Project Acronym: E Awareness
Grant Agreement number: 297261

Project Title: Europeana Awareness

Subtask D3.3 – Analysis of Europeana Content for local history and genealogy users/re-users

Date of submission	20 June 2013
Author(s)	Marie-Claire Dangerfield; Hans de Haan
Dissemination Level	PU

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1	13 June 2013	Hans de Haan	EF	Editing ch. 5-7
2	19-20 June2013	Hans de Haan	EF	Editing
3	22 June 2013	Rob Davies	MDR	Review
4.	29 June 2013	Hans de Haan	EF	Integrating reviewer's remarks in report

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

1. Introduction

Workpackage 3 of *Europeana Awareness, Developing New Partnerships*, addresses several key stakeholder sectors which are not yet optimally exploited by Europeana and is taking action to raise awareness among them of the potential benefits of developing such a relationship. The four key sectors identified are:

• Public libraries
• Local archival collections
• Broadcasters
• Open Culture re-users

This approach to them, in each case, entails - with modifications adapted to the various starting points - :
> • a relatively brief phase of inputs from and discussion with the participating stakeholders; based on the state-of-the-art examples and previous work followed by more structured assessment and negotiation of the possibilities;
> • piloting and testing of joint activities with the sector or cultural domain, utilizing available tools and processes;
> • a period of dissemination and mainstreaming with the sector, conducted via appropriate cooperation with the public media campaign in WP1;
> • accompanied by identification of processes for operationalisation of promising services, through the appropriate project or other instrument within the Europeana 'ecosystem'.

There are numerous potential target audiences for the rich data held by local archives including, researchers, local historians, genealogists/those conducting family research and tourists. WP3 will seek to work specifically with some private online genealogy companies to place Europeana related information on their sites worldwide. Europeana Awareness essentially begins a process, which is potentially of very large proportions, of making the best possible use of this rich and highly important record of Europe's cultural heritage within Europeana.

Deliverable D.3.3 will partially draw on the results of *Deliverable 3.2, Assessment of the role of local archival collections* (3 January 2013) and present the results of a further analysis of the Europeana content and technical possibilities to assess whether in what form Europeana might develop services of relevance and interest to Europe's genealogy and local history markets either through its own portal or in partnership with commercial services. The more extensive research on the commercial possibilities and commercial partnership will take place in Work Package 3, task 3.2.5, MS20.

2. The questions and the method

The main question to be answered in this report, *Deliverable 3.3 Analysis of Europeana Content for local history and genealogy users/re-users,* is in what way

Europeana's content can support the finding of genealogical data and possibly enrich these by adding extra information or pictures.

Therefore the main purpose of the enquiry will be to investigate Europeana's technical ability to produce the necessary information for the (amateur) genealogist and get an insight in the possible content available in Europeana and the extra content needed to specifically cater for the needs of genealogy.

- We have discussed this with a representative of the Dutch Bureau of Genealogy and have used this information to address the problem.

- In order to get a realistic view on the Europeana material we have staged a search and described the results.

- We have looked into the possible technical changes that had to be made in order to be able to cater for the genealogist's needs.

- In the *Annex* we have described the rights labeling issues separately in the light of the Europeana Data Agreement.

3. Digitised genealogy: loving the search game

Genealogy has been one of the forerunner services in archival digitisation because of its strong attraction to the general public. People simply love to find out about their ancestry. To most amateur genealogists the search is what they go for; they are looking for their ancestors as if it is a game. The weight of genealogy and local history is also shown in a fairly recent, 2011, Polish survey report on digital libraries. The survey indicates that a relatively stable group of users of digital libraries in Poland consists of persons interested in local history and amateur genealogists. This is a dominant user group, accounting for 60 per cent of all users, divided more or less evenly between local history and genealogy.[1] These recent Polish results are very much in line with the outcome of a 2002 Canadian survey that showed 33 per cent of its users were interested in history and 29 per cent in genealogy.[2] These numbers all underline the importance of genealogy and local history to the general public. The success of the Europeana 1914-1918 road shows on WWI memorabilia and family war experiences fall definitely into this category too.
This is following the dictum that all history in the end is local history and all interest is personal or family interest.

Finding out one's personal past and trying to find out who was who, where and when, has always appealed to people. People are literally trying to find their place in history and are willing to spend time and money in doing so. As a result genealogy was one of the forerunners in text digitisation and public dissemination, creating new and

---

[1] Gorny,. M., Mazurek, J, (2011), "Key users of Polish digital libraries", *Electronic Library,* Vol. 30, pp.543-556, available at: http://dx.doi.org/10.1108/02640471211252238
[2] Cherry, J.M. and Duff, W.M. (2002), "Studying digital library users over time: a follow-up survey of Early Canadiana Online", Information Research, Vol. 7 No. 2, (2002); available at: http://inforrnationr.net/ir/7-2/paper123.html

easier access to genealogical data. This activity seems to be growing in the western world, especially in the 50+ age category, a user group that combines both time and money to spend on genealogy.

At the same time genealogy and ancestry can be big business. The selling of Ancestry Com., the world's largest on line genealogy/ancestry company, in 2012 for € 1,2 billion to an investment company is an example.[3] Ancestry is planning to increase world coverage, and has entered the European genealogy market too.[4] It offers already a number of localised websites, such as Ancestry.de for Germany, Ancestry.co.uk for the United Kingdom, and the same for France, Sweden, Italy. Mundia (by Ancestry) as a brand covers other countries in the world that do not have a localised name, e.g. the Netherlands.[5] At the same time Ancestry offers its users a wide range of genealogical products to facilitate search and documentation. It is, without doubt, a very well organised genealogical business with a great business case.

In order to give an indication of the money involved the following example: the most extensive and expensive Ancestry monthly subscription in the USA costs 35 USD a month. Ancestry has 2,000,000 million paying subscribers (though not all have the most extensive subscription), so there is a continuous monthly cash flow running into the millions.

Making money with on-line genealogy can be divided into three basic categories:

1. **Providing access to source information** – making money by opening up the sources (birth registers, church registers etc.). This, however, presupposes having digitised content;
2. **Providing internet services** – creating and exploiting services like enriching data and providing pictures, creating educational services for schools, connecting genealogy to tourism and heritage etc.;
3. **Creating specialised programs** – programs to support and facilitate genealogical searching.

These categories also hold for non-commercial usage. When public money has been used for digitisation of public sources, it is clear that commercial usage will be limited to specific areas defined by policy. Providing services and specialised programs can be left to the market, although simple and straight forward searching tools used by libraries and archives usually are part of the public service.

As always the legal rights (IPR) of all material will have to be clearly stated. Genealogical information can be public and free of charge. It can also be public information that has to be paid for. The same goes for private information and private property rights. Generally all rights have to be cleared in advance.

---

[3] http://www.forbes.com/sites/ericsavitz/2012/10/22/ancestry-com-to-sell-to-permira-for-32-a-share-in-cash/
[4] http://www.ancestry.com/cs/us/affiliate
[5] http://www.mundia.com/nl/

4. Genealogical information

What are people looking for when searching for genealogical material?
The search usually starts at a very basic level with: names and places, dates, occupations. A combination of name and place, or two names is often the best way to start. Information will usually be found in metadata, scanned sources and publications. Additional information can be found in property documents, military files, occurrences in judicial files/court records, notarial records, church registers and land records (Cadastre).

Contextual information, which is usually more of a historical nature, is considered an extra, but is not necessary for the pure genealogist. It will, however, be an interesting extra feature in painting the family picture more fully. This type of information actually borders on local history and its easy digital availability to genealogists may open up digital local history more efficiently in combination with genealogy.

5. Europeana and genealogy : methodology

Genealogy and local history are growing fields of interest in the cultural heritage sector and make up the majority of traffic to archival institutions and local history centres. It is therefore important to see what types of records Europeana holds in relation to this user group, how many records are accessible through the Europeana portal of this type and to look at how it would be possible to amass more records of importance to this user group and what impact this would have on Europeana, its structure and its holdings. Genealogy searches will usually have a plug-in on their web-service linking types of records or suggesting what records may be useful to look at next or that may link to the content from record to record, e.g. 'Person X in registry A, may also be found in registry B, however if the name is common it may not be of any use to the researcher'.

Currently the Europeana portal does not have the facility for the end user to conduct two parallel searches without some technical knowledge. Premium services, like Ancestry, also provide interactive services, and social media. Europeana does not.

Genealogy and local history is a highly subjective field, its goals known to the researcher alone. There is no way of tracking the number of names, locations or events held in Europeana on a micro level. In order to get a better practical look at how genealogy works in Europeana, we have created a number of searches using various genealogical concepts.

We have researched the type of records related to genealogy and local history in the Europeana portal using a macro approach. This means that the records were looked at first using key word searches related to genre type, such as census, birth registries etc. in multiple languages (Dutch, English, French, German, Italian and Spanish) and the number of records retrieved recorded under these key word searches. In addition to this numbers of these records were broken down into content providers (e.g. National Aggregators or Projects), then to data provider level, the country of origin and number of records by media type. This method of content analysis allowed for the broadest number of genres to be

analysed across Europeana's 26.7 million records. An individual micro content analysis via individual records across databases would not yield definitive results due to the fluid nature of genealogical material and subjectivity of searches. The table below provides an illustrative example of the results of a keyword search using the phrase 'census'.

Keyword	Language	Number of Records	Provider	Data Provider	Country	Media Type
Census	English	908,002	Irish Manuscripts Commissi (906686)	Picture the Past (420)	Ireland (906692)	Text: 907.307
			CultureGrid (597)	Boston Public Library (182)	UK (655)	Image: 684
			BHL Europe (214)	Leodis (109)	Europe (260)	Video: 11
			The European Library (198)	Repozytorium Cyfr... (49	Spain (97)	
			Federacja Bibliot (89)	Bodleian library (39)	Poland (89)	
			Hispana (60)	Uppsala Universitet (37)	Germany (50)	
			Bayerische Staatsbibliothek (31)	Narodna bibliotek (32)	Sweden (36)	
			Europeana 1914 - ... (25)	Bayerische Staatsbibliothek (31)	Serbia (32)	
			HOPE - heritage o... (21)	Universidad Complutense (30)	Hun gary (30)	
			EuroPhoto (20)	Gredos (27)	Netherlands (19	
			EUscreen Project (10)	Bałtycka Biblioteka (25)	Belgium (9)	
			EuropeanaLocal De... (8)	Lincolnshire Council (24)	France (9)	
			Judaica Europeana (6)		Greece (5)	
			CARARE (5)		Italy (5)	
			askaboutireland (5)		Norway (4)	
			Hellenic Aggregat... (4)		Austria (3)	
			Norsk Kulturråd (4)		Denmark (2)	
			OAPEN (4)		Slovenia (2)	
			Linked heritage (3)		Bulgaria (1)	
			Athena (2)		Finland (1)	
			Békés Megyei Tudá... (2)		Latvia (1)	
			The Great War Archive (2)			
			Biblioteca de Catalunya (1)			
			EFG - The European Film Gateway (1)			
			Swedish Open Cultural Heritage (1)			
			University College Dublin (1)			
			Vlaamse Kunstcollectie (1)			
			Регионална библио (1)			

Fig.1 Example of genealogy search based on the keyword 'Census'

This search shows that there are 908,002 records related to 'census' in an English language key word search. The majority of these come from the Irish Manuscript Commission via the National Archives of Ireland, as genealogy is the major traffic provider

for Irish Archives due to the Irish global diaspora and the subsequent interest in Irish Cultural heritage and genealogy.

Similar searches where done for 51 other key words. What this revealed, is that there are genre based records of use to genealogy and local history researchers, but not in the numbers which might be expected, given the level of digitised materials held in national and local institutions.
Far higher numbers should occur by the researchers of this document. Therefore there is a discrepancy between expected numbers and actual numbers found in Europeana, despite the high numbers of such material available in Europeana through different national archives portals and genealogy services.

This discrepancy has to be analysed in more detail with our content providers, as it is an important issue in limiting the possibilities and value of the Europeana content for genealogical search. A point in this case might be the lack of sufficient metadata provided.
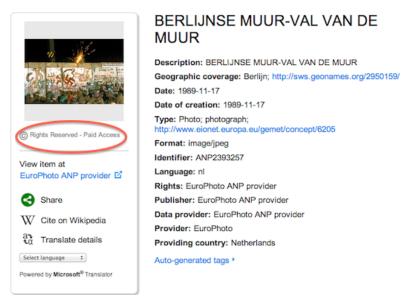
6. Possible contributions to Europeana and licensing/rights clearance

There are a several ways in which institutions could provide content to Europeana, which will be discussed below, and after these alternatives the recommendations from a technical perspective will be made.

Europeana provides all its metadata under a CC0 license, which means that the metadata, not the digital object, can be re-used by the public in any manner it chooses through the use of Europeana's API. This becomes an issue for partner institutions which may view http://www.ancestry.com/cs/us/affiliate genealogy material as a means of monetisation of content which often funds the digitisation of genealogical and local history material. If the institution then provides that metadata to Europeana under the current licensing agreement, the metadata it provides is then open for public re-use. This is thought to affect their ability to generate income from their own content. This issue may be a block for institutions in supplying Europeana with content related to genealogy and local history. As mentioned before, this has to be investigated, as it effects possible developments in genealogy and local history with Europeana content.

Europeana provides twelve licensing options for content contributed to Europeana, and is now engaged in a project to increase the provision of correct licensing labels for all records. Currently there are eight Creative Commons licenses available, and three labels available indicating Rights Reserved. The Rights Reserved labels are the ones most pertinent to genealogy material, because as previously mentioned, a lot of material related to genealogy is available as a paid service and as such often the material can only be found behind a 'paywall'. The most important of the Rights Reserved labels are 'Rights Reserved Paid Access' which points to the digital object as being available via a donation or subscription service, however the metadata surrounding the object is still available as CC0 on Europeana. The application of this license may be one of the best solutions for genealogy and local history data providers since it does not impact their financial ambitions.

Fig. 2 Example of Rights Reserved Paid Access Label

This possible impact on revenue may be one of the greatest hindrances to content provision in relation to this subject. Paying for genealogical information is a common feature in archives and archives use it to cover their costs. Digital services, as is the case with Ancestry, can also be paid for. They definitely facilitate the genealogist, who does not have to travel extensively and can find the genealogy material easier. [6] But tariffs or fees or legal charges vary widely, and also within countries. So do the services provided. In the German Bundesarchiv for instance, all simple research has to be done by the visitor himself and any deeper research has to be handed over to specialised institutions or companies, which of course, charge money.[7]

As the various ways of operating can be licensed equally well, it should be possible to increase Europeana's holdings of this type of material without harming the possibility of organisations in making money themselves. In either case, a substantial increase in visitors of all kind could be expected.

In saying this, it is assumed that 'Rights Reserved Paid Access' labels would be sufficient to encourage more providers to contribute material once they are aware that this option exists. Obviously this requires Europeana doing more education on rights labelling in relation to institutions that have content relevant to this area of research.

7. Metadata and Genealogy

Searches for genealogy and local history focus on micro searches, e.g. a person's name, a location, year or specific event, and as such do not fall under general genre searches. This

---

[6] There is a wide range of fees, legal charges etc. that can be applied. In the Amsterdam city archive, considered expensive by Dutch genealogists, a special action to look for non-scanned material costs €14,95 per quarter of an hour. For an example of legal charges see:
[7] http://www.bundesarchiv.de/benutzung/recherchedienste/index.html.de

means that the metadata provided from content partners and subsequent metadata enrichment by the Europeana Foundation as part of their regular publication cycle is of paramount importance. The richer the metadata provided for these records, the more rewarding the search results for the end user, which benefits both visitors to Europeana and its partner institutions.

Metadata criteria that can be useful to genealogy and local history searches include:
- extended descriptions with names
- locations
- employment information
- comprehensive titles
- geo-tagging
- the provision of a high quality image to the object where the information originates.

Europeana is currently improving its provision of geo-tagging within records as part of its enrichment process. However the provision of descriptions, subject fields, names, and the location used for Europeana's geo-tagging enrichment is at the moment provided at the discretion of the providing institution. The Europeana project *LoCloud (local content in a Europeana cloud)* is planning to address this issue through its microservices.[8]

The standard of the metadata received is also entirely dependent on the providing institution and the contributing National and Project Aggregators. Europeana's metadata schema (the *Europeana Data Model*) does have some mandatory elements, but these are flexible in order to adapt to the wide range of materials and institution standards found in Europeana's 26,7 million objects.

An object's visibility is only as good as the metadata that accompanies it, so the richer the metadata the more likely it is to come up in a search. So in genealogy and local history more comprehensive data is needed to be effective. The field that is most likely to generate effective search results is having a key word visible in the title, description or subject field. These are represented in the *Europeana Data Model* by the presence of *dc:description, dc:coverage, dc:type, dc:title, dcterms:spatial* (indicating place) and *dc:subject* using *Dublin Core* elements as part of the *ORE:Aggregation* field. However only two of these elements are mandatory at this time. Records with more than two mandatory elements filled out are more likely to be found during a genealogy search. This is essential to genealogy.

A problem with the description of genealogical records, including their metadata, also lies with the enormous amounts of available data to be processed and described. To show this, we give an example of one of the actual records found in Europeana within search 'census' as described before in chapter 5 of this report, below.

Fig. 3 Copy of Irish Census Return in the Europeana Portal

Here you can see that there is limited metadata provided with these records. They feature a descriptive title referring to specific locations in Dublin. The description is related to the description of the collection rather than the description of the object. The record within the Europeana Portal only acts as a gateway to the digitised object housed in the National Archives of Ireland, which does contain information pertaining to the contents of the document. This is given to Europeana through the click-through link submitted in the metadata by *europeana:isShownAt* field, the image accompanying the record is submitted through the *europeana:isShownBy* field
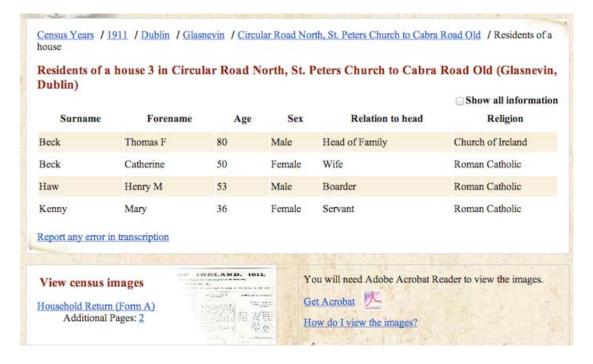

Fig.4 Irish census record as presented by the National Archives of Ireland

However, in this instance the data provider has chosen to given Europeana limited metadata, which effects its return and subsequent traffic on the home portal because the information is purely limited to having previous knowledge of the location of the family on the night of the census in 1911. Since the Irish census is based purely on location rather than permanent residence on any given census night, family members may not be where they are expected. The names of the persons featured in the archival record are not in the metadata provided by the Irish National Archive: therefore this source is not an effective tool for genealogy research.

If the National Archives of Ireland had provided richer metadata, including the information already available on their own site, e.g. names of occupants, this record would have been of greater use to genealogy research. This example also shows that more comprehensive metadata is available.

It may, therefore, be recommended that institutions with genealogy or local history material may provide comprehensive metadata to Europeana, since limited metadata affects the search results, and are less beneficial to end users.

8. Recommendations

Considering the absolute necessity to get richer metadata, meaning all mandatory fields correctly filled out, the quality, and the possible financial implications, it does not appear feasible to focus on primary genealogical search (micro level) through Europeana, at least for the time being.
Europeana cannot facilitate genealogy research easily on a micro level due to the structure of the portal, and the mandatory elements of the metadata required. However, records of national importance, e.g. census records, military registries etc., would be an excellent addition to the Europeana portal, because they are usually already in the national public domain. They would therefore adhere to the current licensing model of the Europeana Foundation, which explicitly stimulates re-use of material by other parties, and would not interfere with the local archives' funding models. The same holds for pictures or other items of any kind that may be linked to persons or families in the genealogy search. Keyword here is enrichment. Europeana and its partners could provide this already within the Europeana technical framework via the portal and via the API.

If Europeana, however, is to accept genealogy and local history records on a micro level around 2018/19, which seems to be the earliest date possible considering the work to be done, the following recommendations are made:
   a. The application of the 'Rights Reserved Paid Access' label to records that appear in full behind a pay wall. Europeana may have to conduct more outreach activities with archival institutions to ensure that they are aware that this is an option for such content (probably the second phase of *Europeana Awareness* WP3i may allow for some of this);
   b. In order to increase genealogy and local history holdings it should be strongly recommended that the fields *dc:subject, dc:coverage, dc:type,*

dc:description, dc:title and *dcterms:spatial* are all made mandatory to fill out, thereby increasing the visibility of the records and allowing better access for end-users and better re-use. This would involve enhancing the *EDM*;

c. In the case of content being submitted, genealogy and local history institutions can restrict the metadata they submit to the minimum. However in doing so this will restrict the visibility of the records;

d. In both cases genealogy institutions should work with the Ingestion team at Europeana to decide the best method of metadata for their collections.

It is, however, a strong recommendation, that, while genealogy and local history researchers are a growing research group, the Europeana Foundation should not alter its current strategy of providing metadata under a CC0 license or accept records, which circumvent its public domain charter, its mission statement, change its access policies relating to the public or its relationship with its data providers and partners in order to cater to a single user group.

Annex: Europeana Rights labelling

Understanding what one can and cannot do with the content represented in Europeana is important if users are to use the repository to its greatest advantage. Currently, there are 12 rights statements that data providers may use to provide rights information about the digital objects (and correspondingly the previews), which should help users to understand the conditions of re-use for each object. These rights statements are explained in the *Europeana Licensing Framework*[9] and in the *Guidelines for the Rights in Objects Submitted to Europeana*.[10]

The available rights statements are:

Public domain: the public domain mark (PDM) must be applied to all content that is in the public domain. Works that are labelled as being in the public domain can be re-used by anyone without any restrictions.

The Creative Commons Zero Public Domain Dedication: if a data provider wants to waive all its rights in a digital object, they can apply this CC0 waiver to the works in question. By applying this waiver, all rights to the content are waived and it can then be re-used by anyone without restrictions.

Creative Commons licences: these provide options for copyright holders to allow others to re-use digitised objects exposed through Europeana under certain conditions. These conditions range from relatively open (only requiring attribution in the case of re-use or redistribution) to relatively restrictive (only allowing the non-commercial redistribution of verbatim copies). Creative Commons licences can only be applied by the rights holder or an entity that has been authorised by the rights holder to apply them. In general, this means that data providers will only be able to CC-license digital objects when they are also the rights holders for these objects.

The following six Creative Commons licences can be used as rights statements for digital objects that are described in Europeana:
- Creative Commons – Attribution (BY);
- Creative Commons – Attribution, Share Alike (BY-SA);
- Creative Commons – Attribution, No derivatives (BY-ND);
- Creative Commons – Attribution, Non-Commercial (BY-NC);
- Creative Commons – Attribution, Non-Commercial, Share Alike (BY-NC-SA);
- Creative Commons – Attribution, Non-Commercial, No Derivative (BY-NC-ND).

Rights Reserved: Europeana has also developed three standardised rights statements that can be applied by data providers to indicate that digital objects can be accessed on their website but may not be re-used by third parties. By applying one of the Europeana rights statements, the data provider is indicating that rights to

---

[9] Source: *Europeana Licensing Framework*
[10] Source: *Guidelines for the Rights in Objects Submitted to Europeana*

the digital object are reserved and that it may not be re-used without additional permissions by the data provider or rights holder. These statements are:

- Right Reserved-Free Access (applicable when users have free, direct and full access to the digitised object);
- Rights Reserved-Paid Access (applicable when users need to pay to gain access to the digitised object itself);
- Rights Reserved-Restricted Access (applicable when users are limited in accessing a digitised object by other means than the need to pay for it, for example when registration is required).

Unknown: the Unknown rights statement can be applied to digital objects when the data provider does not have conclusive information pertaining to the rights status of the digital object. This label may only be used when the copyright status of the work described remains unknown after a diligent search has been carried out.