GIS defined immovable geographic object: point, polygon or line (such as landscape, inhabited places, buildings, natural objects (mountains, river, etc, administrative areas, etc.).

A place name can be understood as an historical identifier for several and (or) as a kind of immovable heritage (based on CIDOC_CRM explanation on “material products of our minds”). In order to deal with existing disparities geodata from different databases were exported and saved within the HPN Thesaurus in separate datasets rather than combining them into one array.

Polygons were another challenge for the HPN microservice development. The final decision was made to represent a historical place as a point and not a polygon. The following aspects were considered important for this approach:

- The lack of polygons in geodata systems (e. g. only one database provided by LoCloud project partners had polygons).
- The complexity of defining historical polygons, especially in case of towns and cities, which requires more comprehensive historic research of sites and places, a sufficient amount of written sources and detailed maps.

6.6 Available documentation

All documentation about HPN microservice is available to developers at: (http://hpn.aksprendimai.lt/unitedgeo/) and also on the LoCloud project website at: (http://support.locloud.eu/LoCloud%20Historical%20Placenames%20Microservice).

The documentation includes:

HPN microservice description: http://www.locloud.eu/Media/Files/Deliverables/D3.5.-Historical-place-names-service
User documentation: http://support.locloud.eu/HPN%20user%20documentation or http://hpn.aksprendimai.lt/unitedgeo/documentation/
Technical documentation: http://support.locloud.eu/HPN%20Technical%20Documentation or http://hpn.aksprendimai.lt/unitedgeo/web-service/
Developers documentation: http://hpndocs.aksprendimai.lt/

6.7 Conclusion

The development of the HPN microservice was based on Open Source. The code is freely available for any interested parties and hosted on Github: https://github.com/justinasjaronis/hpn. It is published under General Public Lisence v3.0. The microservice can be re-used in the future collaboration and similar projects. Further development of the HPN infrastructure should consider:

- Enriching HPN Thesaurus content.
- Investigating different GIS data management models from contemporary point based model to polygon based model.
- Creating HPN "toolbox" (e.g. by historical maps vizualization, historical geo-information analysis tool, etc.).
- Enabling interoperability between LoCloud HPN microservices and another similar tools (e.g. Pleiades Plus).
- Applying HPN microservice and its funcionalities within Europeana.
- Introducing HPN microservice to international research communities (e. g. DARIAH).

7 Wikimedia and crowd sourcing

7.1 Development overview

The Wikimedia service has been developed using the PHP programming language and has been implemented as a harvester like application, which provides another metadata source to MORe. The concept behind the implementation is to be able to harvest content from any kind of Wikimedia installation based on the contributing user using Wikimedia's API. The API provides a number of metadata elements (these are not fixed – see Figure 10).


```
▼<response version="0.92">
  ▼<file>
    <name>Tora PM 78918 E.jpg</name>
    <title>File:Tora_PM_78918_E.jpg</title>
    ▼<urls>
      ▼<file>
        http://upload.wikimedia.org/wikipedia/commons/c/c0/Tora_PM_78918_E.jpg
      </file>
      ▼<description>
        http://commons.wikimedia.org/wiki/File:Tora_PM_78918_E.jpg
      </description>
    </urls>
    <size>287189</size>
    <width>600</width>
    <height>400</height>
    <uploader>PMRMaeyaert</uploader>
    <upload_date>2014-09-30T15:05:14Z</upload_date>
    <sha1>4d020dcc5ca41ad763f1f3c4a164c47ef5884041</sha1>
    ▼<date>
      <span style="white-space:nowrap"><time class="dtstart" datetime="2013-05-05 13:01:42">5 May 2013, 13:01:42</time></span>
    </date>
    ▼<author>
      <a href="http://commons.wikimedia.org/wiki/User:PMRMaeyaert" title="User:PMRMaeyaert">PMRMaeyaert</a>
    </author>
    <source><span class="int-own-work">Own work</span></source>
    <permission/>
  </file>
  ▼<description>
    ▼<language code="default">
      <div class="description mw-content-ltr en" dir="ltr" lang="en" style=""><span class="language en" title=""><b>English:</b></span> Spain, Torà, province, Lleida,
      Plaça de la Font, square of the fountain; www.pmrmaeyaert.eu;© Paul M.R. Maeyaert; polmayer@yahoo.es reference , PM_78918_E_Tora.jpg</div> <div class="description
      mw-content-ltr ca" dir="ltr" lang="ca" style=""><span class="language ca" title=""><b>Català:</b></span> Spain, Torà, provincia Lleida, Plaça de la Font; © Paul
      M.R. Maeyaert; polmayer@yahoo.es; reference , PM_78918_E_Tora.jpg</div>
    </language>
  </description>
  ▼<categories>
    <category>All media needing categories as of 2014</category>
    <category>Media needing categories as of 30 September 2014</category>
    <category>Uploaded with UploadWizard</category>
  </categories>
  ▼<licenses selfmade="1">
    ▼<license>
      <name>CC-BY-SA-4.0</name>
    </license>
  </licenses>
</response>
```


Figure 10: Metadata elements of API

The application maps these elements directly to EDM (Figure 11) and produces a SIP (submission information package) in line with the MORe specification. This SIP package can then be ingested, enriched and published according to the MORe workflow.

```
▼<rdf:RDF xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:edm="http://www.europeana.eu/schemas/edm/"
 xmlns:dcterms="http://purl.org/dc/terms/" xmlns:ore="http://www.openarchives.org/ore/terms/" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:wgs84_pos="http://www.w3.org/2003/01/geo/wgs84_pos#" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:skos="http://www.w3.org/2004/02/skos/core#">
 ▼<edm:ProvidedCHO rdf:about="http://more.locloud.eu/object/test/4d020dcc5ca41ad763f1f3c4a164c47ef5884041">
    <dc:title>File:Tora PM 78918 E.jpg</dc:title>
   ▼<dc:description xml:lang="en">
      English: Spain, Torà, province, Lleida, Plaça de la Font, square of the fountain; www.pmrmaeyaert.eu;© Paul M.R. Maeyaert; polmayer@yahoo.es reference ,
      PM_78918_E_Tora.jpg
    </dc:description>
   ▼<dc:description xml:lang="ca">
      Català: Spain, Torà, provincia Lleida, Plaça de la Font; © Paul M.R. Maeyaert; polmayer@yahoo.es; reference , PM_78918_E_Tora.jpg
    </dc:description>
    <dc:language>en</dc:language>
    <dc:language>ca</dc:language>
    <dc:subject>All media needing categories as of 2014</dc:subject>
    <dc:subject>Media needing categories as of 30 September 2014</dc:subject>
    <dc:subject>Uploaded with UploadWizard</dc:subject>
    <dc:type>IMAGE</dc:type>
    <edm:type>IMAGE</edm:type>
    <dc:creator>PMRMaeyaert</dc:creator>
    <dc:date>5 May 2013, 13:01:42</dc:date>
    <dc:publisher>PMRMaeyaert</dc:publisher>
    <dc:identifier>4d020dcc5ca41ad763f1f3c4a164c47ef5884041</dc:identifier>
    <dcterms:extent>287189</dcterms:extent>
    <dcterms:extent>600</dcterms:extent>
    <dcterms:extent>400</dcterms:extent>
    <dcterms:issued>2014-09-30T15:05:14Z</dcterms:issued>
  </edm:ProvidedCHO>
 ▼<edm:WebResource rdf:about="http://upload.wikimedia.org/wikipedia/commons/c/c0/Tora_PM_78918_E.jpg">
    <dc:format>JPG</dc:format>
    <dc:rights>CC-BY-SA-4.0</dc:rights>
  </edm:WebResource>
 ▼<ore:aggregation rdf:about="http://more.locloud.eu/object/test/4d020dcc5ca41ad763f1f3c4a164c47ef5884041#aggregation">
    <edm:aggregatedCHO rdf:resource="http://more.locloud.eu/object/test/4d020dcc5ca41ad763f1f3c4a164c47ef5884041"/>
   ▼<edm:object>
      http://upload.wikimedia.org/wikipedia/commons/c/c0/Tora_PM_78918_E.jpg
    </edm:object>
    <edm:rights rdf:resource="http://creativecommons.org/licenses/by-sa/4.0/"/>
    <edm:dataProvider>test</edm:dataProvider>
    <edm:provider>LoCloud</edm:provider>
    <edm:isShownBy rdf:resource="http://upload.wikimedia.org/wikipedia/commons/c/c0/Tora_PM_78918_E.jpg"/>
    <edm:isShownAt rdf:resource="http://commons.wikimedia.org/wiki/File:Tora_PM_78918_E.jpg"/>
  </ore:aggregation>
</rdf:RDF>
```


Figure 11: EDM format

7.2 Impact and benefits for the community

The impact of this application for the community is that it allows Wiki owners to deliver their content easily to Europeana. Such an approach presents users with a serious issue to do with the fact that Wikimedia does not enforce the existence of proper metadata according to a standard schema. Routing this content through MORe allows the validation and enrichment of this metadata using MORe's enrichment micro-services.

7.3 Usage of micro service within LoCloud

7.3.1 Public consumption

The usage of the Wikimedia application targets the Wikimedia owners that have content that could be valueable for Europeana. This includes content that falls into the cultural heritage domain and that contains at least some basic metadata such as: title, description, an image, etc.

7.4 Final status

- Version 0.1 of the Wikimedia application was originally developed as a standalone application that harvests content from Wikimedia using Wikimedia's own XML format and was based on the contributing user.
- Version 0.2 of the Wikimedia application included an internal mapping mechanism that would map Wikimedia's internal XML format to Dublin Core.

- Version 0.3 of the Wikimedia application was enhanced with the feature to harvest records from Wikimedia and produce SIP format packages. This would allow the application to be integrated into MORe as another metadata source. The advantage of this is that it provides seamless integration with MORe and it is easier for the users to understand and use.
- Version 0.4 of the Wikimedia application would map directly to EDM. This will allow users to skip one step (transformation from Dublin Core to EDM).

7.5 Lessons Learned

The lessons learned from developing the Wikimedia service include the overcoming of a certain number of challenges. These include:

- The extraction of all available metadata because Wikimedia API does not always provide all available metadata in a structured way but instead through microformats (as part of the main body of the page).
- The mapping of the available metadata to EDM. This requires a rule based algorithmic approach to disambiguate certain elements.
- The integration into the aggregation infrastructure so that users can use the service easily.

7.6 Available documentation

The Wikimedia related documentation is provided through MORe. Please see http://support.locloud.eu/Wikimedia for further information and the API guide.

7.7 Conclusion

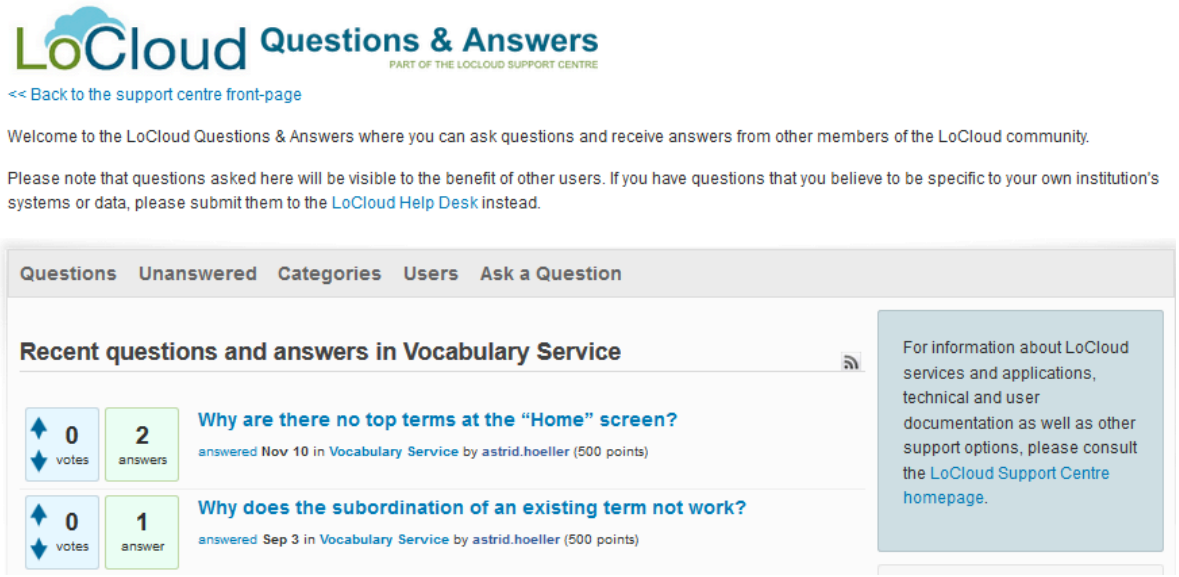
The Wikimedia application utilizes Wikimedia's API so that it can provide a more standardized way of communicating with Wikimedia. This makes the application more sustainable as it is easier to harvest from all Wikimedia installations and requires minimum changes to the system. The license of the application is based on Wikimedia Commons Licensing (http://commons.wikimedia.org/wiki/Commons:Licensing).

The service is integrated into MORe as another metadata source provider. MORe trained users can directly utilize it by giving the URL of their Wikimedia installation. This also implies that less documentation and training material is required.

8 Related links

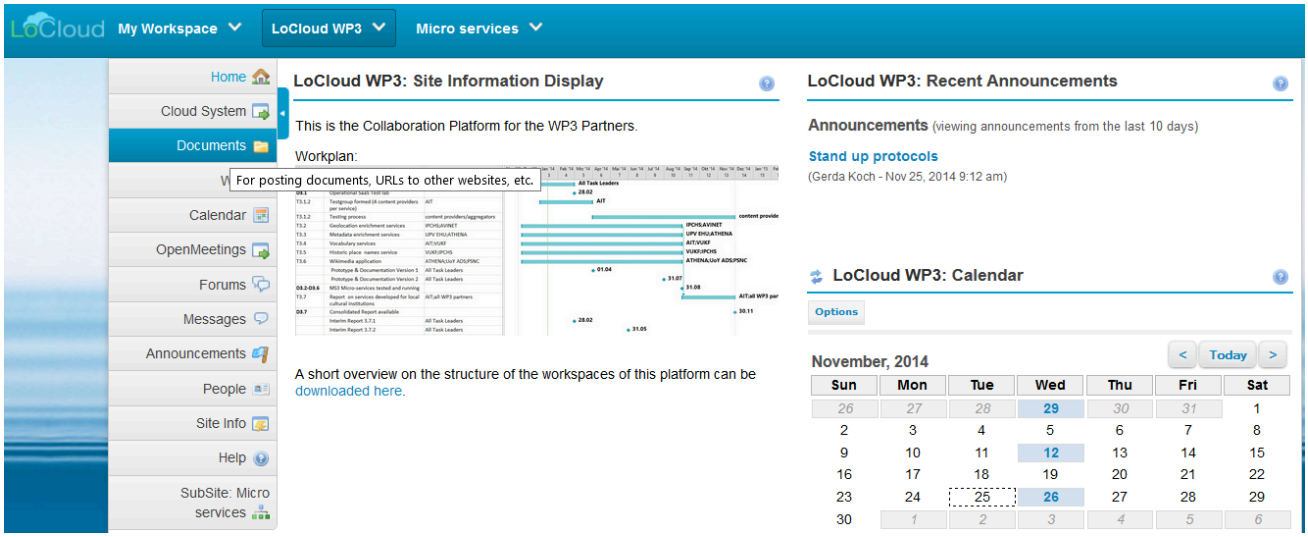
All documentation on the various microservices can be found in the LoCloud Support Portal:
Support Platform: http://support.locloud.eu/tiki-index.php

The FAQ Section is part of the LoCloud Support Portal and provides additional information on frequently asked questions about the various microservices.
FAQ Section: https://support.locloud.eu/qna/

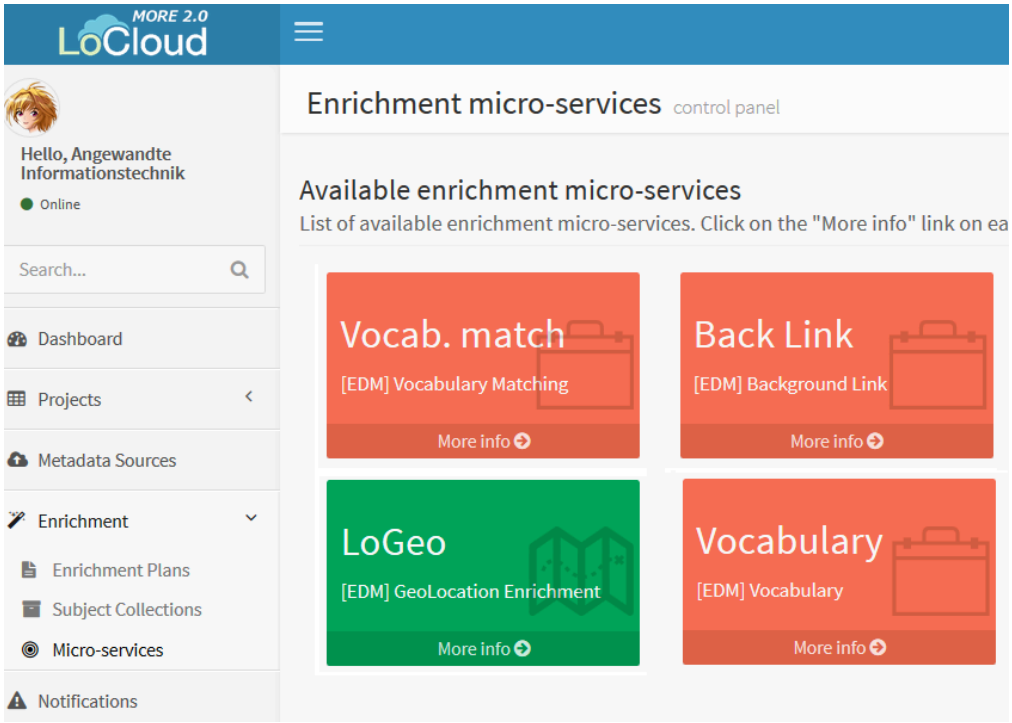


WP3 public deliverables, especially the deliverable documentation per microservice may be downloaded from the LoCloud project web site at:
http://www.locloud.eu/Resources/Deliverables

WP3 has used a dedicated documentation repository for work package collaboration on a SAKAI Platform that has been established in the LoCloud testlab and can be reached via: http://lc004.ait.co.at:8080/portal



The LoCloud aggregation environment can be accessed via http://more.locloud.eu/. It is accessible for content providers and aggregators.



9 References

Benda O., Höller A., Koch G., LoCloud Deliverable 3.4 Vocabulary services, http://www.locloud.eu/Resources/Deliverables, September 2014.

Bergheim S.R., Zakrajsek Z., Vodeb V., Stare J., Grilc A., LoCloud Deliverable 3.2 Geocoding Enrichment Services, Geolocation API (LoGeo API) and Geocoding application, http://www.locloud.eu/Resources/Deliverables, September 2014.

Gavrilis D., Dallas C., Makri D., LoCloud Deliverable 3.6 Wikimedia Application, http://www.locloud.eu/Resources/Deliverables, September 2014.

Koch W., Benda O., Koch G., LoCloud Deliverable 3.1 Operational SaaS Test lab, http://www.locloud.eu/Resources/Deliverables, March 2014.

Laužikas R., Jaronis J., Vosyliūtė I., LoCloud Deliverable 3.5 Historical place names service, http://www.locloud.eu/Resources/Deliverables, September 2014.

Soroa A., Otegi A., Agirre E., Agerri R., LoCloud Deliverable 3.3 Metadata Enrichment services, http://www.locloud.eu/Resources/Deliverables, September 2014.

10 Glossary

Term	Description
API	In computer programming, an **a**pplication **p**rogramming **i**nterface (API) is a set of routines, protocols, and tools for building software applications.[10]
CH (object)	**C**ultural **H**eritage (object)
CIDOC CRM	The **CIDOC Conceptual Reference Model (CRM)** provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.[11]
DARIAH	DARIAH, the Digital Research Infrastructure for the Arts and Humanities, aims to enhance and support digitally-enabled research and teaching across the humanities and arts.[12]
gazetteer	A gazetteer is a geographical dictionary or directory used in conjunction with a map or atlas.[13]
GIS	A **g**eographic **i**nformation **s**ystem (GIS) is a computer system designed to capture, store, manipulate, analyze, manage, and present all types of spatial or geographical data.[14]
LOD	**L**inked **O**pen **D**ata[15]
microservice	a particular way of designing software applications as suites of independently deployable services[16]
MLA	**M**odern **L**anguage **A**ssociation[17]
MORe	The LoCloud **M**etadata & **O**bject **Re**pository aggregation platform[18]
NED	**N**amed **E**ntity **D**isambiguation[19]

[10] http://en.wikipedia.org/wiki/Application_programming_interface 20th November, 2014
[11] http://www.cidoc-crm.org/ 20th November, 2014
[12] https://www.dariah.eu/ 20th November, 2014
[13] http://en.wikipedia.org/wiki/Gazetteer 20th November, 2014
[14] http://en.wikipedia.org/wiki/Geographic_information_system 20th November, 2014
[15] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData 20th November, 2014
[16] http://martinfowler.com/articles/microservices.html 20th November, 2014
[17] https://www.mla.org/ 20th November, 2014
[18] http://more.locloud.eu/ 26th November, 2014

Term	Description
polygon	A plane figure that is bounded by a finite chain of straightline segments closing in a loop to form a closed chain or circuit[20]
REST	**R**epresentational **s**tate **t**ransfer[21]
SKOS	SKOS (**S**imple **K**nowledge **O**rganization **S**ystem) is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists and taxonomies within the framework of the Semantic Web[22]
UI	**U**ser **I**nterface
VM	**V**irtual **M**achine[23]
WMS	A **W**eb **M**ap **S**ervice (WMS) is a standard protocol for serving georeferenced map images over the Internet that are generated by a map server using data from a **GIS** database.[24]

[19] http://en.wikipedia.org/wiki/Entity_linking 20th November, 2014
[20] http://en.wikipedia.org/wiki/Polygon 20th November, 2014
[21] http://en.wikipedia.org/wiki/Representational_state_transfer 20th November, 2014
[22] http://www.w3.org/2004/02/skos/ 20th November, 2014
[23] http://en.wikipedia.org/wiki/Virtual_machine 20th November, 2014
[24] http://en.wikipedia.org/wiki/Web_Map_Service 20th November, 2014