Project Acronym: EAwareness
Grant Agreement number: 297261
Project Title: Europeana Awareness

Report: Investigating the Genealogy Services Market

Revision	Final 3.0
Date of submission	15 January 2015
Author(s)	Yvette Hoitink, Heritas
Dissemination Level	Public

Report: Investigating the Genealogy Services Market

Project co-funded by the European Commission within the ICT Policy Support Programme

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
1.0	5 December 2014	Yvette Hoitink	Heritas	Draft
2.0	31 December 2014	Yvette Hoitink	Heritas	Second draft
3.0	15 January 2014	Yvette Hoitink	Heritas	Final version

Statement of originality:

This milestone contains original unpublished work except where clearly indicated otherwise.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

1 Introduction

Work package 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, utilising 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'.

Deliverable 3.5 presents the results of Work Package 3, task 3.2.5, MS20, to conduct discussions with a range of commercial suppliers of online services to the genealogy and local history sectors, with a view to delineating sustainable partnership agreements.

2 Goals and methods

This subtask explores the possibilities for collaboration between Europeana and commercial suppliers of online services to the genealogy sector. The report builds on the findings of subtask D3.2: "Assessment of the role of local archival collections"[1] and subtask D3.3: "Analysis of Europeana Content for local history and genealogy users/re-users."[2]

Online searches were conducted to find information about the genealogy market and identify the major companies. Representatives of these companies were approached by visiting

---

[1] Rob Davies, Carol Usher, "D3.2: Assessment of the role of local archival collections," PDF report, version 3.0, 30 June 2013, "WP3 Developing new partnerships," *Europeana Professional* (http://pro.europeana.eu/web/europeana-awareness/outcomes/ : accessed 14 November 2014)

[2] Marie-Claire Dangerfield, Hans de Haan, "Subtask D3.3 – Analysis of Europeana Content for local history and genealogy users-re-users," PDF report, revision 5.0, 3 July 2013, "WP3 Developing new partnerships," *Europeana Professional* (http://pro.europeana.eu/web/europeana-awareness/outcomes/ : accessed 24 November 2014)

stands of these organisations at a genealogy fair and via email and LinkedIn. Negotiations were then held via email and Skype. Publicly available resources about these organisations, such as press releases, were consulted for information about the companies including existing partnership agreements.

Interviews were also held with representatives of Archives Portal Europe (APE) to find out their expectations for partnerships with commercial suppliers. Since Archives Portal Europe was already negotiating with one of these parties directly, the information about possible collaborations with that party was obtained through APE. The preliminary findings were discussed with the genealogy suppliers, APE, the Europeana Awareness project and the Central Bureau for Genealogy, a knowledge centre for genealogy in the Netherlands. Their feedback was incorporated in the final report.

Note about formulations in the report: This report uses terms as defined in the Europeana Glossary.[3] The word "data" is used to describe content plus metadata. To protect the commercial interests of the suppliers, the report present the results of conversations with suppliers on an aggregated level that cannot be traced to an individual supplier.

3 Overview of the genealogy services market

Genealogy is a growing market. A survey by Ancestry.com, which included respondents from Australia, Canada, Germany, Sweden, the UK and the US, found that interest in family history has doubled since 2008 and is expected to double again by 2015. In 2014, 36% of online adults had used the internet to learn more about their family history.[4]

Genealogy In Time Magazine publishes an annual overview of the Top 100 Genealogy websites. The 2014 report includes an overview of the 100 most visited genealogy websites and identifies the three companies that dominate the genealogy market:[5]

Company	Most visited websites	# Websites in Top 10	# Websites in Top 50	Estimated % of genealogy traffic
Ancestry.com	Ancestry.com (1) Find a Grave (2) Ancestry.co.uk (6)	5	16	32%
MyHeritage	MyHeritage.com (4) Geni.com (5) MyHeritage.no (20)	2	13	8%
BrightSolid (DC Thomson Family History)	FindMyPast UK (18) Genes Reunited (30) ScotlandsPeople (48)	0	6	4%

---

[3] "Glossary," *Europeana* (http://pro.europeana.eu/glossary : accessed 23 December 2014)

[4] Ancestry.com, "Ancestry Global Family History Report," presentation, *Slideshare* (http://www.slideshare.net/ancestry/global-family-history-report : published 26 November 2014)

[5] "Top 100 Genealogy Websites of 2014," *Genealogy In Time Magazine* (http://www.genealogyintime.com/articles/top-100-genealogy-websites-of-2014-page01.html : accessed 24 November 2014)

Together, these three companies own 70% of the top 10 websites and 35% of the top 100 websites. Of the 15 new websites in the Top 100, 12 were European, which suggests that genealogy is now growing more rapidly in Europe than in the US, traditionally the biggest market for genealogy.[6]

Another major player is *Familysearch*, the third most visited website, which is not operated by a commercial company but is a non-profit organisation supported by the Church of Jesus Christ of Latter-day Saints.

Introduction of the four major genealogy suppliers

Ancestry.com

Ancestry.com is the largest genealogy company in the world with approximately 2.7 million paying subscribers, providing access to approximately 15 billion records and 60 million family trees.[7] In 2013, Ancestry.com made $540,000,000 in revenue from subscriptions and other products and made a net loss of $80,000,000.[8]

Ancestry.com's corporate headquarters are in Provo, Utah. Ancestry.com has several European websites that focus on European collections, including Ancestry.co.uk (United Kingdom), Ancestry.de (Germany), Ancestry.se (Sweden) and Ancestry.it (Italy).

Examples of partnerships of Ancestry.com:
- Joint agreement with Familysearch for a five-year effort to digitise and index more than one billion records from around the world.[9]
- Janus project with German archives, whereby Ancestry digitises and indexes German collections and offers digital copies to the archives.[10]

Familysearch

Familysearch is a non-profit family history organisation supported by the Church of Jesus Christ of Latter-Day Saints. For members of the LDS church, doing family history is a part of acting on their beliefs that family relationships are forever. Besides the Familysearch website, the Familysearch organisation also has almost 5,000 FamilySearch Centers worldwide.[11]

---

[6] "Top 100 Genealogy Websites of 2014."

[7] " Ancestry.com LLC Reports Third Quarter 2014 Financial Results," press release, *Ancestry.com* (http://corporate.ancestry.com/press/press-releases/ : accessed 18 November 2014)

[8] "Ancestry.com LLC Reports Fourth Quarter and Full Year 2013 Financial Results," press release, *Ancestry.com* (http://corporate.ancestry.com/press/press-releases/ : accessed 18 November 2014)

[9] "FamilySearch and Ancestry.com Working Together to Make More Records Available Online," blog post, *Familysearch* (https://familysearch.org/blog/en/familysearch-ancestrycom-working-records-online/ : accessed 19 November 2014)

[10] "Vorteile einer Ancestry Partnerschaft," *Janus – ein Ancestry Project* (http://www.janus-projekt.de/digitalisation/partnerschaft : accessed 19 November 2014)

[11] "About Familysearch," *Familysearch* (https://familysearch.org/about : accessed 19 November 2014)

The LDS church has been creating image copies of archival holdings for over 100 years, first on microfilm but now digitally.[12] Their main repository, in a granite mountain vault in Salt Lake City, contains over 2.4 million rolls of microfilm, presenting over 3.5 billion images. The microfilmed collections are being digitised and made available on Familysearch.org.[13]

Examples of partnerships of Familysearch:
- Partnership with National Archives of Ireland and FindMyPast to provide access to the Irish 1901 census.[14]
- Partnership with OCLC to catalogue the microfilm holdings of Familysearch in Worldcat.[15]

MyHeritage

MyHeritage is a family history network with a focus on combining technology and genealogy. MyHeritage has over 75 million subscribers who created 27 million family trees. Their website, which is available in 40 languages, provides access to 5.5 billion historic records.[16]

Examples of partnerships of MyHeritage:
- Strategic partnership with Familysearch, whereby MyHeritage provides its SmartMatching technology to Familysearch, which matches records to family trees, and Familysearch provides access to 2 billion records to MyHeritage.[17]
- Partnership with BillionGraves to digitise cemeteries throughout the world.[18]
- Partnerships with RootsMagic, Aldfaer, GenealogieOnline and FamilyHistorian to embed SmartMatching technology into tree building software.[19]

DC Thomson Family History

DC Thomson Family History is a British-owned genealogy company. The company has 18 million registered users and provides online access to over 1.8 billion genealogical records.[20]

---

[12] "The Future Looks Exciting for FamilySearch's Record Digitization Program," *Familysearch* (https://familysearch.org/blog/en/future-exciting-familysearchs-record-digitization-program/ : accessed 23 November 2014).

[13] "Granite Mountain Records Vault," *Familysearch* (https://familysearch.org/archives/about-granite-mountain/ : accessed 25 November 2014)

[14] "Pre-1901 Irish census records online for the first time," press release, *FindMyPast* (http://blog.findmypast.co.uk/ : accessed 25 November 2014)

[15] " Genealogists can now find FamilySearch and WorldCat records through both online resources," *Familysearch* (https://familysearch.org/blog/ : accessed 25 November 2014)

[16] "Media Kit – MyHeritage.com," *MyHeritage* (http://blog.myheritage.com/media-kit-myheritage-com/ : accessed 1 December 2014)

[17] "MyHeritage and FamilySearch Enter Into a Significant Strategic Partnership," press release, *Familysearch* (https://familysearch.org/node/2410 : created 14 October 2013)

[18] "BillionGraves Partners with MyHeritage to Digitize the World's Cemeteries," blog post, *Familysearch* (https://familysearch.org/blog/ : created 2 May 2014)

[19] "RootsMagic Adds MyHeritage Matching Technologies for Powerful Automatic Research Capabilities," press release, 25 November 2014; "Family Historian Genealogy Software Integrates MyHeritage Matching Technologies for Automated Discoveries," press release, 24 November 2014; "MyHeritage Bolsters Leadership in the Netherlands with Strategic Partnerships and First National TV Advertising Campaign," press release, 14 November 2014; all available via *MyHeritage* (http://blog.myheritage.com/media-kit-myheritage-com/ : accessed 4 December 2014).

Examples of partnerships of DC Thomson Family History:

- Agreement with Familysearch to collaborate on a wide range of projects including digital preservation and technology, whereby 13 million records from Familysearch were made available on Findmypast.com.[21]
- Partnership with the National Records of Scotland and the Court of Lord Lyon whereby DC Thomson Family History operates the website ScotlandsPeople.[22] This partnership was the result of a tender that was won by BrightSolid (the parent company of DC Thomson Family History) in 2009.[23]

4 Data provided by genealogy suppliers

Online access to genealogical information is provided in a number of ways. The different modes of access can have legal and technical implications for collaboration projects.

Modes of access

Archival descriptions and finding aids

Repositories create archival descriptions of their holdings and create finding aids that describe the contents of record groups and how to use them. This allows users and suppliers to identify record groups or individual records. These descriptions are high-level (e.g. 'Vienna population cards 1850-1895') and do not facilitate micro-searches that genealogists typically use, such as searching for names and dates. These descriptions are usually created by the cultural heritage organisations that hold the original objects.

Digital objects

Using scanners and cameras, digital images are created of the records. Scans can be made from the originals, or from preservation copies of the sources such as microfilms. Most genealogy suppliers have their own digitisation projects in collaboration with cultural heritage organisations, which includes a license agreement for use of the resulting digital objects. In other cases, cultural heritage organisations digitise their own content, which they may license to genealogy suppliers. The license may restrict re-use of the digital objects and may be for a limited time only.

Indexes and databases

Indexes are created from sources by manually compiling a list of all the people, places or organisations in a source. Databases are more elaborate than indexes and transform the

---

[20] "About," *DC Thomson Family History* (http://www.dcthomsonfamilyhistory.com/about/ : accessed 1 December 2014)

[21] "DC Thomson Family History and FamilySearch.org to Make Billions of Records Available for People to Search," press release, *Familysearch* (https://familysearch.org/node/2412 : created 16 October 2013)

[22] "About ScotlandsPeople," *ScotlandsPeople* (http://www.scotlandspeople.gov.uk : accessed 25 November 2014)

[23] "Brightsolid wins ScotlandsPeople tender," press release, Alan Stewart, blogger, *Grown Your Own Family Tree* (https://growyourownfamilytree.wordpress.com/ : created 8 September 2009)

information in the source into structured metadata where the different elements of information are entered in separate fields. Indexes and databases are usually created by humans (paid or volunteers). They can be created from the originals, from microfilm or from scans. Most genealogy suppliers create their own indexes and databases and own the rights to these sets of metadata.

Transcripts and abstracts

A transcript is a literal copy of a document, with all spelling, grammar, and punctuation exactly as it was found.[24] Optical Character Recognition (OCR) technology can be used to create full-text versions of typed sources, although recognition is not perfect. Making transcripts of hand-written sources is usually done by humans; automatic recognition of these manuscripts is still in the research phase.[25] Abstracts are summaries that record all important details from a whole document.[26] Abstracts can be included in a database. Most genealogy suppliers do not create transcripts or abstracts because of the investment needed.

Family trees

A family tree, the colloquial term for a database that stores the research results of a genealogist, is an authored work, combining information from a variety of sources. Most genealogy suppliers allow users to create genealogical databases, either individually or collaboratively (Wiki-style) and discover and attach historical records to their trees.

Genealogy suppliers usually offer access to historical records in one of three ways:
- *Index/database plus images.* This is the level of access that genealogy suppliers strive for. The indexes and databases allow the user to find the record that they need. By providing the digital image in addition to the image, the user can then read the whole record himself. The added value of a costly transcript or abstract would be small, especially in the case of records that are easily legible.
- *Index/database only.* This is the case where a license for the images has not been obtained or if the source has not scanned yet.
- *Images only.* This happens when content has been digitised but has not been indexed yet. Since the discoverability of images is low, providing access to image-only datasets is usually a temporary situation until the index has been completed. Because of the time involved in creating an index, publishers may choose to publish the images without the indexes in the meanwhile.

---

[24] Mary McCampbell Bell, "Transcripts and Abstracts," in Elizabeth Shown Mills, editor, *Professional Genealogy : a manual for researchers, writers, editors, lecturers, and librarians* (Baltimore: Genealogy Publishing Co, 2001), chap. 16, p. 293.

[25] See for instance, Lambert Schomaker, *Monk* (http://www.ai.rug.nl/~lambert/Monk-collections-english.html : accessed 1 December 2014) for an academic project that aims to use artificial intelligence to improve access to historical collections which are difficult to process by traditional OCR methods.

[26] Bell, "Transcripts and Abstracts," 294.

Report: Investigating the Genealogy Services Market

European data currently available from genealogy suppliers

The genealogy suppliers provide access to content about billions of people. Examples of European data from genealogy suppliers that may be of interest to Europeana are:

Country	Source	Size	Index /database	Digital Images	Ancestry	Familysearch	MyHeritage	DC Thomson FH
			Mode		Available from			
Belgium	Civil Registration and Church records, 1582-1910	>19M images		X		F		
Croatia	Church Books 1516-1994	>1.4M images		X		F		
Czech Republic	Censuses, 1800-1945	>3M images		X		F		
Denmark	Estate records, 1436-1964	>2.6M images		X		F		
Germany	Hamburg passenger lists 1850-1934	>8M persons	X	X	P			
Germany	Bavaria World War I personnel rosters, 1914-1918	>8M persons		X	P			
Hungary	Civil Registration 1801-1980	>5M images		X		F		
Italy	Civil registration, 1806-1900	>10M persons	X				P	
Netherlands	Population registers, 1574-1940	>2.5M images		X		F		
Lithuania	Census lists, 1795-1900	>400K persons	X		P			
Portugal	Catholic church records, 1459-1913	>10M images		X		F		
Spain	Testaments	>1M images		X		F		
Sweden	Church records 1500-1941	>35M persons	X	X	P			
Sweden	Baptisms 1611-1920	>26M persons	X			F	P	
United Kingdom	1911 census of England and Wales	>33M persons	X	X	P	F	P	P
United Kingdom	WWI Service medal and award rolls, 1914-1920	>6M persons	X	X	P			
United Kingdom	British Newspaper archive	>9M pages	X	X				P
United Kingdom	FreeBMD Birth, marriage, death index, 1837-1915	>140M persons	X			F		

X = available online, F = Free access, P = Paid access, M = millions, K = thousands

5 Assets in collaboration agreements

Assets in partnership agreements

Analysis of the existing partnership agreements of the four major genealogy suppliers shows that the following assets can be involved in partnership agreements:

- Access to physical objects
- Digital images of records
- Catalogue information (archival descriptions)
- Indexes and databases

- Exposure to reach a wider audience
- Technical capabilities
- Money

For example: German archives provide access to the original records in return for the digital images. MyHeritage provides SmartMatching Technology to RootsMagic in return for exposure. Note: press releases rarely specify if money was involved in these transactions.

Analysis of potential assets of Europeana and genealogy suppliers

The first step in considering partnerships between Europeana and genealogy suppliers is an analysis of the assets that each party could bring to an agreement. The next table summarises the current situation.

	Europeana	Genealogy Suppliers
Access to physical objects	No, only through participating cultural heritage organisations.	No, only through cultural heritage organisations.
Digital images of records	No, only through data providers.[a]	Sometimes (depends on license and rights situation).[c]
Catalogue information (archival descriptions)	Yes, all metadata is provided as CC-0.	Yes, but probably not in a format that is directly usable for Europeana.[d]
Indexes and databases	Yes, all metadata is provided as CC-0. Genealogical metadata is limited.[b]	Yes (sometimes licensed, often created themselves).[e]
Exposure to reach a wider audience	Yes. In 2013, Europeana received 1.8 million visitors that generated 6.9 million pageviews.[27]	Yes. The suppliers each have millions of subscribers or registered users.
Technical capabilities	Research reports are publicly available.	Technical infrastructure for handling genealogical data. Matching technology to find records based on personal information. Multilingual interfaces.
Money	No. Europeana would be interested in new ways to fund its infrastructure and does not want to pay for providing access to (commercial) genealogical data.	None of the genealogy suppliers were interested in paying for the increased exposure, beyond investing employee time if the business case is solid.

Ad a (digital images provided by Europeana): License agreements between Europeana and the data providers do not give Europeana any rights to the digital objects.[28] Instead, Europeana displays the thumbnails that link to the digital objects on the websites of the data providers. Because Europeana does not have the rights to the digital objects, any

---

[27] "Traffic & Usage: 2013," *Europeana Statistics Dashboard* (http://statistics.europeana.eu/ : accessed 29 November 2014)

[28] "The Europeana Licensing Framework," PDF, "Publications," *Europeana Professional* (http://pro.europeana.eu/publications : accessed 18 November 2014)," p. 9.

agreements to provide the digital objects to genealogy suppliers need to be made between the genealogy supplier and the data provider, not by Europeana.

Ad b (Indexes and databases provided by Europeana): Deliverable D3.3 of the Europeana Awareness program contains an analysis of Europeana Content for local history and genealogy users/re-users. The report identified several sources in Archives Portal Europe (APE) that would be of potential interest to genealogy users, but found that the metadata was often limited to high-level archival descriptions. Information like personal names and geographical names were not included in the metadata supplied to APE, which makes it unsuitable to search for names.[29]

Archives Portal Europe has started adding person-based metadata in the form of descriptions of archival creators using the Encoded Archival Context – Corporate Bodies, Person, Families (EAC-CPF) metadata standard.[30] Since only the archival creators are described, not the persons mentioned in the records, the number of EAC-CPF records ingested into APE has been small. At the end of 2014, less than 20,000 people and entities were described in APE.[31]

Europeana would be allowed to distribute the existing metadata, since Europeana provides all its metadata under CC-0 license. However, this metadata does not contain much detailed genealogical information and would be of limited use to genealogy suppliers, limiting their business case to present Europeana data on their platforms.

Ad c (digital images held by genealogy suppliers): In many cases, the genealogy suppliers only have a license to use the images on their own website. Europeana would require previews, not the full images, as described in the Data Exchange Agreement.[32] Whether the license that the genealogy supplier has would allow for previews on other websites will have to be ascertained for each case individually.

For any digital objects supplied to Europeana, the metadata should indicate the rights information relating to the digital objects.[33]
- *For works that are not in the Public Domain*, the "Paid Access – No Reuse" statement can be attached to digital objects that require a subscription or one-time fee.[34]
- *For works that are in the Public Domain*, Europeana prefers that the "Public Domain Mark" be attached to the digital objects.[35] Since a significant part of the data published by genealogical suppliers involves paid access to Public Domain Content,

---

[29] Dangerfield, "Subtask D3.3."

[30] Karin Bredenberg, "Records creators - Use of EAC-CPF in Archives Portal Europe," article, *APEx* (http://www.apex-project.eu/index.php/en/articles : accessed 29 December 2014)

[31] *Archives Portal Europe*, home page (http://www.archivesportaleurope.net/ : accessed 29 December 2014).

[32] "Europeana Data Exchange Agreement," *Europeana Professional* (http://pro.europeana.eu/web/guest/data-exchange-agreement : accessed 4 December 2014)

[33] "The Europeana Licensing Framework," 7.

[34] Marie-Claire Dangerfield, Paul Keller, "Europeana digital objects to have valid rights statement by July 2014," blog post, *Europeana Professional* (http://pro.europeana.eu/pro-blog/ : published 26 February 2014)

[35] "The Europeana Licensing Framework," 7.

attaching this rights statement won't be acceptable to the suppliers. Possible deviations from the Europeana guideline to attach the "Public Domain Mark" to objects in the Public Domain need to be determined on a per-case base. Exceptions can be made, for example if national law allows copyright claims over digitised works or if the works were digitised in a Public-Private Partnership where limits to commercial re-use were part of the agreement.

Ad d (catalogue information and archival descriptions held by genealogy suppliers): Catalogue information and archival descriptions held by genealogy suppliers are usually not in the format required by Archives Portal Europe (Encoded Archival Descriptions, an XML-based international standard), so a conversion will be necessary. APE has developed a "Data Preparation Tool" to aid with conversion.[36]

Ad e (indexes and database held by genealogy suppliers): A potential issue with supplying indexes and databases by genealogy suppliers is that Europeana requires that all metadata, including genealogical metadata, be provided under a CC-0 license. The report from subtask 3.5 recognised that not all parties may be willing to contribute their metadata under CC-0, since that would limit options to monetise the material.[37]. The interviews with genealogy suppliers revealed that even without the requirement for CC-0, the investments needed to create these indexes limit their willingness to share these indexes.

Deliverable D3.3 of the Europeana Awareness program also found that the structure of the Archives Portal Europe does not currently facilitate genealogy research easily on a microlevel.[38] One option to include detailed genealogical metadata would be to encode it as EAC-CPF, a format supported by APE. However, the EAC-CPF standard is meant for creating authority records which provide descriptions of entities (corporate bodies, persons and families).[39] By contrast, genealogical metadata does not consist of authority files, but one person may appear multiple times in the same record set or in different record sets where the information about the person can be different in each record. It is up to the researchers to identify the different records as the same person and aggregate the information.

Without using EAC-CPF, the infrastructure of Europeana is not set up to ingest rich genealogical metadata even if genealogy suppliers were willing to share this data.

6 Models for collaboration

Based on the analysis of the different assets, several models for collaboration were examined in more detail.

---

[36] "Tools and Manuals," *APEx* (http://www.apex-project.eu/index.php/en/outcomes/tools-and-manuals : accessed 29 December 2014)

[37] Dangerfield, "Subtask D3.3."

[38] Dangerfield, "Subtask D3.3."

[39] "International Standard Archival Authority Record For Corporate Bodies, Persons and Families," 2nd ed., 1 April 2004, PDF, *International Council on Archives* (http://www.icacds.org.uk/eng/ISAAR%28CPF%292ed.pdf : accessed 29 December 2014), p. 8.

Contributing Europeana data to genealogy suppliers

One option for collaboration between Europeana and genealogy suppliers is by providing Europeana data to the suppliers. Europeana would contribute digital images and/or indexes in return for exposure and/or money.

The report "Analysis of Europeana Content for local history and genealogy users/re-users" found that although Europeana provides access to a significant number of objects related to generic genealogical search terms (e.g. "census"), detailed genealogical metadata is absent.[40] In discussions with genealogy suppliers, they indicated that indexes, not the images, were the most valuable part of the genealogical data, because of the investment needed to create them. Their subscribers are searching for their ancestors' names, so content that does not include names in the metadata will not be found. The current Europeana data, lacking these descriptive metadata, does not support name searches and for that reason is not that interesting to genealogical suppliers.

Another issue with this collaboration model is that in most cases, Europeana does not have the license to distribute the digital objects so any collaboration that contributes data from Europeana to genealogy suppliers would be limited to the metadata or would need to involve the data providers as well.

There are some Europeana datasets where Europeana does have the license to distribute the digital objects and the metadata, such as the personal papers and memorabilia of some 7,000 people involved in World War I which were collected in the *Europeana 1914-1918* project.[41] Datasets like these could be provided to genealogy suppliers by Europeana.

Before committing to such a collaboration, the genealogy suppliers would weigh the investments needed to harvest the data against the added value of providing access to the data. Since genealogy suppliers are used to dealing with datasets covering millions of persons, the value of a dataset like *Europeana 1914-1918* with 7,000 people will be limited to them so is unlikely to lead to a positive business case.

Because Europeana does not have the license to distribute the digital objects for the majority of datasets and investigating the possibilities for re-use of individual datasets was outside the scope of this project, this has not been investigated further at this time.

Contributing data from genealogy suppliers to Europeana

Another option is to contribute the data from genealogy suppliers to Europeana. In this case, the genealogy suppliers would contribute digital images, archival descriptions and/or indexes in return for exposure on Europeana. Chapter 4 identified record sets that could complement Europeana's current offerings.

---

[40] Dangerfield, "Subtask D3.3."
[41] "The First World Centenary: the site that brings all sides together launches in Berlin," press release, 28 January 2014, *Europeana Professional* (http://pro.europeana.eu/pressrelease/29jan : accessed 29 December 2014)

Report: Investigating the Genealogy Services Market

The business case for the suppliers to present their data on Europeana would be to attract new (paying) clients to their website.

For the genealogy suppliers to contribute their data to Europeana, the genealogy supplier and Europeana would need to enter into a Data Exchange Agreement (DEA).[42] Whether the genealogy supplier would need authorisation from the cultural heritage organisation that supplied the content or metadata or whether the cultural heritage organisation would need to co-sign the DEA depends on the license that the genealogy supplier has from the cultural heritage organisation.

Europeana has the following requirements for the rights to the metadata and the digital objects: [43]
- All metadata, including genealogical metadata, needs to be provided as CC-0.
- The metadata should include a rights statement that indicates the rights information relating to the digital objects.

The data from genealogy suppliers could be shared on different levels, which have implications for the legal and technical feasibility of the collaboration.

Sharing level	Acceptable to supplier?	Acceptable to Europeana?
Archival description	Possibly, not very interesting because it not expected to drive much traffic towards their website.	Probably not; not engaging enough for visitors.
Archival description + thumbnails of records	Possibly. Could drive traffic towards their website. Depends on license agreements with the cultural heritage organisations and rights situation.	Yes, if acceptable rights statement is attached. Feasible with current infrastructure.
Archival description + index/database	Probably not if the metadata needs to be provided under a CC0-license, because that would limit their options to monetise the material.[44]	Maybe, but not feasible with current infrastructure.
Archival description + index/database + thumbnails of records	Probably not because of investments in creating indexes and requirement to provide metadata as CC0. Depends on rights situation.	Yes, but not feasible with current infrastructure.

This analysis demonstrates that using the current infrastructure, providing access to digital images and archival descriptions could be a feasible collaboration model that is acceptable to both parties. Two types of datasets are the most obvious candidates to start with:
- Datasets for which the original works are not in the Public Domain, which can be made available using the "Paid Access – No Re-use" statement.
- Datasets for which the original works are in the Public Domain which are available without restrictions or for non-commercial use, which can be made available using the

---

[42] "The Europeana Licensing Framework," 9.
[43] "The Europeana Licensing Framework," 7.
[44] "The Europeana Licensing Framework," 9.

> "Public Domain Mark" or "Out of copyright – Non Commercial Reuse" rights statement.

Participating in a Europeana infrastructure for genealogy

The current Europeana infrastructure is not set up to deal with rich genealogical metadata, but the genealogy suppliers already have such an infrastructure in place. Preliminary discussions with the genealogy suppliers showed that they would not be interested in contributing to the Europeana infrastructure since their focus is on their own infrastructure.

Some of the parties would potentially be interested in joining a Best Practice Network for genealogical data, provided that a business case could be made. A business case could include funding by Europeana or increased exposure of their data.

Collaboration on technology for multilingual access

Providing multilingual access is a prime concern of Europeana since the users of the data may not speak the same language as the creators and describers of the data. Genealogy suppliers face the same challenges. Sharing technology to handle this multilingual access challenge could be another opportunity for collaboration between Europeana and genealogy suppliers.

The need for multilingual access to Europeana data was addressed in the *EuropeanaConnect* project. The project made Europeana more multilingual by giving users the possibility to search in more than one language and to have texts automatically translated.[45] A Task Force within EuropeanaTech investigated using semantic enrichment strategies to match concepts across language boundaries.[46]

Most of the genealogy suppliers also provide multilingual access to their collections, suggesting that sharing technology or best practices might provide another opportunity for collaboration. A survey of the websites and discussions with the suppliers showed that multilingual access to the genealogy suppliers' websites is currently limited to translated interfaces and translated editorial content. The contents of the searchable databases (metadata) are mostly provided in the original language.

Since the technology for multilingual access used by the genealogy suppliers is not more advanced than the technology currently being used and researched by Europeana, a collaboration on this topic is not likely to be of interest to Europeana. The outcomes of the Europeana research projects are already available to the public, including genealogy suppliers, so cannot be used as incentive to collaborate with Europeana.

---

[45] "Europeana Connect," *Europeana Professional* (http://www.pro.europeana.eu/web/europeanaconnect : accessed 2 December 2014)

[46] "EuropeanaTech Task Force on a Multilingual and Semantic Enrichment Strategy: final report," PDF, 7 April 2014, "Report on multilingual semantic enrichment," *Europeana Professional* (http://pro.europeana.eu/web/network/europeana-tech/ : accessed 4 December 2014)

7 Contributing descriptions and images to Europeana

The analysis in the previous chapter showed there was one promising model for collaboration between Europeana and genealogy suppliers that can be realised with the current infrastructure: Contributing archival descriptions and thumbnails to Europeana by genealogy suppliers, in return for additional exposure of their data. Three out of four suppliers expressed interest in such a collaboration model. The fourth supplier did not have the necessary licenses to share their data at this level.

A demo was created of what the 1911 UK census would look like in the context of Archives Portal Europe. The 1911 UK census is a source that is available on the websites of all four suppliers. The original work is protected by Crown Copyright in the UK, which would make the "Paid Access – No Re-Use" rights statement acceptable to Europeana. This demo can be used by suppliers and Europeana to discuss the merits of such a collaboration.

The demo shows how the hierarchy of records in the 1911 census is presented. By navigating the tree on the left, the visitor can browse the records. Thumbnail previews of the digital objects can be clicked to view the original records on the website of the genealogy supplier. On the website of the supplier, the image may be viewed immediately (in the case of free access), or visitors may have to subscribe or pay-per-view first (in the case of paid access).

Demo version of the 1911 census in Archives Portal Europe

8 Conclusions and recommendations

Genealogy suppliers have a lot to offer to Europeana: their data complements Europeana's current offerings. They have an existing network of cultural heritage organisations that supply content to them. Collaborating with these genealogy suppliers can make the data of these organisations available in Europeana. This contributes to Europeana's "Aggregate" strategic track wherein Europeana will "extend [its] network of content providers and encourage the development of aggregators that fit the needs of different countries, domains and users."[47]

Three main factors limit the collaboration opportunities:
- *Licenses*: Europeana and genealogy suppliers do not always have the rights to redistribute their data.
- *Rights statements:* As a guideline, Europeana requires that Public Domain material be marked using the "Public Domain Mark," which does not allow for access restrictions. In many cases, commercial suppliers want to restrict access to paying users and won't be willing to provide the material using the "Public Domain Mark."
- *Technical:* The Europeana infrastructure is not set up to deal with detailed genealogical metadata.

Two feasible collaboration opportunities were identified. For the short term, Europeana could ingest collections of genealogy suppliers at the level of archival descriptions and thumbnails. That is the level of access that is currently supported by Archives Portal Europe, so the existing infrastructure is capable of handling this information. The holdings would be presented in catalogue form with previews (thumbnails) with a rights statement that explains the rights information for the digital objects on the genealogy supplier's website. These can either be digital objects that are not in the public domain, or digital objects that are in the public domain where the supplier provides free access, as any other rights statement would undermine Europeana's commitment to the Public Domain.[48] The genealogy suppliers would have to make an assessment of the investments necessary to supply their data in the format required by APE and compare that to the expected extra exposure to determine if a positive business case can be made.

For the longer term, a Best Practice Network could be set up to create standards, processes, infrastructure and collaboration agreements for ingesting genealogical data into Europeana. Genealogy suppliers can be approached to act as knowledge partners and data providers in this Best Practice Network.

Specific datasets where Europeana does have the license to distribute the digital objects may lend itself to other collaboration opportunities. This can be investigated on a per-case basis.

---

[47] "Europeana Strategic Plan 2011-2015," PDF-file, *Europeana Professional* (http://pro.europeana.eu/publications : accessed 29 December 2014)

[48] For information about Europeana's views on the Public Domain, see the "Europeana Public Domain Charter," PDF, April 2010, *Europeana Professional* (http://pro.europeana.eu/publications : accessed 31 December 2014)