DELIVERABLE

Project Acronym:	**Europeana Cloud**
Grant Agreement number:	**325091**
Project Title:	**Europeana Cloud: Unlocking Europe's Research via The Cloud**

Deliverable 1.5 (1 of 4): Expert Forum Case Studies Report

Revision: 2

Authors:

Prof. Susan Schreibman (Trinity College Dublin)
Vicky Garnett (Trinity College Dublin)

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

Revision History

Revision	Date	Author	Organisation	Description
1	15/07/2013	Susan Schreibman and Vicky Garnett	Trinity College Dublin	First draft of report sent to RCAB
2	25/07/2013	Susan Schreibman and Vicky Garnett	Trinity College Dublin	Draft revised following comments from RCAB

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

D1.5 (1 of 4) Expert Forum Case Studies Report

Table of Contents

1. *The Purpose of the Expert Forum*

The Expert Forums take place as part of Work Package 1 of Europeana Cloud. There are four Expert Forums in total, each addressing a particular area of research.

Expert Forum 1 took place in Month 5 of the project (June 2013). The forum aimed to investigate the uses for Europeana by researchers in both the Humanities and the Social Sciences by developing case studies that reflect the typical needs of researchers at whom Europeana Cloud will be aimed. The case studies developed through the outcomes of this Forum examined both Europeana in its current form, and at Europeana as it could be as it moves into developing cloud services. The outcomes of this Expert Forum feed into Deliverable 3.1 in Work Package 3.

Expert Forums 2 and 3 will take place in Months 9 and 10 of the Project (October and November 2013) and will look specifically at tools that could be developed within Europeana Cloud for the Humanities (Expert Forum 2) and the Social Sciences (Expert Forum 3).

Expert Forum 4 is scheduled to take place by Month 30 (July 2015) of the project, and will provide a broad review of the tools and content access and use services provided by Europeana Cloud, and develop recommendations for future work, including how the engagement of researchers will continue beyond the lifetime of the project to ensure their future use and uptake of the Europeana Research platform.

This report focuses on the outcomes of Expert Forum 1 “Case Studies”.

2. *Identifying Disciplinary Areas to Develop Case Studies*

In identifying research areas targeted in this Expert Forum, it was decided not to invite experts from many different disciplines as it was felt researchers may have too little in common to develop useful case studies. Therefore four key research areas were selected that would between them use a variety of data types.

The academic areas selected were Humanities disciplines History (focusing on context-based scholarship) and Archaeology (focusing on images, 3-D, and images), with broader areas of Social Sciences (focusing on datasets) and Ontologies (focusing on the metadata itself). All the focus groups, in addition to specific datatypes, dealt with metadata in Europeana.

3. *Identifying the Experts*

The invitees to the forum were identified to ensure the broadest range of expertise and experience. They were also selected to create a European-wide perspective. Participants came from institutions in Denmark, Greece, Ireland, Lithuania, The Netherlands, Norway, Portugal, Sweden, and the UK.

The group comprised of experts from eCloud-related institutions and those from institutions who were not directly involved with the Europeana Cloud project. In addition a member of the eCloud Research Communities Advisory Board, Professor Rob Kitchin, was invited to participate.

4. Designing the Expert Forum

The intention of the day was to establish an overview of the research that can be conducted using Europeana currently, and develop this to include future possibilities that might be realised as Europeana moves into a Cloud-based service.

The results of Expert Forum 1 will feed directly into the work being conducted in Work Package 3. In particular, they will be analysed for use in Deliverable 3.1:

> *Deliverable 3.1*
> *In close alignment with the tasks in WP1, KUL will lead on developing personas (descriptions of typical researchers that we address with this project), scenarios and use cases that describe in details what kind of tool a researcher would typically use, and how that use would fit in his/her typical workflow. This will result in Deliverable 3.1 (M6).*

The aim of this Expert Forum, therefore, was to task the experts from within their groups to develop scenarios for specific researchers at key stages in their careers and within certain disciplines, focusing on the resources and tools they use, and the methodologies they employ.

In order to meet this objective, the day was broken up into three main sessions to create a structure in which a healthy discussion could be achieved that would result in useful scenarios that could be developed into case studies.

Session 1 - Europeana Treasure Hunt

The first session served as both an ice-breaker and gamified the task of introducing researchers to Europeana, both in terms of Europeana's content, functionality, and special exhibitions. As with the eCloud kick-off meeting, participants were divided into groups with cloud icons. These icons divided participants into the domain groups for Archaeology (Stratus), History (Cumulus), Ontologies (Stratocumulus), and Social Sciences (Altocumulus). But for this ice-breaker, sub-teams formed which had one cloud per group so that participants had the opportunity to meet participants outside their subject area.

The task itself was entitled 'Europeana Treasure Hunt'. Teams were asked to complete three Europeana-based challenges within 15 minutes. All teams were asked to create a Europeana Profile. This allowed those who had not used Europeana before to familiarise themselves with the user profile function for collecting and saving searches and results.

The three tasks were developed to highlight search functionality and metadata present in the current instantiation of Europeana. The exercise can be found in Appendix III – Europeana Treasure Trail.

Multiple Search Terms for the Same Named Entity

The purpose of this task was to highlight Europeana search functionality and associated issues when metadata for key word terms is not controlled, such as in the case of the Great War. This is an

issue not simply within a single language, and is only compounded when searching across multiple languages. One group found 27 cognate terms for this event, each resulting in a different data set being returned

Minimum Metadata

Many of the items within Europeana do not contain the full complement of metadata. In some cases, only the name of the item and the source is provided. The purpose of this task was to enable Expert Forum members to familiarise themselves with the kinds of metadata typically provided by contributors, and to get a feel for the shortage or completeness of metadata typically provided.

Europeana-Whacks

The 'Europeana-Whacks' task was based loosely on the practice of trying to find a 'Google-whack' whereby a search term produces only one result in Google. The purpose behind the 'Europeana-Whacks' task was to allow the Expert Forum members the opportunity to see if they could find something very specific within the database. Often a researcher doesn't want to manually filter through the results to find the most relevant item. This task was to establish if that can indeed be done, and what difficulties the task may present.

Session 2 - What Research Can Europeana Support in its CURRENT form?

In order to identify areas for development to make Europeana more useful for Researchers in future, it was necessary for Expert Forum members to understand the current scope of Europeana. Each group was asked to create a case study in which a fictional Researcher might engage with Europeana as it is currently implemented. To do this each group was invited to complete a Case Studies template (found in **Error! Reference source not found.**).

The information the groups were asked to complete included:

- The research task/ goal of the hypothetical researcher
- Their career/ experience level (both disciplinary and within Digital Humanities)
- Discipline Research Area
- Resource/Data Type
- Tools to be used
- Methodologies
- Problems they might encounter

Participants were instructed in this session to *only* develop case studies that utilised the current functionally of Europeana resources; i.e. metadata.

Session 3 – Future Possibilities of Europeana as a Research Platform

The last session of the day further developed the case studies developed during the second session. Only now, participants were instructed to imagine that there were no restrictions as to what Europeana might provide in terms of content types. Moreover, they were instructed to envisage the types of services that could be useful in the analysis of these data types especially within a cloud platform.

5. Expert Forum Feedback

After each breakout session there was ample opportunity for the group to reconvene as a whole and discuss the observations and outcomes generated by the disciplinary-based groups. These conversations have been consolidated under the following broad subject headings.

Insufficient Metadata

The overall response to Europeana in its current form is that the majority of items do not have sufficient metadata to make it a truly usable tool for research. The users in the Expert Forum commented that in the majority of cases, searches for a targeted term resulted in a vast number of unrelated hits or, when the hits were relevant, the lack of metadata of individual items made the result set almost impossible to work with. Moreover, images were frequently missing and the date provided would often be inaccurate (or at least not relevant), while lexical ambiguity would bring up 'wildcard' results.

The group as a whole found the lack or partial lack of metadata frustrating. The overwhelming consensus was that this was one of the major factors that prevented the usefulness of Europeana as anything more than a discovery tool.

User Ranking

One method for addressing this issue was an incentive scheme, whereby a 'star-ranking system' could be used to rank the metadata provided. Users of Europeana could rate an item for the quality of its metadata from 1 star (poor quality) up to 5 stars (excellent quality). The ranking system would take into consideration the completeness of the data, the quality of the data (including provenance and providing institution), as well as a description of the item, having at least metadata for these fields comply with the best practices in digital curation.

While the individual items could be ranked, the contributing organisations could also then be ranked according to the quality of their metadata. This would be reviewed annually to encourage contributors to improve where needed.

Multi-lingual Resources

At present there are no translational services in Europeana. As metadata can be submitted in the native language of the contributing organisation, a user must search in every language to be sure she has done a thorough search for a particular term. For example, if a researcher types in 'World War One', then she will receive results with the specific phrase 'World War One' in the metadata, but won't get results with 'La Première Guerre Mondiale' (French) or 'Pierwsza Woyna Swiatowa' (Polish).

A key recommendation of the Expert Forum is that multilingual metadata should be dynamically generated by the system as it is impractical to ask content providers to provide multi-lingual search terms. It was strongly recommended that a thesaurus or translation tool be built into Europeana that can automatically link multilingual metadata to enable researchers to find all items associated with a search term.

Date Ranges not Recognised

Many users, particularly in History and Archaeology, found when entering date ranges into the search field that only the specific years they entered were being returned. For example, a query on items from the '1800s' returned only items from the year 1800, or items in which the number '1800' was specifically mentioned as a quantity. Similarly, when entering a date range such as '1850 – 1950', only those two years would be specified in the results.

Spatial and Temporal Mapping of Results

Participants unanimously recommended that Europeana needed better browsing functionality. Suggestions included spatial mapping and temporal mapping via maps and timelines, respectively. Simple conceptual maps, such as tag clouds or other visualisations would also be desirable.

More Transparent Citation Methods

Forum participants agreed that a more transparent and reliable citation method was needed to both allow users to cite the results of their searches so that they could be replicated or validated by other Researchers, and to cite specific items found within Europeana.

Tool Development

There was overwhelming consensus that Researchers from all disciplines would want to export result sets into specialised, often domain-specific tools. Therefore, it was strongly recommended that Europeana not develop specialised tools, but rather focus on

- generalised tools (as mentioned above) to aid in discovery
- tools that allow for (meta)data export into a variety of formats
- tools that provide more sophisticated and targeted filtering of results than are currently available.

The group felt strongly that Europeana should not concern itself with developing discipline-specific tools. While these would be expensive to develop and would serve only a small subset of users, it would also require that Researchers learn yet another piece of software. Moreover, it is also likely that in the time it would take to develop these tools, discipline-specific tools will have moved on, placing Europeana in a game of constant catch up.

Europeana as a Teaching Resource

A strong recommendation was made for Europeana to strengthen its resources for teachers, allowing for lesson plans and interactive functions that could be used within the classroom, or be imported into an eLearning environment such as BlackBoard or Moodle.

6. *Case Studies*

The following case studies present hypothetical researchers working in typical research scenarios across the EU as developed during the breakout sessions.

Case Study 1: Early Stage Historian using Europeana in its Current Form (History)

Mary is an early stage History researcher at MPhil level, who is reviewing the background information for her thesis proposal. She intends to discuss how the physical form of printed bibles changed during the reign of Queen Victoria. Mary typically uses search engines in browsers or in specialised databases, such as her library subscribes to, but she is not familiar with the technology behind these search engines. She has no experience of XML or programming. She is comfortable with drag-and-drop facilities, and tools such as 'Evernote'.

Much of her data is in the form of images, for example of the Bible, bindings and the metatdata that would support such images. She also makes use of text taken via OCR software from Bibles in the Victoria Era, text of Bible commentaries, reading plans and other materials, Sunday school teaching materials such as books and pictures, Church inventories and information from bookseller' catalogues and advertisements.

In conducting her research, Mary plans to use referencing tools, both to save the results of searches, but also to populate her final written thesis. These include open source applications such as Evernote, Zotero and Endnote. She also plans to use geo-referencing tools in her research, such as CreateMap, and also wishes to publish any images she may find. She therefore needs to ensure that her sources for the images allow for reproduction in her thesis without breaching copyright laws.

The methodology for her research will typically involve communication with researchers in similar fields, data capture and analysis, publishing and dissemination, and data structuring through means such as cataloguing and indexing. She will employ methods that will allow her to use existing data for collation, image-segmentation and text recognition.

She therefore needs Europeana to integrate well with open-source tools such as EverNote and EndNote for citation purposes, as well as to allow use of image files in a drag and drop function. The records also need to be trusted. The problems she is encountering, however, are mainly due to a lack of insufficient records of analogue items that would typically fall into the scope of her research. While searching for items in Europeana that were produced during the Victorian era, she is also experiencing difficulty in finding items within a range of dates. Her searches for items produced during the '1800s', or from '1837-1901' produce limited results, often restricted to the specific years she has typed, and not to the wider range, as she hopes. The provenance information in the metadata of many of the items she is finding is not complete, and therefore does not provide her with sufficient background to either reliably use the items as data sources for analysis, or to use as evidence to support her arguments.

Case Study 2 – Experienced Computer Scientist using Europeana in its current form (Ontologies)

Jonathan is an experienced computer scientist working in development. He wants to be able to write SPARQL queries to achieve two key goals: 1) for discovery/info seeking, and 2) for visualisation of information e.g. geographical based information. He wants to then create a list of dataset collections, which have more than 10 numbers of items with certain rights qualification. The project aims to visualise the datasets of Europeana in order to inform Europeana content strategy but also to assure researchers as to the quality and depth and coverage of the research they are performing. This should result in tools that visualise coverage, number of records e.g. related to a particular area.

Jonathan will be working mainly with RDF schema, and wants to use SPARQL for the analysis, and CARARE for the visualisation of his results.

Currently, the metadata on Europeana in its current form is almost useless for this method of analysis. Jonathan finds that his research is precisely the research needed to review the current state of the Europeana dataset, however the dataset itself is the biggest restriction to his research. He finds that he can't cite the state of the data, for example, he doesn't know what has been excluded from the metadata, as it does not comply with VoID, the Vocabulary of Interlinked Datasets, which sets a standard for metadata use. This would be especially useful in Jonathan's research using RDF.

Case Study 3 – An Experienced Lecturer in Sociology using an enriched Europeana (Social Sciences)

Nicola is a Senior Lecturer in Sociology at a University in 2017. She uses Europeana to develop lesson plans for her students on her BA Hons in Sociology course. She is teaching a module on migration, both immigration and emigration, asking her students to particularly look at the reasons people move, where they move from and to, and who they are. Parts of the course are delivered online through eLearning.

Much of the data she uses is taken from trusted Europeana contributors providing migration data on both Immigration and Emigration stories from the past 50 years. Europeana has already taken care of any data protection issues in regards to privacy choice and consent. The majority of this data is in text form, which she extracts for analysis by her students during lectures. Before she can do this, however, she likes to review the content of the data she finds on Europeana using an open source tool that is compatible with the Europeana infrastructure.

She also searches for items such as photographs, letters, diaries, travel manifests, travel documents, and other items that might be associated with migration patterns.

The tools she regularly makes use of are face-recognition searches in photos and video, such as 'imagevision', mapping of results, a timeline of results, the full-text search within the content of the items and the export function, particularly for use in spreadsheets (e.g. MS Excel) and another open source software package of her choosing.

Nicola knows that the data she is extracting is trusted, as it complies with the best practice for the storing and maintenance of personal information across the EU. She knows that Europeana regularly runs queries with the contributors to check that they are compliant, and she is also able to run this check herself at intervals.

She is able to select the most useful items on Europeana, thanks to the star-ranking system that has been introduced. She makes sure to rank any items she uses herself, in order to assist her fellow researchers in selected high-quality items. She also knows which contributors across the EU have the highest rankings for most consistent metadata, and often this reduces the amount of time she spends searching for items, as she can select which contributors' content to search. She is also able to filter the results of her searches down to gather the most relevant items.

Case Study 4 – An early-stage career Archaeologist using an enriched Europeana (Archaeology)

Franz is a German postdoctoral fellow working on a project that will digitally recreate the Parthenon using 3D data, images and texts.

He recently completed his Doctoral Degree and has a reasonable knowledge of digital tools, but it still learning new technologies. He therefore relies on the basic tools available in Europeana that allow him to export the data he needs into the digital tools with which he is familiar. He enjoys the ability to preview 3D media objects before he analyses them for content, and is able to find the relevant metadata he needs by filtering the results and using the mapping and timeline visualization functions in his search.

As well as his native German, Franz speaks English, Greek, and some Ancient Greek and Latin. However, he is not familiar with other languages in Europe, and therefore is not able to directly translate search terms. Thankfully he is able to type a search term into Europeana in one of the languages he does speak, and is able to find useful metadata from any country, as Europeana has an automatic translation function that searches for multilingual terms.

7. Conclusions

Recommendations for Europeana

It is clear from this Expert Forum that participants do not consider Europeana a truly useful or robust resource without better quality metadata and a suite of user tools to aid browsing and analysis. Moreover, it is essential that this metadata can be viewed, filtered, and manipulated in a variety of ways.

Key recommendations include:

- encouraging Europeana contributors to provide complete metadata in future
- the development of metadata enrichment tools for current incomplete metadata
- providing simple visual tools to enhance browsing (eg maps, timelines, tag clouds)
- providing export mechanisms in a variety of metadata formats for Researcher-generated metadata sets

Recommendations for Future Expert Forums

The three-session format of the day proved successful in terms of gathering feedback, encouraging discussion, and gaining wide community participation. Despite the very different disciplines and levels of expertise, the feedback from the disciplinary groups was generally in agreement.

The length of the Forum worked well, as it was long enough to promote debate, but succinct enough to keep the participants focused on the task ahead of them. This particular forum took place over one full day, and due to the focused delivery was able to keep to the allotted time and allow participants to travel home afterwards. However it is also proposed that for future Expert Forums, two half-day sessions (beginning with lunch on Day I) could be more desirable. Consideration will have to be given to the division of sessions and tasks across the two days in order to maintain momentum of debate and keep the participants focused. However, the evening break could also serve to refresh the participants for the session the following morning.

Gathering participants was by and large the most challenging aspect of the Expert Forum delivery. The organizers therefore strongly recommend that identification and invitation of participants in future Expert Forums should be conducted at least four months in advance. It is also recommended that complete invitation lists with the responding RSVPs should be kept for reference when coordinating subsequent Expert Forums.

Appendix I – List of Participants

Name	eCloud status	Inst	Research Field
Christine Morris	Non-eCloud	Trinity College Dublin	Archaeology
Erik Champion	Non-eCloud	Digital Humanities Lab Denmark	Archaeology
Ingrida Vosyliute	Non-eCloud	Vilnius University Faculty of Communication (VUFC)	Archaeology
Anthony Corns	Non-eCloud	Discovery Programme	Archaeology
Norman Rodger	eCloud WP1	University of Edinburgh	Archaeology
Mary Rowlatt	eCloud WP1	MDR Partners	Archaeology
Matt Munson	Non-eCloud	Gottingen Centre for Digital Humanities	History
Mark Sweetnam	Non-eCloud	Trinity College Dublin	History
Wilko Hardenberg	Non-eCloud	Rachel Carson Center, LMU	History
Thomas Baldwin	eCloud WP1	The European Library	History
Kees Waterman	eCloud WP1	NIOD	History
Rob Kitchin	eCloud RCAB member	NUI Maynooth	Social Science
Thoa Pam	Non-eCloud	Dublin Institute of Technology	Social Science
Jorge Ricardo da Costa Ferreira	Non-eCloud	Universidade Nova de Lisboa	Social Science
Stefan Ekman	eCloud WP1	University of Gothenberg	Social Science
Louise Edwards	eCloud WP1	The European Library	Social Science
Christian-Emil Ore	Non-eCloud	University of Oslo	Ontologies
Costis Dallas	eCloud WP1	University of Toronto	Ontologies
Øyvind Eide	Non-eCloud	University of Oslo	Ontologies
Alexander O'Connor	Non-eCloud	Trinity College Dublin	Ontologies
Torsten Reimer	Non-eCloud	JISC	Ontologies
Susan Reilly	eCloud WP1	LIBER	Ontologies
Owain Roberts	eCloud WP1	National Library of Wales	Ontologies
Agiati Benardou	eCloud WP1	DCU Athens	eCloud staff
Susan Schreibman	eCloud WP1	Trinity College Dublin	eCloud staff
Vicky Garnett	eCloud WP1	Trinity College Dublin	eCloud staff

Appendix II – Expert Forum Running Order

Expert Forum - Case Studies

18th June 2013
IIIS Seminar Room (6.002, Arts Block)
Trinity College Dublin

9.25	House-keeping	Vicky Garnett, TCD
9.30	Session 1 About eCloud and the Aims of the Day	Hosted by Prof. Susan Schreibman, TCD Dr. Agiati Benardou, DCU Athens
9.50	Europeana 'Treasure Hunt' game	Vicky Garnett, TCD
10.20	Session 2 About Europeana	Prof. Costis Dallas, University of Toronto
10.40	**Coffee Break**	
11.00	Session 4 - Case studies - Europeana NOW - 1hr 30mins	Introduced by Vicky Garnett
12.30	Session 4 Reporting back on Scenarios	Facilitated by Vicky Garnett
13.00	**LUNCH**	
13.45	Session 5 Potential uses of Europeana - Discussion	Introduced by Susan Schreibman/ Agiati Benardou
14.15	Session 6 Revisiting case studies and presenting with how Europeana could/should be used	Introduced by Susan Schreibman
15.45	Reporting back	Facilitated by Susan Schreibman / Agiati Benardou
16.30	Day ends	

Appendix III – Europeana Treasure Trail

The Europeana Treasure Hunt!

Before you start:

- Log in to www.europeana.eu and create a 'My Europeana' profile for your team. To do this, you will need to create an account using an email address.
- If you wish, you could create this for one individual, as that person may wish to use Europeana at a later stage.
- If a team member already has a Europeana profile, feel free to use that, but don't use any previously saved searches for the Treasure Hunt!
- **Use the 'My Europeana' function to save all your searches.**
- **You can only use one computer.**

You have 15 minutes in total to answer the following questions. The team with the most points at the end of the hunt wins.

1. **Multiple Search Terms**
   How many different search terms does Europeana use for the major European war that began as a result of the assassination of Archduke Franz Ferdinand of the Austro-Hungarian Empire?
   1 point for each search term found.

2. **Minimum Metadata**
   Using at least 3 different search terms, how many examples of items can you find that have a minimum of 10 metadata fields completed.
   1 point for each example

3. **Europeana-Whacks.**
   Using between 2-5 words, what is the fewest number of results you can find for a search term.

   You score **1 point for each item produced** in a single search (for example, if a search term gives you 10 results, you score 10 points). The aim of this task is to score as few points as possible, as points scored in this round will be deducted from your overall score. Select your three lowest scoring search terms (and give the number of items found for each). Scores of '0' will not be accepted, and will instead incur a 5-point deduction from your overall score.

   Rules: Do not use proper nouns, including place names or people's names. Search terms MUST be between 2-5 words long.