



Deployment and Maintenance of Europeana DSI core services - SMART 2016/1019

CONTRACT NUMBER - 30 - CE - 0885387/00-80

DELIVERABLE

C.3 Data access pattern report M12

Revision	1.0
Date of submission	31 August 2018 (M12)
Author(s)	Dasha Moskalenko, Europeana Foundation
Dissemination Level	Public



Co-financed by the European Union
Connecting Europe Facility

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
0.1	15-8-2018	Dasha Moskalenko	Europeana Foundation	Final version
1.0	21-8-2018	Victor-Jan Vos, Julia Schellenberg	Europeana Foundation	Final review

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.

'Europeana DSI is co-financed by the European Union's Connecting Europe Facility'.

TABLE OF CONTENTS

Executive summary	4
1 Introduction	5
2 Data access questions	6
3 Data access methods	7
4 User retention on Europeana Collections	8
5 Data access measurements	10
5.1 Quantitative measurements based on user searches and visited content	10
5.1.1 Top 10 blog posts	10
5.1.2 Top 10 galleries	11
5.1.3 Top 10 items accessed	12
5.1.4 Items shared on social media that received most engagement	13
5.1.5 Top 10 sources of traffic to Europeana Collections	15
5.2 Quantitative usage of the API	17
5.2.1 Content accessed via SPARQL endpoint	17
5.2.2 Content accessed via OAI-PMH harvesting	17
5.3. Link maintenance activity	18
6 Analysis of the data access measurements	18
7 Conclusion	20

Executive summary

This report describes the data we gather and the methods and tools we use to evaluate data access patterns of our users. The results gained shall help us to assist our users in finding the objects relevant to their needs.

The report provides an overview of the most popular blog posts, galleries, items and social media shares and analyzes those to find subjects our users were most interested in (between April and July 2018). Detailed statistics on API data access patterns are not yet fully extractable and will complement the evaluation in future reports.

1 Introduction

The purpose of this report is to show how users interact with content and metadata on Europeana Collections, thus their most common data access patterns. Further, to analyse content and metadata accessed by users with the aim to anticipate and serve our users the objects they need.

The research question that we address in this report is “How can we increase the amount of returning users on Europeana Collections?”

This question is interesting to explore because returning users are fulfilled by our service. These users are satisfied with the experience and are likely to use the product or service again and promote it by word of mouth to their friends and acquaintances.

In the period April to July 2018 Europeana Collections had 11.6% returning users. This number is relatively low, and we aim to increase it in the period of Europeana DSI-4.

2 Data access questions

The research question that we address in this report is “How can we increase the amount of returning users on Europeana Collections?”

This question is interesting to explore because returning users are fulfilled by our service. These users are satisfied with the experience and are likely to use the product or service again and promote it by word of mouth to their friends and acquaintances. Thus, returning users are cheaper to the business, as the business doesn't need to spend as much money on advertising and giveaways.

Word of mouth promotion is now more valuable than ever. We are overwhelmed by choice in this time. For every product and service, there are multiple competitors all competing for the attention of their target market. Each is shouting into the void how great his or her product or service is and how it is the answer to all your problems. However, how often have you tried a product or service and realized that the company did not fulfill its promise? You left unsatisfied and vowed not to use this service or product ever again. These dissatisfied users also take it upon themselves to spread the word and warn others in their surrounding not use this product or service. The promotion of this negative first-hand experience can hurt a business.

On the positive side, satisfied users are also very vocal about their positive experience. They give your business the most valuable promotion by telling others of their first-hand experience with your product or service and how it changed their lives for the better.

Many businesses have adopted and integrated this word of mouth promotion into their online platform by offering a review section and showcasing it prominently next to the service or product. This section is highlighting to users how satisfied these people are and that you can be one of them. You also see this promotional aspect manifesting in the form of quotes from satisfied customers.

There is even a well-read article that touches on this written by Kevin Kelly called “[1000 true fans](#)” which emphasizes the importance of having a critical mass of very engaged audiences that love your product or service and will be faithful to it in use and word of mouth promotion.

In this report, we will analyze data access patterns to understand what data engages our users and how we can capitalize on that to offer and provide an even better experience that will motivate our users to return to Europeana Collections.

3 Data access methods

We collect data on data access patterns from the following sources:

Google Analytics: this tool gives an overview of the items, blog posts, exhibitions and galleries that received most visits on Europeana Collections over a certain period of time and the main sources of traffic. This gives us a good overview of the most popular content we hold, what our users are searching for on Europeana Collections (e.g. what search terms they use to find the information they are looking for) and in search engines to land on Europeana Collections.

API key logs: another source of information is the use of our API keys¹. We looked at how many active API users² we have (users that were either active for more than five days or made more than five accesses in a month) and which APIs are the most popular. For this activity we rely on our logs.

SPARQL: besides offering access for API users to Europeana Collections (through the Search API³), Europeana also offers a SPARQL⁴ endpoint as a more advanced way for researchers to search through Europeana content. The SPARQL endpoint facilitates a more precise interaction with our data and enables one to search and correlate data based on specific metadata fields and data patterns⁵. Logging and analyzing SPARQL usage offers a granular view on how users, in this case researchers interact with our data and which they find most relevant for their research.

OAI-PMH: not all users of the API interact with our data in a synchronous way. Some require that data is collected beforehand and indexed on their side so that they can provide a better service to their users. For this purpose, Europeana offers a OAI-PMH service⁶ which allows API users to selectively gather a large amount of metadata records based on datasets and when the metadata records have been last updated (or created). This also makes it possible for API users to keep data up-to-date with the data that Europeana holds. Logging and analysing which datasets are being accessed and the frequency of their update gives us a general idea of what content is seen to be more relevant and the level of demand.

¹ An API user may reflect one or more actual users given that an API key can be associated to a single platform/portal such as Europeana Collections (which has its own API key).

² The Europeana REST API receives an average of 45 million requests from about 200 different users every month. From these we consider as active users the ones that demonstrate either a relatively frequent or high use of the API as they reflect services or applications that contribute to the reuse of the Europeana content in several of its markets.

³ Europeana Search API. View at <https://pro.europeana.eu/resources/apis/search>

⁴ Europeana SPARQL API. View at <https://pro.europeana.eu/resources/apis/sparql>

⁵ E.g. what are the French 18th-century painters with at least five artworks available through Europeana

⁶ OAI-PMH service. View at <https://pro.europeana.eu/resources/apis/oai-pmh-service>

Broken links: we also analyse how many broken links there are on Europeana Collections that may be preventing users from accessing the content they need and damaging their experience. We can assume that the less broken links we have the better experience we can offer for our users.

4 User retention on Europeana Collections

User retention is the ability of a business to retain users. It measures how often users return to the website over the certain period of time. It is both a measure of user loyalty and the capacity of the business to keep their users satisfied by providing a good product and service.

The following aspects impact user retention:

- Product
- Price
- Customer service
- Appealing website layout

In the time period from April to July 2018 Europeana Collections had 1,000,527 unique visitors in total. 133,432 (12.1%) of those were returning users.

In this period the user retention for returning users was:

Day 1 - 25.61%

Day 5 - 4.56%

Day 10 - 2.63%

25.61% of the returning 133,432 users returned to Europeana Collections the day after their initial visit. 4.56% visited Europeana Collections within five days and 2.63% within ten days after their initial visit.

In this period the user retention for new users was:

Day 1 - 0.05%

Day 5 - 0%

Day 10 - 0%

0.05% of the new users returned to Europeana Collections one day after their first visit.

These numbers tell us that over the past four months we did not gain any returning users, “true fans” on Europeana Collections. To verify this we compared the number of returning users we had in April 2018 with the number we had in July 2018.

2018	Total unique users	Returning users %	Returning users
April	283,354	13.4%	70,708
May	273,131	14.1%	73,291
June	243,060	13.8%	63,718
July	236,753	13.2%	58,852

Since April 2018 our percentage of returning users decreased by -0.33%

We also compared April to July 2017 with April to July 2018 to see if we gained returning users over the course of this year.

2017	Total unique users	Returning users %	Returning users
April	195,861	12.8%	50,167
May	255,986	12.6%	64,008
June	231,787	13.2%	61,882
July	255,877	11.1%	56,261

Comparing July 2017 and July 2018 we gained +2.63% returning users.

The table below lists the decrease or increase of returning users in the years between 2014 and 2018 so see developments over time.

Time period compared	Returning users %
July 2017 with July 2018	+2.63%
July 2016 with July 2017	-1.85%
July 2015 with July 2016	-1.62%
July 2014 with July 2015	+15.03%

Looking at the table above its interesting to highlight the steep increase in returning users in 2015 which was due to the launch of the new Europeana Collections interface which included the segmentation by thematic collections. Since then we have seen a decrease in returning users till this year 2018. The increase in returning users this year could be

because we brought in the content from the 1914-1918 stand alone website to Europeana Collections and with that its users.

5 Data access measurements

5.1 Quantitative measurements based on user searches and visited content

Looking at the items, galleries and blog posts that received the highest number of unique visits from our users illustrate which type of data they are most interested in. We can use this information to highlight the most popular data on our home page and thematic collection pages by for example creating top ten lists. Offering new content on a regular basis and highlighting the greatest hits is one approach to giving users a reason to come back to our website.

5.1.1 Top 10 blog posts

The table below lists blog posts visited most by our users in the period between April to July 2018. Most viewed are recent editorials, however a number of older editorials with a specific focus remain popular. Most popular are descriptive blog posts that tell a story about famous historical figures or events.

Blog post	Publication date	Number of visits	Topic
Marco Polo – the man who brought China to Europe	January 2014	2,673	Asia, China, Geography, Italy, Maps
Reading habits in the past	June 2018	2,426	Rise of Literacy, Literature, Manuscripts
World Cup of Art	June 2108	2,338	Art, Competitions
Text Announcement in Manuscripts and Early Printed Books	July 2018	1,782	Rise of Literacy, Literature, Manuscripts
Francis Bacon, Shakespeare and Secret Societies	January 2013	1,671	History, Literature, Science
Did Columbus really see mermaids?	January 2012	1,499	Art, History, Geography

Pola Negri: Woman of the World	April 2018	1,076	Film, Migration
Czechs and Slovaks fighting for independence during World War One	July 2018	1,067	World War I, History
Love across borders: the German student and the English patriot	May 2018	1,059	World War I, Migration, Love
Pride in Europeana: Victorian Dragqueens	June 2016	1,027	LGBT History

Note: The pageview figures in this report show figures of users accessing blogs post in isolation. These figures do not account for viewing the contents of the blog through the overall blog feed, which accounts for 15% of the overall traffic to the Europeana Blog.

5.1.2 Top 10 galleries

The table below lists galleries visited most by our users in the period between April to July 2018. Most viewed are recent editorials, however older editorials remain popular.

Most popular galleries are themed on a specific subject or event. Galleries offer a visual overview of a subject that is digestible easily and quickly. We assume that the reason that galleries organised around a particular subject are most visited because users don't require much cognitive power to gain a surfaced understanding of the featured subject.

Gallery title & URL	Publication date	Pageviews	Topic
Art Nouveau Posters	April 2017	1,330	Art, Art Nouveau
Famous Migrants	December 2017	1,089	Migration, History, People, Photography
1914-1918: Serbia	June 2018	1,015	World War I
Albert Edelfelt	January 2018	972	Art
The Magic Lantern	May 2017	957	Photography
Clothes for the ballroom	April 2017	928	Fashion
Ballet	April 2017	872	Dance, Photography, Fashion
Artists' Self-portraits	April 2017	861	Art

European Landscape and Landmarks	May 2017	737	Art, Photography
Leeuwarden	June 2018	729	Art, Maps & Geography, Photography, European Capital of Culture

5.1.3 Top 10 items accessed

The table below lists the top 10 items and number of times they were viewed on Europeana Collections, during the period of April to July 2018.

It appears that many of the top 10 items accessed lack a thumbnail and good quality data which prompts the question why these are the most accessed items. We hypothesize that this could be due to the spam traffic bots that regularly target Europeana Collections. This is a common issue that many websites struggle with. We will investigate this further if we can validate the initial hypothesis.

Object title & URL	Pageviews	Topic
Portrait of the son of the Catalan singer-songwriter Joan Manuel Serrat, Queco, born in Madrid on May 8, 1969 of his relationship with the Mercedes Doménech model	3,265	Photography, Music
Wibo van der Linde, the 39-year-old television presenter and head of informative programs at the TROS-TV, will marry at the end of October with the 19-year-old Saskia Joustra from Friesland.	1,851	Netherlands, Celebrity
Elżbieta Starostecka i Włodzimierz Korcz z synem Kamilem w domu PAP	1,156	Photography
Short Legends / author, Antonio Osete Osete Pérez, Antonio (1876-1934)	1,081	Photography
Matrimonio di Antonio Cabrini con Consuelo Benzi	1,045	Photography
Alte Frau Balthasar Denner	978	Art
Diego Abatantuono e la moglie Giulia Begnotti seduti accanto ad un bancone da bar	959	Photography

Udslidt H. A. Brendekilde	944	Art
Escultura sedente	890	Art, Archaeology
El balandrito - Cuadro Sorolla Bastida, Joaquín	884	Art

5.1.4 Items shared on social media that received most engagement

Engagement is a great indicator of popularity. Items that a lot of people find interesting usually receive a lot of engagement. Social media engagement measures the public shares, likes and comments, showing how many people interacted with content. It is a common measure of social media performance.

It is interesting to see that social media posts that are performing well are related to current events like #OnThisDay, participatory posts such as voting on the best cover image and posts that offer a new way of looking at a topic (victorian drag queens, heritage and breastfeeding.)

Post & URL	Publication date	Engagement	Topic
August is here and we need a new cover image. Which of these fruits and vegetables would you like to see on your summer plate? 🍅🍆🥒🍎🍏🍌 Vote by liking your favourite.	1/08/2018	1,488	Monthly Cover Image voting
Welcome in June! We've launched new collection - Europeana Manuscripts. You can visit it here http://bit.ly/2InjyNL To celebrate, we'd like you to chose one from these four images for our cover. Vote by liking your favourite.	1/06/2018	1,192	Monthly Cover Image Voting, Europeana Manuscripts
On the Anniversary of the Stonewall riots, the highlight of LGBT+ #PrideMonth 2018, we dive into Wellcome Collection's rich stores of queer heritage and take a look at an album of Victorian-era photographs from the James Gardiner Collection. Who would think this innocent photograph of a beautiful Victorian lady depicts the same person who stood trial in a case that attracted considerable attention from the Victorian-era London public. Read the story	28/06/2018	1,067	Photography, LGBT+ in cultural heritage, Pride

of Stella http://bit.ly/2tHocx1			
Today is #InternationalCatDay and we have a purr-fect gallery for you, featuring paw-some images of cats from the collections of cultural institutions across Europe 🐾🐾 http://bit.ly/2OTkIUH	8/08/2018	930	International Cat Day/Europeana Gallery
And the winner of the #WorldCupofArt is... Denmark 🇩🇰🏆  Summer evening on Skagen's Southern Beach Peder Severin Krøyer Skagens Museum http://bit.ly/2NTn9m1 For more 🎨👉 Europeana Art: http://bit.ly/2NrDIEj For more ⚽👉 Europeana Sport: http://bit.ly/2GObqRM	18/07/2018	891	World Cup of Art, Europeana Art, Europeana Sport
Alphonse Mucha was born #OnThisDay in 1860. Read our blog to find out how he made his artistic breakthrough http://bit.ly/2nRcg9l	24/07/2018	874	Art Nouveau. Art
#OnThisDay in 1878, Eadweard Muybridge took a series of photographs to prove that all four feet of a horse leave the ground when it runs. This study became the basis of motion pictures. Images: A galloping horse and rider, Eadweard Muybridge, Wellcome Collection, CC BY Source: http://bit.ly/2MvtbbM	15/06/2018	771	Photography, Art
Happy #WorldEmojiDay! What's your current mood? Ours: 🎧📖🎵🍷🎬😄 Did you know that Europeana gives access to over 13 M openly licensed images you can play with? Check them out and get creative 😊 http://bit.ly/2zLQ2NV	17/07/2018	678	World Emoji Day, Openly licensed images
Did you know that Europeana.eu gives access to more than 11,5 million openly licensed images 📖 from the collections of	5/07/2018	661	Openly licensed images, Art

cultural institutions across Europe? Explore them all here: http://bit.ly/2Kzlyn9 #OpenAccess #OpenGLAM #OpenContent			
Nowadays, reports of photos of women breastfeeding their babies getting removed from social media with complaints of indecency are commonplace. And there's not a great deal of coverage (pun intended) in our art heritage either. Breastfeeding heritage for #WorldBreastfeedingWeek on our blog: http://bit.ly/2vaQKQ5	5/08/2018	660	World Breastfeeding Week, Breastfeeding Heritage
Happy #InternationalMuseumDay to all museums and cultural institutions contributing to @Europeanaeu! We're really happy to have so many great partners across Europe and beyond! #museumday 🎉👏👏 #EuropeForCulture	18/05/2018	650	#InternationalMuseumDay

5.1.5 Top 10 sources of traffic to Europeana Collections

Below are statistics from Google Analytics on the top sources of traffic to Europeana Collections, including the average number of pages visited per session. From these statistics, we can see that Google is our primary source of traffic (c. 62% of all traffic).

Traffic Source	Number of users (proportion of overall) April - July 2018	Pages per session
Google - Organic Search	678,146 (61.72%)	2.81
Direct Traffic/No Source	206,091(20.65%)	3.81
Referral Traffic	87,050 (8.63%)	5.35
Social	33,495 (3.32%)	4.47
(Other)	2,619 (0.26%)	5.05
Email	1,603 (0.16%)	5.54

Looking at the table above it is interesting to note that the highest numbers of traffic is coming from Google - Organic Search but these people view an average of 2.81 pages per session. Whereas the traffic resulting from links to exhibitions and galleries disseminated via email is much lower, these users visit more pages during their session. Looking at these

results we can conclude that people who have permitted us to contact them via email have a higher level of engagement and therefore are already “true fans” of Europeana Collections or will be easier to convert to “true fans.”

Top Referral Traffic Sources	Number of users (proportion of overall referral traffic) April - July 2018	Pages per session
Noticias.universia.es	8,151 (9.02%)	3.31
Search.creativecommons.org	6,762 (7.48%)	2.97
bibliotecaspublicas.es	5,254 (5.81%)	2.20
blog.europeana.eu	3,745 (4.14%)	8.33

Social media and our own end-user blog are also prominent contributors to our traffic on Europeana Collections. We get the highest level of user engagement from users coming from our own blog and Pinterest. These users stay two to three times longer on Europeana Collections than users coming from other platforms.

Top Social Traffic Sources	Number of users (proportion of overall referral traffic) April - July 2018	Pages per session
Facebook	17,150 (51.21%)	3.26
Pinterest	10,635 (31.75%)	6.17
Twitter	2,497 (7.46%)	4.49

5.2 Quantitative usage of the API

Users also access data via the Europeana REST API. Since April 2018, EF has collected monthly usage statistics for the Europeana REST API. From these statistics we derived two metrics to determine the number of active users: 1) number of API keys that exceeded the average of 5 calls a day; 2) number of API keys that were active for more than 5 days in each month. In the table below you find information of the months 1 April to 31 July 2018.

Period	Number of API keys that...	
	exceeded the average of 5 calls a day	were active for more than 5 days in each month
April	62	64
May	70	76
June	65	62
July	64	61

5.2.1 Content accessed via SPARQL endpoint

The SPARQL endpoint was developed as a pilot service and therefore does not yet track usage. As part of our efforts to bring this service into a full production state, we will start logging all user requests and collect the relevant content statistics for future reports.

5.2.2 Content accessed via OAI-PMH harvesting

The OAI-PMH service is still in an alpha state and therefore does not yet track usage. This will be addressed once this service is moved into beta state by adding logging of all user requests. We will collect the relevant content statistics for future reports.

5.3. Link maintenance activity

As reported in the previous C.3 report (M8) we worked on broken links as part of a bigger analysis of Europeana datasets to improve data according to the Europeana Publishing Framework⁷ (EPF). We identified that about 1,200 datasets (about 13 million records) are not compliant to any of the EPF tiers (e.g. have an insufficient thumbnail size, lack of direct links to digital media) and about 60 datasets (1.2 million records) of those have broken links. The majority of these datasets were de-published since September 2017 and all datasets larger than 5,000 records were free of broken links in April 2018. During a follow up analysis of Europeana datasets in August 2018, a number of datasets larger than 5,000 records have been identified where links must have been broken after we completed the last analysis. We will work with the data partners that submitted the datasets over the next 3 months to fix them or otherwise depublish them after these three months.

The process for detecting broken links in the Europeana Collections galleries is now semi-automated. When a displayed item in a gallery encounters an error in retrievability, the data partner services team gets notified and can take appropriate action.

6 Analysis of the data access measurements

We collected quantitative data from a number of sources that help answer the question “How can we increase the amount of returning users on Europeana Collections?”

Data shows that at this time we have a small percentage of returning users. In the period April to July 2018 the percentage of returning users on Europeana Collections was 11.6%. We also didn't gain any returning users in this period but actually had a -0.33% decrease in returning users since April 2018.

Looking at the top ten blog posts that are most viewed by our users we see that these are all descriptive posts that tell a story about famous historical figures or events. As highlighted in the previous C.3 report, it is the stories that connect the data that our users find most interesting and valuable to read.

⁷ The EPF clarifies the relationship between the quality and openness of data and the impact it can have. Implementing this framework means that EF and all its data partners work towards high content quality so it is compliant with the higher tiers of the framework. In the EPF four tiers of participation were developed, to allow data partners to decide what they want to achieve with their digital collections, based on their own agendas and capabilities. The higher the quality of data provided by data partners, the more benefits they will create for their audiences. View at <https://pro.europeana.eu/post/publishing-framework>

Similarly, the most popular galleries are ones organised around a specific subject or event. We assume it is because galleries tell a story visually so users don't require much cognitive power to gain a surface understanding of the featured subject.

The social media posts that receive the most engagement are related to current events like #OnThisDay, participatory posts such as voting on the best cover image and posts that offer a new way of looking at a topic (victorian drag queens, heritage and breastfeeding.)

The top ten sources of traffic tell us that users who land on Europeana Collections from a google search results page have a lower engagement level than users who follow a link from a newsletter or an email that arrives directly in their mailbox. In practice this means that we should more actively and prominently invite people to join our mailing list and pay attention that the content and calls to action in our newsletters and emails actively convert people to visit Europeana Collections.

Also, traffic coming from our end-user blog and Pinterest have a higher quality of engagement. Reasons for this could include that people subscribed to our blog RSS feed are like the users that subscribed to an email list. These are the "true fans" of Europeana Collections. Or our blog posts include prominent and active calls to action that draw the eyes of the user and lead them back to items on Europeana Collections. Pinterest is a highly visual content website. Everything is catered towards designing an attractive visual presentation of content. Since our other social media accounts have an average engagement whereas on Pinterest the engagement is twice as high we would conclude that Pinterest as a platform closely aligns with the expectation and needs of our users when viewing content from Europeana Collections.

7 Conclusion

There are a few conclusions that we can draw from this research regarding the changes that need to be made in our interface and editorial strategy to increase the amount of returning users on Europeana Collections.

First, we need to offer our users a browsable interface similar to Pinterest since this is the platform they seem to favor.

Second, it is key to highlight the content / data that people are looking for on the Europeana Collections home page and the thematic collection landing pages. These stories are surfaced when connections are made between the data about a subject, person, time period or an event in European history.

Third, we need to actively use our direct communication channels such as newsletters and email lists to guide people back to the Europeana Collections website. These people have entrusted us with their email address and we need to provide them timely and effectively with value by directing them to new and popular cultural heritage content and data made available on Europeana Collections.