![europeana]

Deployment and Maintenance of Europeana DSI core services - SMART 2017/1136

CONTRACT NUMBER - LC - 00822914

DELIVERABLE

C.2 Users and usage report M5

Revision	1.0
Date of submission	31 January 2019
Author(s)	Dasha Moskalenko, Julia Schellenberg, Mónica Marrero, Aleksandra Strzelichowska, Ash Marriott, Antoine Isaac, Hugo Manguinhas
Dissemination Level	Public

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
0.1	10/01/2019	Dasha Moskalenko, Julia Schellenberg, Mónica Marrero, Aleksandra Strzelichowska, Ash Marriott, Antoine Isaac, Hugo Manguinhas	Europeana Foundation	First draft
0.2	25/01/2019	Harry Verwayen, Andy Neale	Europeana Foundation	Review
0.3	30/01/2019	Dasha Moskalenko, Julia Schellenberg, Mónica Marrero, Ash Marriott, Aleksandra Strzelichowska, Antoine Isaac, Hugo Manguinhas	Europeana Foundation	Final version
1.0	31/01/2019	Harry Verwayen, Andy Neale	Europeana Foundation	Final review

The European Commission does not guarantee the accuracy of the information and accepts no responsibility or liability whatsoever with regard to the information in this document. Neither the European Commission, nor any person acting on the European Commission's behalf, is responsible or liable for the accuracy or use of the information in this document.

This deliverable contains original unpublished work except where clearly indicated. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

1. Executive summary

This user and usage report details the main challenges we face on Europeana Collections, our hypotheses for possible improvements and the conclusions of the preliminary research performed during this phase of Europeana DSI-4.

We aim to increase traffic coming to Europeana Collections. A positive trend is that traffic increased with 16% from 2017 to 2018. But the ambition is higher. We therefore need to investigate what is within our span of control to positively influence this trend. Looking at top channels of traffic we received most traffic from organic search (70.6%). We investigated what influence improving SEO has on increased indexing and therewith findability of content via organic search. We could not establish a direct correlation as efforts into SEO didn't increase the number of pages indexed by Google. This informs us that we should continue our experiments with SEO in the upcoming months (e.g. submitting adapted sitemaps to Google). On the other hand, we saw that other channels have the potential to generate traffic to Europeana Collections which are more in our direct control and that can be influenced by our various communication and promotional activities. Top channels of traffic were direct (17.7%), referral (8.3%), and social traffic (2.8%). While we have little control over direct traffic, we see potential to generate traffic via referral sources and social media. We will further invest in diversifying traffic sources through partnerships, social media, and PR activities, in particular looking at the impact of specific campaigns or events.

Secondly, we want to increase the user return rate by improving user satisfaction on Europeana Collections. Search (34.2%) is the most prominent action users perform on Europeana Collections and user feedback has shown that users have difficulties to find the content they are looking for. We made major changes in the search algorithm. From the measures obtained, there wasn't any significant impact in terms of how users interact with the search functionality. We will investigate deeper and experiment further with possible search improvements. Another path we are looking into is offering the item (e.g. newspapers) and item page in the preferred language of users. We are experimenting with applying automatic translation to our exhibitions. After this experiment we can evaluate if this had positive impact on user return rates. Finally, improving the quality of the item page might stimulate the user return rate. In December, we released the new item page design which is expected to lead to a more fulfilling user experience. We performed initial tests with users showing that key information was easily findable. To make educated guesses if the new item page had a positive effect on the user return rate we will track several metrics (bounce rate, average time on page) in the upcoming months.

2. Introduction

Europeana Collections is Europeana's main service for end-users to access content. Europeana Collections with its twelve thematic collections (e.g. Europeana Art, Europeana 1914-1918, Europeana Fashion, Europeana Music) provides access to the full Europeana dataset (ca. 58 million items) via its search functionalities (search box and filters) and via browse entry points, exhibitions, and galleries.

We have several mechanisms in place to get to understand who our users are (e.g. user research on user groups) and to evaluate the usage of the platform, including performance indicators (e.g. number of visitors, retention of visitors). We also look at direct user feedback and undertake usability studies to find out where users are experiencing difficulties with Europeana Collections, with the aim of improving the platform over time.

Overall, the current users seem to receive Europeana Collections positively. A high level measurement that we use to track general user satisfaction with Europeana Collections, the NPS score, showed a very good score of 36[1] on the question: 'How likely are you to recommend Europeana to a friend or colleague?' As for example the evaluation of Europeana shows, users appreciate the diversity of the content on Europeana and its attempt to provide reliable information on for example copyright. Other feedback is more critical. A lot of the more critical feedback pertains to the quality of data (broken links, missing titles, etc). Data quality improvement is a key objective of Europeana and a project on its own, therefore out of scope for this report.

In the following document, we explore what we know about the users of Europeana Collections, and investigate what we can learn from usage patterns, with two specific objectives - we aim to:

1. **increase the traffic to Europeana Collections measured in number of visits, and**
2. **increase the user return rate to Europeana Collections measured in the number of returning visitors as a percentage of the total over time.**

To investigate this we have taken the following hypotheses as the starting point:

Problem statement: traffic
- *Hypothesis 1:* Investing resources in search engine optimization activities results in better indexing of pages by Google and other search engines.

---

[1] Measured in December 2018. A NPS that is positive (e.g. higher than zero) is felt to be good while an NPS of +50 is excellent.

- *Hypothesis 2:* Promotional activities such as campaigns, social media contests, GIF IT UP competition and transcribathon events result in views on Europeana Collections.

Problem statement: user return rate
- *Hypothesis 1*: Improving the search algorithm and thereby the accuracy of the search results on Europeana Collections will increase the user return rate.
- *Hypothesis 2:* Translating the language of the item and item page will result in a higher user return rate.
- *Hypothesis 3:* Improving the quality of our item pages will stimulate the user return rate.

This report states what we have learned so far. We were able to establish some hypotheses as 'correct', which allows us to continue making improvements towards our stated aims. Some of our hypotheses we have concluded to be incorrect, or partially incorrect. In these cases we will be developing new hypotheses that we will investigate the upcoming reports (Users and usage report, M10). In many cases, we feel that more data needs to be collected to establish whether our hypotheses are correct or not. Increasing traffic and increasing retention are complex research questions that we need to investigate further in order to make progress.

3. Data collection

For the evaluation of the usage of Europeana Collections this report draws from data available via Google Analytics[2]. The data looked at, excludes Europeana offices IP addresses based in The Netherlands (filter view) to allow for little deviations as possible (as of use by Europeana staff during maintenance and development).

To look at usage related to search performance we use data from our logging infrastructure. Other search-related insights (such as the additional information on the Entity Collection) come from the Europeana database itself.

We refer to relevant user research activities we have done over the past couple of years. Our tools include: user interviews, surveys, usability tests on low fidelity prototypes, and heatmaps to monitor interactions on Europeana Collections.

We also list the direct feedback received from users. Every page on Europeana Collections offers a user feedback button allowing visitors to directly leave comments on the website.
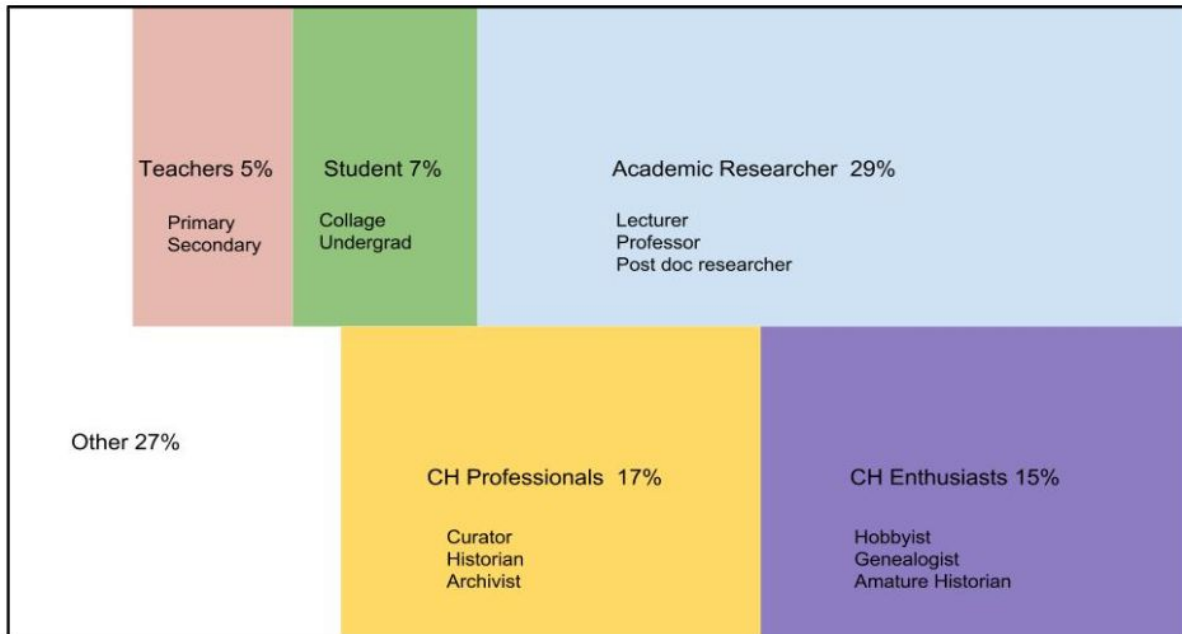
---

4. User groups

We have reviewed and updated[3] our "Personas" for Europeana Collections. We started with an analysis of previous surveys and regrouped users into various categories:

Europeana Collections percentages of user groups gathered from 859 users in 133 surveys from 2015 to 2018.

As part of this review we then interviewed 10 people, each for an hour to understand their goals, frustrations and key tasks. This information combined, with the information we gathered from previous surveys, was recorded in "Persona" cards.

"Personas" are a tool designed to help product teams to empathize with their users and consider their needs when designing the interface and developing the functionalities that determine the user experience of Europeana Collections.

Results
We established that our user base consists of four primary user groups[4]: teachers, researchers, API users and cultural enthusiasts (our assumption is that the needs of CH professionals and students would be met by meeting the needs of the cultural enthusiast).

---

[3] https://docs.google.com/presentation/d/1XcG-zTQo2Krp3Otk_LIx3a9-Vv_fXE1n2vJCrxFwAP0/edit?usp=sharing
[4] https://drive.google.com/open?id=1SCkKLsL6Glv4JpjyBVaP1uXAwcDr994X

An overview of our personas detailing main goals, tasks, and frustrations ect. can be found in the Annex. These attitudes and behaviours have informed the development of the hypotheses in this report.

5. Problem statement: traffic

Europeana Collections (including the Europeana blog and the Transcribathon website) is not reaching the target of 500,000 monthly visits. Currently, we reach approximately 400,000 visits each month on average.
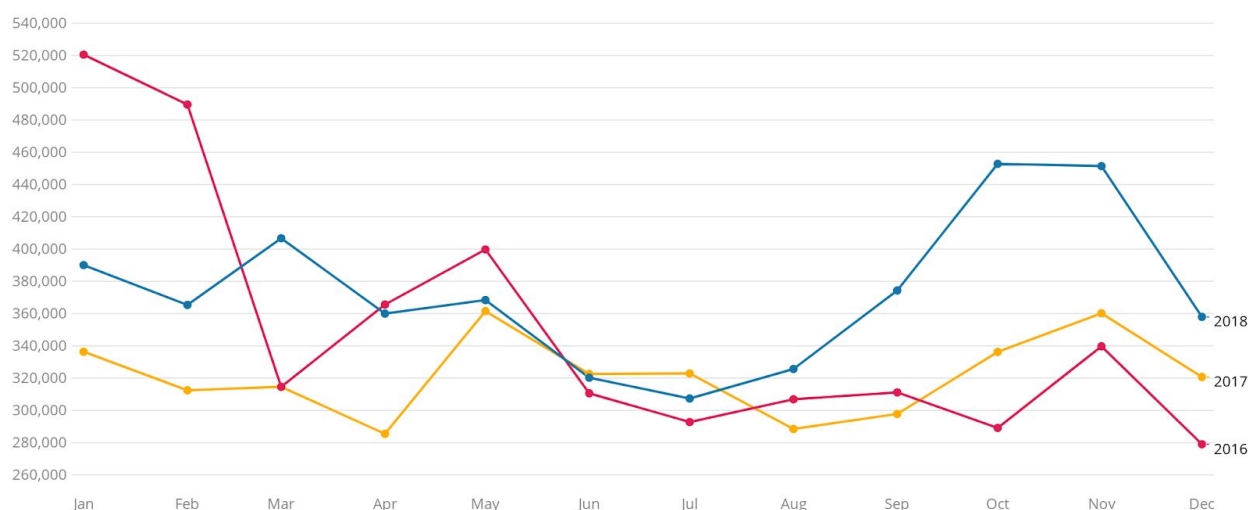
Traffic to Europeana Collections, Europeana blog and the Transcribathon website between 2016 and 2018

A first observation we should make is that while we are not yet reaching the KPI of 500,000 visits, our results are already significantly better than the number of visits we received in 2017 and most of 2016. Overall, we can see that the traffic increased with 16% from 2017 (3,858,018 total traffic) to 2018 (4,479,925 total traffic).

This begs the question what affected this increase. We believe that a large part of the increase in traffic between July and November 2018 can be explained by several activities we undertook that generated additional traffic. These are the GIF IT UP competition, Rise of Literacy and Migration campaigns (as part of the European Year of Cultural Heritage (EYCH)) and Europeana's 10th birthday. The traffic drop in December 2018 happens each year and we assume that the Christmas holiday season causes the drop. In order to understand better what influences traffic we broke the problem down into several hypotheses.

Hypothesis 1: Investing resources in search engine optimization activities results in better indexing of pages by Google and other search engines and increases traffic to Europeana Collections.

With 92%, Google is the dominant search engine on the market worldwide[5]. Millions of people use it daily to discover content on the internet. In the period between September to December 2018 69% of the Europeana traffic came from this organic search with 99% of searches coming from Google. It is therefore critical that search engines like Google, and to a lesser extend other search engines like Bing, see Europeana Collections as a valuable source of unique information. This could result in a higher ranking of Europeana content in the search results list (and hence potentially in higher traffic on Europeana).
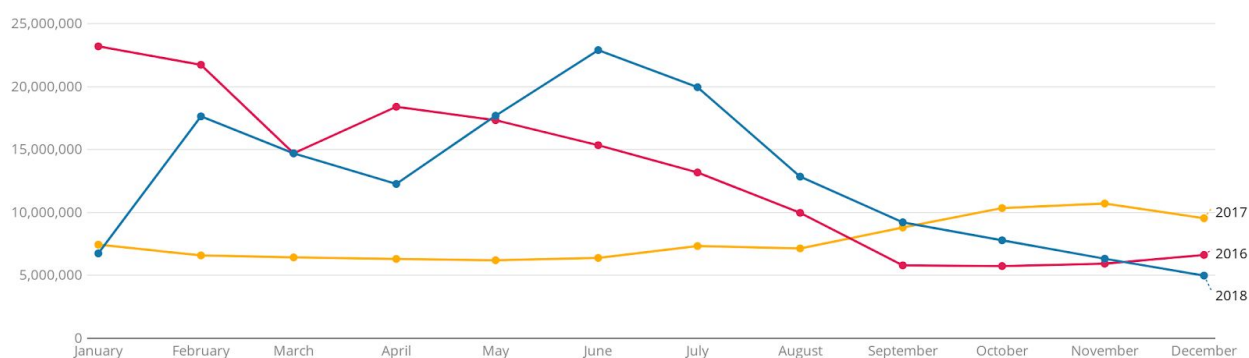
Number of pages indexed by Google for 2016, 2017 and 2018

What we observe is that while indexing by Google went up during April and May 2018 and was higher than 2017 until September, it went down from June. While during that period several efforts were undertaken to improve SEO. Interestingly, the overall traffic actually increased during that same period. Before we reach any conclusions we should investigate specific efforts undertaken to improve SEO.

Technical efforts towards improved indexing
We know that Google looks at the performance of pages such as pagespeed and accessibility when indexing them. Between October and November 2018 we looked at how our pages were performing using a tool called Lighthouse and invested in the following activities:
- Routine compression of editorially-uploaded images, which doubled our pagespeed score on both the homepage and on thematic collections.
- Improvement to accessibility for less-abled persons on the item page, aiming to increase our accessibility ranking. This includes improvements designed to assist users with screen readers.

---

[5] http://gs.statcounter.com/search-engine-market-share

During Europeana DSI-4, we will continue to introduce changes which should contribute to the improvement of SEO as described in the paper '*Current efforts to counter de-indexing of Europeana*'[6]. In the coming months we will seek to conclude work on the following items, which we already started in the past months:

- Build a sitemap for Entity Pages
- Include 'organizations' in the Entity Collection and organization pages in Europeana.eu
- Publish Schema.org metadata alongside our pages, so that search engines can 'read' directly structured metadata associated with our content
- Increase backlinks from big Linked Open Data sources (Wikidata and DBpedia)
- Further investigate the use of metadata quality to favour ranking of higher quality items (a first update apparently did not lead to a conclusive outcome, see other sections)

We will release these changes quarterly to give us the time to judge their impact.

Conclusion

Current numbers on indexing suggest that our hypothesis has to be rejected for now. While we invested in SEO indexing didn't increase while traffic went up. What is possible is that improvements in SEO did increase the perceived value of Europeana material by the (google) search engine, leading to higher ranking but we currently have no data for that. In fact, one of the issues is that Google's ranking algorithm remains a black box. In the upcoming months, we will investigate the accuracy of the hypothesis and what additional data would be necessary to make more accurate conclusions.

Hypothesis 2: Diversifying traffic sources increases visits to Europeana Collections.[7]

We assume that by undertaking various communication and promotional activities we can increase the awareness of Europeana Collections. In the following, we look at the top channels of traffic and what activities had the potential to generate traffic to Europeana Collections.

Traffic to Europeana Collections comes from four primary sources: organic search (via search engines), referral (visits coming via links on other websites), direct (people directly

---

6

https://pro.europeana.eu/files/Europeana_Professional/Governing_Board/Public_Papers_of_Governing_Board_Meetings/Public%20Papers%20of%20Governing%20Board%20Meeting%202018-04-11/Paper%204C%20-%20EF%20Board%20Meeting%2011Apr18%20-%20De-indexing%20of%20Europeana.pdf

[7] Please note that the data gathered for hypothesis two only refers to the Europeana Collections website. It does not include data from the Europeana blog or the Transcribathon website.

typing the url in the browser or using bookmarks), social (from social media platforms), and email (from the newsletters).
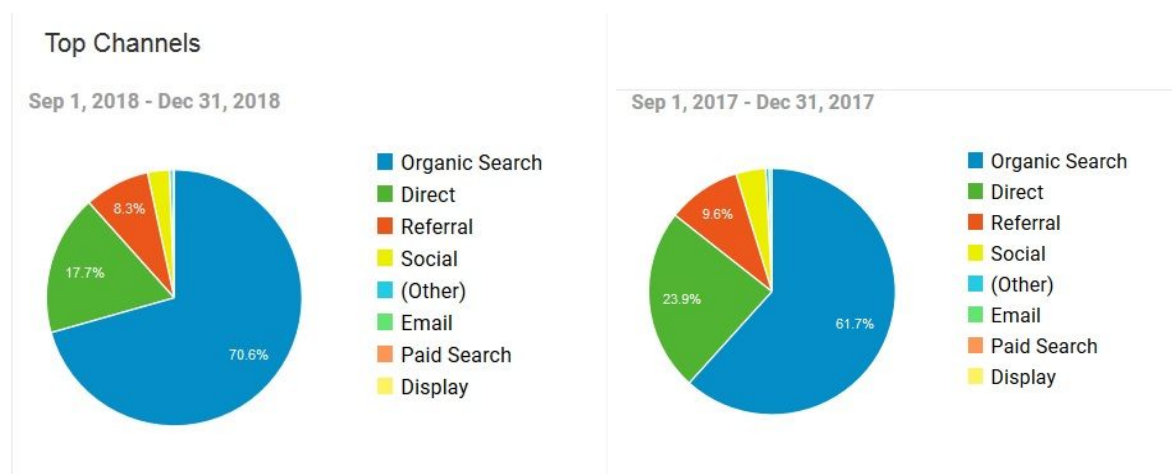
Top channels of traffic in the period between September and December for 2018 (on the left) and 2017 (on the right)

Organic traffic[8]
As we see in the graph above between September and December 2018 most traffic to Europeana Collections with 70.6% (862,374 visits) resulted from organic search. Comparing the same period with 2017 61.7%, (623,015 visits) we can record an increase since last year of 38.4% (239,359 visits) from organic search.

Unfortunately, it is not possible to get further information on the keywords users used to discover the website. Google doesn't provide further information in interest of protecting the privacy of the searcher.[9]

Direct traffic[10]
Europeana Collections also received 17.7% (216,683 visits) direct traffic in this period. Comparing the numbers to the same period in 2017 23.9%, (241,881 visits), there is a decrease in direct traffic of 10.4% (25,198 visits).

---

[8] A visit is considered to come from organic search if a user arrives at our website via a search engine with keywords (unpaid search). Google Analytics automatically recognizes the most popular search engines, and attributes traffic to these sources.
[9] More information: https://analytics.googleblog.com/2011/10/making-search-more-secure-accessing.html
[10] A visit is considered to be direct traffic if a user directly typed an Europeana Collections related URL, or who had bookmarked the site.

Referral traffic[11]

Referral traffic resulted in 8.3% (101,055 visits) of the total traffic in this period. Comparing it to the same period in 2017 this was 9.6% (97,326 visits). We can conclude that no major changes can be observed towards referral traffic.

Source	Sessions ▼↓	Sessions	Contribution to total: Sessions ▼
	129,397 % of Total: 8.81% (1,468,669)	129,397 % of Total: 8.81% (1,468,669)	
1. ■ noticias.universia.es	13,343	10.31%	
2. ■ bibliotecaspublicas.es	7,372	5.70%	
3. ■ blog.europeana.eu	6,492	5.02%	
4. ■ search.creativecommons.org	4,531	3.50%	
5. ■ conted.ox.ac.uk	3,704	2.86%	
6. ■ omnia.ie	3,523	2.72%	
7. ■ noticias.universia.net.mx	2,000	1.55%	
8. ■ en.wikipedia.org	1,680	1.30%	
9. ■ bdh.bne.es	1,570	1.21%	
10. ■ data.collectienederland.nl	1,519	1.17%	

Referral traffic to Europeana Collections in the period between September and December 2018.

Most referral traffic in 2018 with 10.3% (13,343 visits) of the total referral traffic came from a Spanish/Portuguese language university network for Ibero-America[12], a portal for students and universities. There is a link to Europeana Collections on a page about how to download books legally.[13] Comparing to the same period in 2017 this was 4.6% (5,885 visits). There has been a significant increase in visits coming from this referral traffic source with 127% (7,458 visits).

The second highest referral traffic source is the Europeana blog[14] with 5.2% (6,492 visits) of the total referral traffic. Comparing the data to the same period in 2017 this was 5.6% (7,096 visits) of the total referral traffic. We can conclude that no major changes can be observed here.

The creative commons search page[15] is the third highest referral traffic generator in 2018 resulting in 3.5% (4,531 visits) of the total referral traffic. Comparing the data to the same period in 2017 this was 6.7% (8,497 visits). We experienced a decrease of 46.7% (3,966 visits) from this traffic source.

---

[11] A visit is considered to come from referral traffic when a user is referred from another site to Europeana Collections.
[12] http://noticias.universia.es/
[13] http://noticias.universia.es/portada/noticia/2014/02/18/1082572/8-sitios-web-descargar-libros-gratis-forma-legal.html
[14] Europeana blogs are currently maintained under a separate CMS and are therefore considered as referral traffic by Google Analytics - the separation is not visible to the users in their user journey on Europeana Collections. http://blog.europeana.eu/
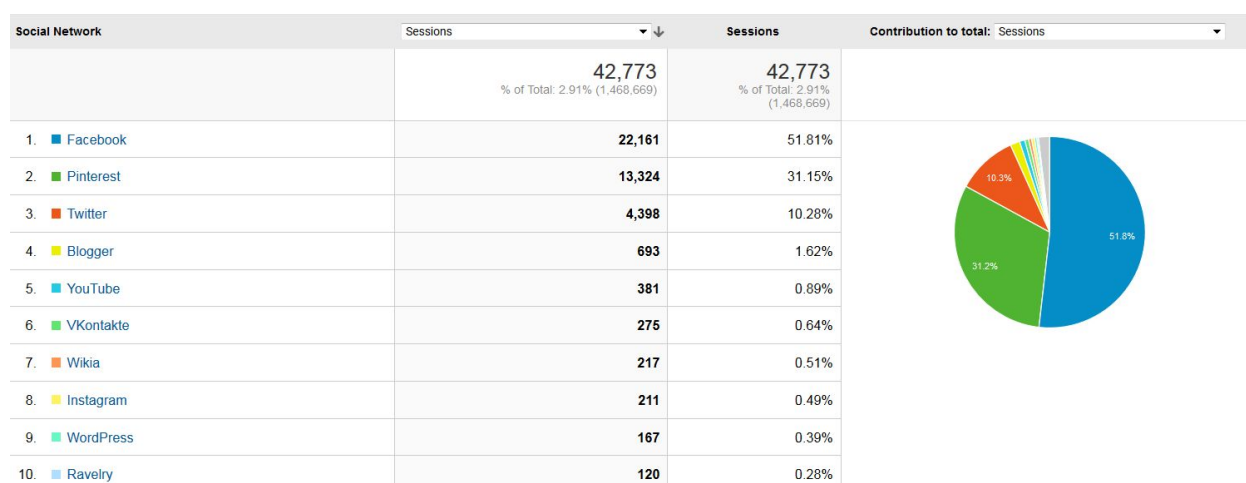[15] https://search.creativecommons.org/

Wikipedia pages are the fourth highest referral traffic generator in 2018 with 1.3% (1,680 visits) of the total referral traffic. Comparing the data to the same period in 2017 this was 1,12% (1,427 visits). Here we experienced an increase in visits within one year of 17.7 % (253 visits).

There are also a few unidentified sources which appear in the top 10 referral traffic list such as bibliotecaspublicas.es. This URL leads to a dead page and does not appear valid.

Social[16]
Social media contributed to 2.8% (33,843 visits) of the total traffic. Comparing the data to the same period in 2017 this was 3.9% (39,458 visits). We experienced a decrease of 14.2% (5,615 visits) from this traffic source.

Social Network	Sessions ▾ ↓	Sessions	Contribution to total: Sessions ▾
	42,773 % of Total: 2.91% (1,468,669)	42,773 % of Total: 2.91% (1,468,669)	
1. ■ Facebook	22,161	51.81%	
2. ■ Pinterest	13,324	31.15%	
3. ■ Twitter	4,398	10.28%	
4. ■ Blogger	693	1.62%	
5. ■ YouTube	381	0.89%	
6. ■ VKontakte	275	0.64%	
7. ■ Wikia	217	0.51%	
8. ■ Instagram	211	0.49%	
9. ■ WordPress	167	0.39%	
10. ■ Ravelry	120	0.28%	

Social traffic to Europeana Collections in the period between September and December 2018.

Most traffic in 2018 from social media to Europeana Collections came from Facebook with 51.8% (22,161 visits). Comparing to the same period in 2017 this was 46.8% (24,531 visits). We experienced a decrease in visits of 9.7% (2,370 visits). While the frequency and quality of our posts remained the same, Facebook's decision[17] to prioritize posts from friends and family over content from 'Pages' might be a reason for the drop. Additionally, multiple data breaches affecting users' details pushed us to remove Facebook pixel from Europeana, placing our visitors privacy and safety over tracking and targeting opportunities.

Pinterest generates the second highest number of traffic from social media in 2018 with 31.2% (13,324 visits). Comparing to the same period in 2017 this was 20.5% (10,754 visits). Here we experienced an increase of 23.9% (2,570 visits). Pins on Pinterest have little description on the item itself which might motivate the user to further explore the item on Europeana Collections. Pinterest users by nature might also be collectors of items

---

[16] Traffic from social media such as Facebook.
[17] https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/

searching for new pins towards their interests, which might motivate them to explore Europeana Collections further.

Twitter is the third highest traffic generator with 10.3% (4,398 visits). Comparing to the same period in 2017 this was 14.6% (7,655 visits). We experienced a decrease of 42.5% (3,257 visits) from this traffic source. A reason for the drop might be the rebranding of the service from Social Media to News Website and the fact that Twitter doubled its character limit, which gives the user a chance to learn more about a piece of content within their feed.

It's worth noting that social media posts in general receive more views on the social media platforms than they contribute visits to Europeana Collections. A social media post about an item usually contains contextual information about the item, leaving users no incentive to click through to Europeana Collections.

Conclusion

While indexing went down we see an increase since last year of 38.4% (239,359 visits) from organic search. This leads us to the point that we need to investigate other hypotheses: it could be that Google sees us as a higher value resource than previously, which results in better ranking of our pages in search results. It could also result from our promotional activities. Or perhaps there was an external factor that contributed to users clicking more on search results leading to Europeana Collections. We would need to investigate further to draw a definite conclusion.

Referral traffic remained fairly stable. Here it's worth highlighting that the partnership we have with the Spanish/Portuguese language university network for Ibero-America where they share a link to Europeana Collections on their resources page has proven to be very successful. It's worth looking into the reasons why this link generates more traffic than links on other third party websites.

There is a decrease in traffic coming from direct and social sources. We need to investigate further what the possible reasons could be. It is worthy to note that Facebook and Pinterest provide the most value for us in term of social traffic to Europeana Collections. Our promotional activities are well received on these platforms.

To conclude, we should continue to investigate further what the success factors are in organic search, and further invest in diversifying traffic sources through partnerships, social media, PR activities. We should also test other possible traffic sources such as paid search and email marketing (until now mainly used for promoting the blog). This would help not only to increase the number of visits, but it could also strengthen our SEO efforts. It is important to investigate why certain partnerships are successful in terms of generating visits to Europeana Collections and how we can utilize the Facebook and Pinterest

platforms more effectively to further increase our click throughs to Europeana Collections. More elaborate conclusions and dependencies will be explored in the upcoming months.

6. Problem statement: user return rate

Europeana Collections has had a stable user return rate fluctuating between 12-14% but it has not increased.

Looking at the data we see a stable user return rate to the website. While over 80% of our users are new users that we gained via search engines and various communication activities only a relatively small percentage comes back. On the following pages, we will explore the possible hypotheses that might influence the user return rate.
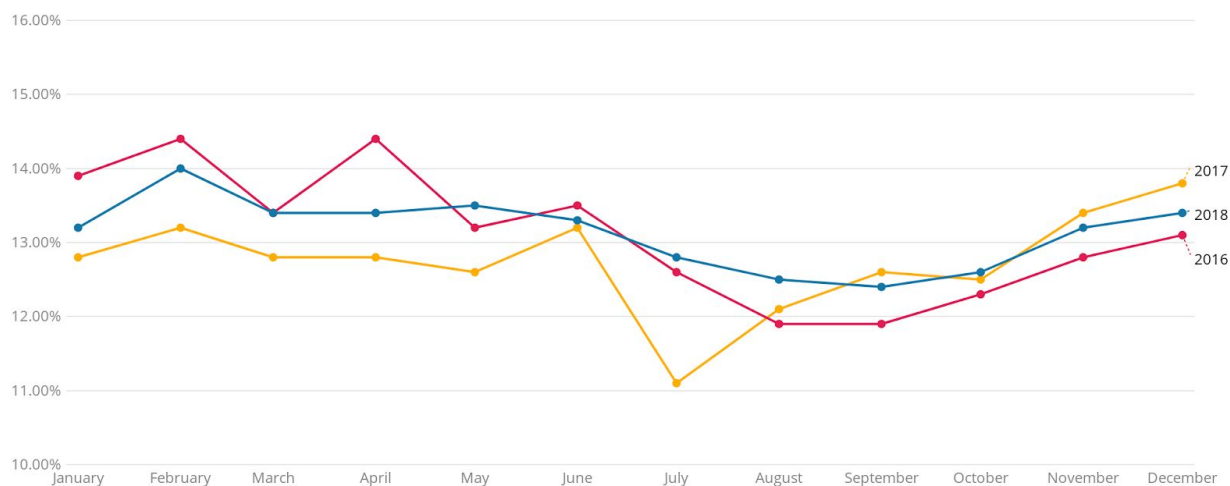
Returning visitors in percentages showing the years 2016, 2017, 2018

Hypothesis 1: Improving the search algorithm and thereby the accuracy of the search results on Europeana Collections will increase the user return rate.

Improvements to the search functionality
During the last quarter, efforts to improve search have focused on the release of a minimal viable product for the newspapers collection, in production since 6th December 2018, with the option of searching the full-text instead of only the metadata. From a search perspective, its release required the creation of a new index in Solr, the indexation of the new contents, and the creation of a new schema and configuration to effectively search in large amounts of text. We have also included new features like the highlighting of the query terms (not yet deployed on the portal) and the faceting of dates by other temporal expressions besides year (by month, by range of days, etc.). This is an important step to support full-text search in our collection, and new enhancements are expected to be done as well as measuring the effectiveness of the search engine in this collection.

We have done this in combination with the update of the search engine to the newest stable version (Solr 6.5), taking advantage of some of the latests features available in the new version to optimize the queries. Additionally, during 2018 a new ranking criteria was included following the recommendations in the Search Improvement Report[18] in Europeana DSI-2, where items retrieved from a search are ordered according to the following criteria (in that order):
1. Presence of digital content (first those with digital contents)
2. Solr score
3. Timestamp (first those more recently updated/added)
4. Completeness of metadata (first those more complete)

Besides, the ranking algorithm used (BM25F) was replaced by the current default ranking algorithm (BM25) in the updated version of the search engine (Solr 6.5), which was released in production the 21st November 2018. The reason for changing the ranking algorithm was the recommendation made to update that algorithm in the Search Improvement Report in DSI-2, and the fact that the previous one was not compatible with the new version of the search engine used. Besides, some of the features included in that algorithm were difficult to adapt to new challenges, as is the case of the search in the full-text of newspapers (in production since 6th December 2018), with different fields and requirements.

---

[18] Europeana DSI-2. D6.3 Search Improvement Report [available at:
https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_DSI-2/Deliverables/d6.3-search-improvement-report.pdf]

Performance and usage pattern metrics

In this report we have measured the performance of Europeana search (see table in the annex) following the criteria already established in previous reports, with the current measures taken after the upgrading of Solr and the introduction of the new ranking algorithm and criteria (21st November 2018). The measures taken since that moment until the 21st of January 2019 are compared with those obtained in the last report where they were calculated, before the previous search specialist left Europeana. The new search specialist, who started on October 2018, focused on the launch of the newspapers collection, and has resumed the evaluation activities although, due to lack of time, a few measures are not reported. However, the search effectiveness is analyzed more in detail and some improvements are suggested. The whole set of metrics are expected to be reported in the following reports, together with the next improvements done in terms of search.

We have slightly changed the methodology to compute nDCG for the general search (as well as the additional metrics calculated, Precision and Reciprocal Rank, in Search Functionality Annex). In previous reports, the data was obtained directly from the logs, taking all queries with clicked results for the evaluation and considering clicked documents as relevant documents. In this case, in order to measure how effective is the system for the user queries, we have took into account only queries with keywords as opposed to queries with only filters, where the user intention may be completely different (get all records that match a condition, no matter in what order).

Component	Evaluation Criteria	Description	Value for the period of January to April 2018	Value for the period of November 2018 to January 2019
General Search	nDCG	Normalised Discounted Cumulative Gain: a measure of how high in the result list clicked items are found. See https://en.wikipedia.org/wiki/Discou nted_cumulative_gain#Nor malized_ DCG. Used as a reporting metric in previous DSIs' Search Improvement Plan Progress Reports.	0.611 (0.56 for queries with query keywords)	0.54 for queries with query keywords.
	Percentage of queries with clicked results	The proportion of queries that are followed by a click on their results, as opposed to queries without clicks (which are then assumed	16.3%	9.45%

		not to have brought relevant results).		
Filters	Frequency of use	Queries using filters	46.58%	61.59%
	Frequency of use	Queries using only filters	46.5%	22.55%
	Frequency of item access after filter applied	Queries using filters with clicked results	8.67%	5.17%
Similar items	Frequency of use	The number of times users have clicked on one of the items suggested as part of the 'similar items' listed for an object being browsed.	0.24%	0.11%
	nDCG	Normalised Discounted Cumulative Gain: a measure of how high the similar item clicked was in the list of all similar items suggested.	0.66	0.53
Auto-complete	nDCG upon completion of autosuggest term by user	nDCG as applied to the entity list supplied by the autocomplete.	0.887 (unchanged until a script is implemented to register users' selection within the list of entities suggested by the auto-complete)	
Entity Collection	Entity Collection coverage	Percentage of searches matchable by entities within the Entity Collection. Used as a reporting metric in previous DSIs' Search Improvement Plan Progress Reports (with a KPI of 30%)	58.7% (based on a small sample of user queries, due to time constraints)	N/A
	Links to Entity Collection	Percentage of the items linked to at least one entity	47.23%	44.98%

Search metrics

Interpretation of new results, improvement measures planned
As a result of the analysis we can see that the search effectiveness is similar to the reported previously, with an important decrease in the number of clicked results. Further analysis should be done to discard that this can be due to an increase of queries coming from our internal services (which do not result in any clicks) or even from external crawlers or automated processes coming from the Europeana Search API.

One puzzling aspect is the apparent lack of effect of the new ranking deployed in summer 2018, while this new ranking was expected to alleviate a lot of users' frustrations. It is possible that these frustrations were in fact not concerning a significant number of queries. We also suspect that we are reaching the limits of the simple evaluation method we have used so far, which assumes that clicks reflect the relevance of search results. We will investigate this point in the coming year.

In terms of the features used, there has been a significant increase in the use of filters, although the types of filters most used remain similar.

Regarding the semantic enrichment, we keep a similar percentage of records linked to entities, which we expect can help users to find relevant records. We plan to continue with the analysis of the impact of the enrichment in the search effectiveness for the coming reports, as well as with the analysis of the auto-complete feature.

Conclusion

We made major changes in the search algorithm and apparently they have not contributed to an increase in the return rate. From the measures obtained, there has not been any significant impact in terms of how users interact with the search functionality. Therefore, either those improvements actually did not change the performance of the search, user interaction issues from the portal are affecting the return rate from the search/browsing pages, or we may be facing issues in the search that we are not able to spot. We will investigate the latter further, revisiting our evaluation methodology.

Hypothesis 2: Translating the language of the item and item page will result in a higher user return rate.

Europeana Collections user interface is translated into 27 European languages that cater to the vast diversity of nationalities and languages available in Europe. To ease the user's journey we translated all the navigation and labelling on the interface and the error messages.

As we receive data from institutions across Europe the metadata describing the items as well as the language of the item itself (e.g. book, newspaper) are in their original language. We understand that this is not ideal for our users as they can only engage with a limited amount of items that are available in the language(s) they can read. It may be disappointing for a user to view an item in a language they don't understand and even discourage them from browsing further or returning to the website.

In collaboration with the eTransation DSI of the European Commission we are currently experimenting with applying automatic translations to particular parts of Europeana Collections. For example, Europeana will use the tool to translate selected exhibitions (e.g.

from the Rise of Literacy and Migrations projects) as a feasibility study to apply automatic translations to exhibitions (and other types of static, user-contributed content) in general.

In 2019 we will also try to run an experiment applying automatic translation to our newspapers collections, which are the main source of (item) content that are available to us for such processing.

Conclusion

Currently it is too early for us to see if translating the language of the item and item page will result in higher user satisfaction and therewith a higher user return rate. We need to further investigate the technical experiment that is planned for the newspaper collection and see if this had a positive effect on user satisfaction and therefore retention.

Hypothesis 3: Improving the quality of our item pages will stimulate the user return rate.

A note on data quality

The key factor that determines the quality of the item page is the quality of the data (media and metadata quality). One of Europeana DSI-4's key objectives is to improve data quality by adding thumbnails, fix broken links and add metadata descriptions to items by enrichment and contextualisation (e.g. adding places, dates, subjects). This does not only provide more context to the item itself, but it also improves the findability of the item both on Europeana Collections and via search engines. In the period between September and December 2018 we received four comments from users via the feedback button reporting issues with the metadata on the item pages such as broken links, and incorrect description. Data quality improvement is a project on its own, therefore out of scope for this evaluation.

We investigated what other factors on the item page impact user satisfaction and therefore user retention.

Item page

Europeana Collections faced a high bounce rate on the item page. In October 2018 it was 68.6%. We assume that improving the quality of our item pages will stimulate the user return rate. In December 2018, we released a new item page with browse options. Support for extended browsing across the site is expected to lead to a more fulfilling experience for users. When discovering more useful material they may be more likely to return if they are getting more from their experience.

All new designs are prototyped and tested with users before implementation. We tested the item page prototype prior to implementation and we tested it again post release to analyze if users can find all the key information.

One click test
In December 2018, we performed a one click test on the item page with existing users that we recruited via our social media channels and the newsletter. Our objective for this research was to find out if our users can easily find all the key information such as, the date it was created, its copyright information, the download button and the name of the data provider on the new item page design.

We recruited 104 people to participate in a remote first click test which examines what a test participant would click on first on the interface in order to complete their intended task.

Results
All of the tasks were completed by the participants within 20 to 35 seconds. The success rates were:

- 38% of the participants clicked on the Download button.
- 21% of the participants clicked on the Created by field.
- 18% of the participants clicked the Place field and 9% on the Location field.
- 28% of the participants clicked the Provided by field and 14% clicked on the Provenance field.

This is a good result since we have to keep in mind that the test participants see this new item page design for the first time. This increases the cognitive load for a participant as it requires them to digest the question and then look through the design to find and click on the field that answers the question.

Heatmap review
In December 2018, we reviewed how our users are interacting with the new item page using a heatmap. A heatmap shows which elements on the page attract the users attention and therefore receive the most clicks. It is important that the key information on the page stands out and therefore is easily finable.
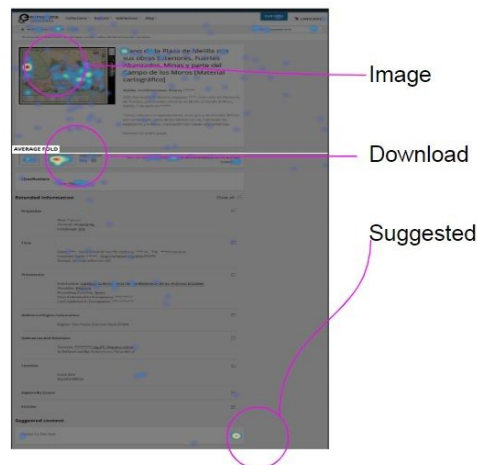
Results
These initial results show that the image, download button, and the suggested items below are easily findable for our users. As these are the primary real estate of the item page this is a positive result.

Items page heatmap

Most clicked in descending order :

Desktop: 435 clicks
Image - 103 clicks
Download- 31 clicks
Suggested - 19 clicks

Mobile: 166 clicks
Cookie- 50
Image- 28
Suggested- 10 tap
Download - 1 tap

Heatmap review of the new item page[19]

Conclusion

In December 2018, we released the new item page. Initial tests show positive results. To make educated guesses if the new item page had a positive effect on the user return rate we will first track if the bounce rate decreased. In October 2018 it was 68.6%. In January 2019, it was 79.5%. At this time the bounce rate on an item page increased by around 10%. This could be due to bugs and performance issues that we faced after the release. We assume that this number will decrease over time. We will also track if the average time on an item page increased. In October 2018, prior to the release of the new item page this was 1 minute 45 seconds, post release in January 2019 this is 1 minute 27 seconds. We can conclude that so far no major changes can be observed here. Additionally, we will track the exit rate on an item page that informs how often users exit the website from an item page. In October 2018, the exist rate on an item page was 53.8%. In January 2018, this is 51%. The exit rate has decreased a little since the release of the new item page on 5 December 2018. It is important to note that these are initial results based on two months. We will need to analyse these metrics and the user return rate in a few months time to draw a conclusive result.

---

[19] Heatmap available here:
https://insights.hotjar.com/h?site=54631&heatmap=3445599&token=e95142d1773762031fc2dde3332218b5&device=desktop&type=click

7. Conclusions

7.1 Traffic

The first challenge we analyzed was the number of monthly views on Europeana Collections. Europeana faces the challenge of not reaching the target of 500,000 monthly visits. Overall, we can see that the number of traffic increased with 16% from 2017 to 2018. On a monthly basis we still see fluctuation in the number of visits comparing to previous years but from 2016 to 2018 there is a gradual and consistent increase.

Hypothesis 1: Investing resources in search engine optimization activities results in better indexing of pages by Google and other search engines and increases traffic to Europeana Collections.
Organic search with 70.6% made up the majority of traffic coming to Europeana Collections with 99% of searches coming from Google. It is commonly assumed that investing resources in search engine optimization (SEO) activities results in better indexing (and higher ranking) of pages by Google and other search engines and an improved user experience for users who are more likely to find content. Our latest SEO actions included for example, increasing pagespeed and accessibility of pages, an attribute we assume will increase the quality of our pages and therewith indexing. Current numbers on indexing suggest that our hypothesis has to be rejected for now. While we invested in SEO, indexing didn't increase. In the upcoming months, we will investigate the accuracy of the hypothesis and what additional data would be necessary to make more accurate conclusions. We will further experiment with submitting adapted sitemaps to Google, publishing Schema.org along our website, and Entity Collections.[20]

Hypothesis 2: Diversifying traffic sources increases visits to Europeana Collections
Other significant sources of traffic to Europeana Collections came from direct traffic (17.7%), referral traffic (8.3%), and social media (2.8%). We assume that our various communication and promotional activities generated traffic to Europeana Collections. With direct traffic being the second most important source of traffic, this indicates a level of awareness for Europeana Collections given users directly search for the site. Referral, as traffic generator was particularly successful with references from universities and libraries as well as platforms with large user bases (e.g. wikimedia). We also generated traffic from social media, particular from Facebook and Pinterest. In the upcoming months, we will further invest in diversifying traffic sources through partnerships, social media, and PR activities, in particular looking at the impact of specific campaigns or events.

---

[20] It is important to note that we are not in control of what Google chooses to index and that all efforts to improve indexing and ranking by Google are based on assumptions.

7.2 User return rate

This report also highlights that Europeana Collections has a stable user return rate fluctuating between 12-14%. While over 80% of our users are new users that we gained via search engines and various communication activities only a relatively small percentage comes back.

Hypothesis 1: Improving the search algorithm and thereby the accuracy of the search results on Europeana Collections will increase the user return rate.
With 34.2%[21] (1,657,740 page views) search is still the most prominent action users take on Europeana Collections. User feedback indicated that users have difficulties to find what they are looking for. We assume that improving the search algorithm and thereby the accuracy of the search results on Europeana Collections will increase user satisfaction and the user return rate. We evaluated quantitative measures of performance and use of the different components involved in the search functionality. We made major changes in the search algorithm. From the measures obtained, there was any significant impact in terms of how users interact with the search functionality. Therefore, either those improvements actually did not change the performance of the search, user interaction issues from the portal are affecting the return rate from the search/browsing pages, or we may be facing issues in the search that we are not able to spot. We will investigate the latter further, revisiting our evaluation methodology.

Note: we expect improved browsability to support increased user satisfaction as well, in particular for the Cultural Enthusiast. But this is out of scope for this report.

Hypothesis 2: Translating the language of the item and item page will result in a higher user return rate.
We receive items from institutions across Europe in many different languages. Users on Europeana Collections can only engage with a limited amount of items in the language they can read. This might discourage users from browsing further or returning to the website. We assume that translating the language of the item page[22] will result in a higher user return rate. Currently, we are experimenting with applying automatic translation to our exhibitions. After this experiment we can evaluate if this had positive impact on user return rates. It is too early for us to see if translating the language of the item will result in higher user satisfaction and therewith a higher user return rate.

---

[21] Search result pages received most page views on Europeana Collections between September and December 2018 e.g. https://www.europeana.eu/portal/en/search?q=&view=grid
[22] Note: currently labels on item pages are automatically translated according to the language preferences of the user. The metadata describing the item is provided by institution and in the language of the providing country.

Hypothesis 3: Improving the quality of our item pages will stimulate the user return rate.

Europeana Collections faces a high bounce rate (68.6%) on the item page. We assume that improving the quality of our item pages will stimulate the user return rate. In December, we released a new item page design with browse options. Support for extended browsing across the site is expected to lead to a more fulfilling experience for users. We performed initial tests with users (one click test and heatmaps) that showed key information was easily findable. To make educated guesses if the new item page had a positive effect on the user return rate we will track several metrics (bounce rate, average time on page) in the upcoming months.

Annex: Europeana Personas

The teacher

Jenny 46 yrs — **Teacher**

Goals
What changes do we want to help the user make?

My goal is to motivate my young students to learn by creating connections with real life

Tasks
What tasks are carried out by the user?
- I use the search feature over thematic collections
- I use similar items feature to find other items
- I make sure the items I use are free to re-use
- I read the description of the items

About Jenny
What does the user believe in?
- I am a secondary school teacher of physics, chemistry, mathematics
- I am very passionate about teaching my student
- I make sure my lessons are fun and interesting for the students
- I believe students are the future

Frustations
Whats keeping the user from achieving their goal?
- I don't know if it's me who can't find it or if it doesn't exist
- I copy paste the URL in a word document to find it later
- I find the inconsistent structure of metadata to be troublesome
- I find unclear information on licensing to hinder my activity

Context of use
What does the user need from us?
- I require high resolution public domain content
- I need to have Science Thematic Collection
- I need to know the information on the rights of use
- I have found out about Europeana through social media posts

The Researcher

Lisa 37 yrs — **Post doctorate**

Goals
What changes do we want to help the user make?

Become a published author by finding and reusing specific items for my research

Tasks
What tasks are carried out by the user?
- I use the search feature over thematic collections
- I download and zoom in to images
- I use filters to refine my search
- I read the metadata and use images
- I use art, maps, manuscripts collections

About Lisa
What does the user believe in?
- I am writing a research paper on the specific topic she is passionate about. With a lot of care, she chooses the content for her research paper.
- In my free times she enjoys reading antique books online
- Sometimes I feel researching alone is a lonely job.

Frustations
Whats keeping the user from achieving their goal?
- I am skeptical about the availability of the metadata
- I get inaccurate search results due to not having appropriate filters like date range
- I come across broken links so can't decide the information to be accurate and up to date
- The quality of metadata needs to be better

Context of Use
What does the user need from us?
- I need to be able to cite the records correctly
- More access to items licensed for reuse
- I need to know the provider, creator of the item
- I found out about Europeana from friends or colleagues
- I visit Europeana once/month

The Culture enthusiasts

Sam 32 yrs	Blogger

Goals
What changes do we want to help the user make?

To gain a wider perspective on the cultural heritage of Europe

Tasks
What tasks are carried out by the user?

- I use the grid and list view
- I copy the link to paste on social media such as Twitter
- I watch videos and search for items based on location
- I visit art, photography, music

About Adam
What does the user believe in?

- I look into Europeana to find an inspiration for my work
- I like to gain followers by sharing Europeana content on social media
- I want to gain a deeper understanding of European cultural heritage

Frustations
Whats keeping the user from achieving their goal?

- I feel neutral about the trustworthiness of items
- To clear filters, I have to go back a page to start a search
- The search results are either too long or too little

Context of use
What does the user need from us?

- I require openly licensed items for reuse
- I search for specific contents
- I found out about Europeana via social media posts
- I visit Europeana once/month

The API users

Sam 35 yrs	Developer

Goals
What changes do we want to help the user make?

Become a creator of a meaningful app by using the records and APIs.

Tasks
What tasks are carried out by the user?

- Read the documentation
- Compare the results against collections
- Find ways to get help from the forum, customer support
- Use the console, examples, and requesting API key

About Sam
What does the user believe in?

- I like the challenge of creating something new. I work on my project side by side my full-time job
- I heavily depend on the documentation to guide me
- My API calls Europeana several times a day
- I use the Search and Record API

Frustations
Whats keeping the user from achieving their goal?

- I feel the documentation is unclear and less detail-oriented
- I feel not having enough examples in the documentation makes things difficult
- I find continuously scrolling the long documentation tiresome
- No range search (eg 1800-1850) is possible for the historical time

Context of use
What does the user need from us?

- A simplified version of API would be useful, as the metadata are rather lengthy
- Avoid redundant depth of API (e.g. many arrays have only one data in [0])
- Order data in terms of guide, title, date to be understandable

Annex: Qualitative user feedback

Between 1 September and 31 December 2018 we received 27 pieces of external feedback. They are categorised as the following:

Type	Number received
Bug report	5 (all were resolved)
Report of a provider's data issue	9 (5 were reported to the provider)
Expressing personal connection to content	2
Positive feedback	4
Negative feedback	1
Providing general feedback	6

Action taken on feedback
Below details some relevant feedback which we took action on, it was not always possible to take action on feedback due to a lack of contact information or a lack of detail. Unless in quotes, the below feedback is paraphrased.

Bug reports	Status
Paragraph breaks are no longer working in Europeana Collections	Fixed by the Metis team and currently in testing
Exhibitions link on Europeana Collections broken for German language	Repaired by Collections team and available
Download button refers user to provider site	Passed to the data quality team to contact partner to correct issue

Positive Feedback	Response/Status
'An nice image of winter days.' - https://www.europeana.eu/portal/en/record/2021672/resource_document_mauritshuis_785.html	We contacted the user back, thanked them for their engagement and promoted some other art they might like.

Negative Feedback	Response/Status
'horrible page, everytime i want to look at a photo i'm taken back to the start page and have to go through entire stack of pics; site is too time consuming to be relevant!'	(This regards an item page before we moved to the new one.) We responded to the user asking if they could provide more information and presenting a willingness to help, but they did not respond.
'For high resolution images, it is actually quite difficult to know how to view them in high resolution. But the display of these images seems to be limited to about 2000 pixels wide. Moreover, for records which have more than one image, it is difficult to click from one image to the next - there should be "next" and "previous" buttons, as was the case in the last Europeana portal.'	We responded to the user regarding these issues. The high resolution images issue is something we are re-considering the UX on, with no current commitment to make any changes as it is not deemed to be a priority issue. The next and previous buttons we currently do not have the intention to implement as our research has shown no other negative reaction with these.
'The new item record page displays a small preview of the content, there is no zoom function or access to larger versions. The technical metadata is not available any more, nor is there a recommendation made on how to cite the content - the "share button" simply offers the https reference - previously the name of the organisation making the content available & the rights licence were included.'	We contacted the user back and left some explanation. The issue they were facing around zoom was due to a lack of processing of the media, which will be corrected when Metis has processed the record. The technical metadata was removed because we do not deem it to provide value to the majority of users, rather it was unfriendly on the page. Regarding the citation of content, this is something considered for our next item page iteration, and improved bibliographic referencing is a related requirement of the Europeana Media Generic Services project.
'The Sketchfab viewer loads rather slowly, the first impression given is that there is a simple image available for this content. It is normal for pages to give an indication (e.g. a timer) to show when content is loading. https://www.europeana.eu/portal/en/record/2048705/object_HA_2148.html'	This issue is not on the Europeana Collections side, and we informed the user that this is possible a temporary issue with Sketchfab.

General Feedback	Response/Status
'I can not see the big picture' https://www.europeana.eu/portal/en/record/9200579/trt58nq3.html?q=paediatric+palm+massage+tuina	We contacted the user to explain to them that they need to have a modern browser in order to view IIIF content, in case that was the issue. We also provided direct links to the larger image file if that was the issue. Development work has been carried out to enable the downloadability of IIIF content which may also have solved this user's issue.
'Great site, but limited to only 10 pages.'	We contacted the user back, and asked if they were referring to the maximum retrievable items of 960 from a search query.
'My search brings up 1,995 results, but will only show me the first 10 pages of them, 10 x 96 = 960 images. this happens both in list and in grid format. Also, scores of the images aren't showing; just the icon for image shows.'	Two issues are here. One is that the user is unhappy with the maximum number of results that we are able to return from our search engine. The other issue is an issue that the provider hasn't provided images. We contacted the user and informed them of these limitations.
'It looks very good! But in old version, which no longer seems to be available "under maintenance", you could zoom into the photo much more than in this new version.'	We contacted the user back and acknowledged that there is a difference in the behaviour of the page, which is a decision we made as we feel it improved the UX. We provided details on how they can obtain the highest quality image.
'The IIIF items on Europeana all seem to be viewable using "leaflet". However, this means that the images cannot be downloaded from Europeana. Is this a temporary measure? Are there plans for Europeana to use some other IIIF viewer that can enable download?'	We contacted the user back to inform them that we have taken active measures to improve this, as IIIF media is now downloadable where available.
'hi, there is only one document on this page. i would like to access the other ones. Also il would be glad to ba in touch with Pierre Lerat about these documents/ https://www.europeana.eu/portal/en/record/2020601/contributions_13274.html?q=who%3A%28Pierre+LERAT%29"	We informed the user that unfortunately we do not have information about this contributor to 1914-1918 and would recommend trying to look them up themselves in their would like to get in touch about these records.
'Do you have advanced research box? Sorry I can't find it.'	We emailed the user to explain how to perform advanced searches on Europeana through our search tips page.

'I would like to see this specific poretract of Theodor Pallady. I am receiving funding to visit the museum at which it is currently housed for this specific purpose . Please confirm that it is still in that museum's collection. Many thanks, *redacted name*'	We emailed the user to advise that we don't have record of the availability of artefacts within the museum and that they are advised to contact the museum directly.
'does this item belong in the Manuscripts section?' https://www.europeana.eu/portal/en/record/2048087/ProvidedCHO_Jersey_Heritage_A_0006195_.html?q=stone+axe	We contacted the user to thank them for their input, and then we corrected the manuscript collection to exclude these irrelevant items.

| Provider Issue	Response/Status
Find Out More Link broken; new permalink: http://sammlungenonline.albertina.at/?query=Inventarnummer=[1837]&showtype=record	We investigated and deemed it to only be an issue of one record in the dataset, and therefore a provider issue. We informed the user appropriately and provided them details on how to report this to the provider.
Translated from Dutch: "Street with children; Street in Dokkum, with children Léon & Lévy. Street in Dokkum with Children. Is incorrect. The 'Description' is correct."	This issue was perceived to be inaccurate at source, and we contacted this user back to inform them how they can request that the data be corrected by the provider.
'I am not seeing any image in https://www.europeana.eu/portal/en/record/9200579'	The provider provided a blank image, and we referred the user to the source to make contact and find out what had happened to the record.
'For some reason, the audiovisual material from BNF doesn't play on Europeana - even though the audio works, as in this example: https://gallica.bnf.fr/ark:/12148/bpt6k13111655/f1.audio It would be great if this bug could be fixed as there is some fantastic material here!'	We found that the MIME type of the media provided by the source is incorrect, and we created a backlog item with the data quality team to contact them to correct this. We contacted the user to let them know we took action.

Annex: Search Functionality

Additional information: general search

In order to measure the effectiveness of the search engine, we keep using nDCG, but this time using the standard implementation of TREC_EVAL[23], tool used in the TREC conferences and in general in the scientific Information Retrieval context. We have also included two new metrics used in that context, Precision at 10 (P@10) and Reciprocal Rank (RR), two common metrics also used in the scientific community that allow us to measure other aspects of retrieval effectiveness: proportion of the relevant results in the first (10) documents displayed to the user, and (inverse of the) rank of the first document displayed that is relevant. We have compared the previous data reported and available (First trimester of 2018) with the new period:

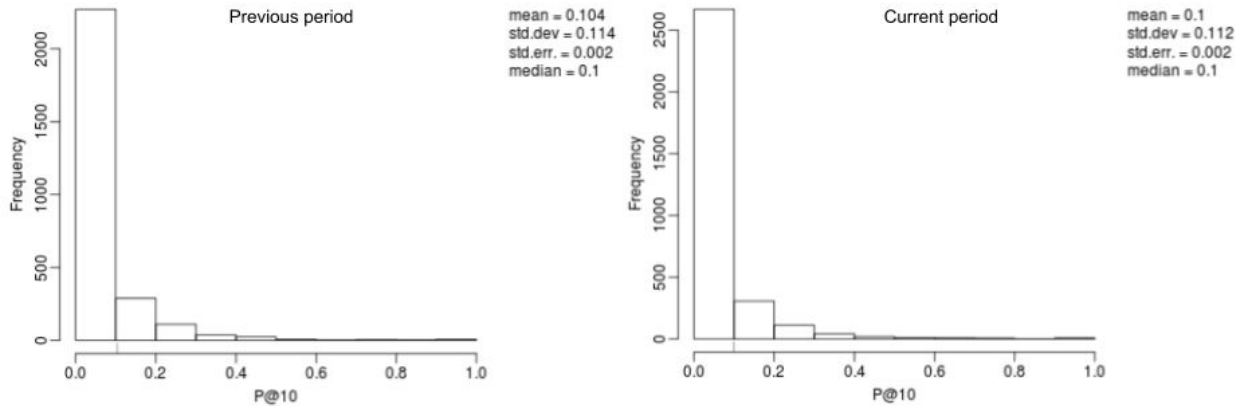|	Value for the period of January to April 2018	Value for the period of November 2018 to January 2019
nDCG	0.56 ∓ 0.012	0.54 ∓ 0.01
P@10	0.10 ∓ 0.004	0.10 ∓ 0.004
RR	0.45 ∓ 0.016	0.43 ∓ 0.014

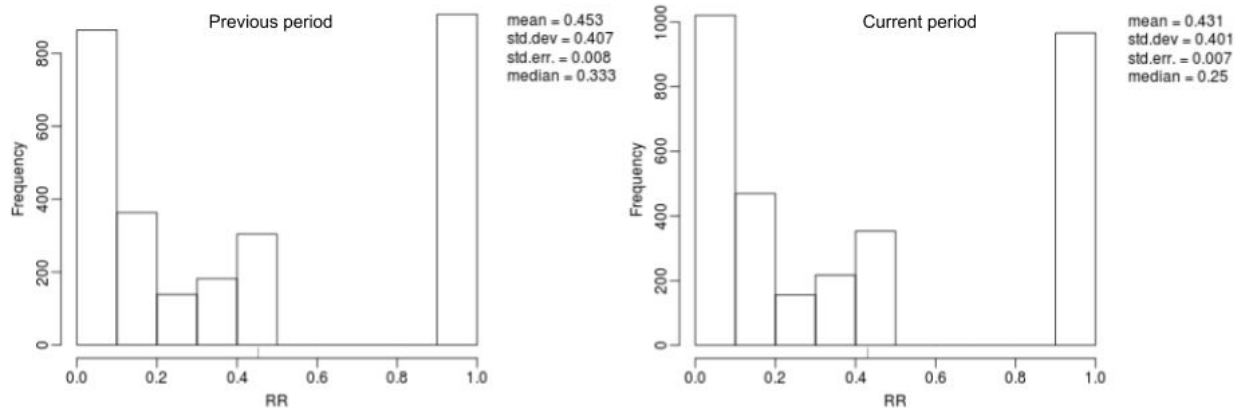*Comparison of effectiveness metrics for the previous and current period*
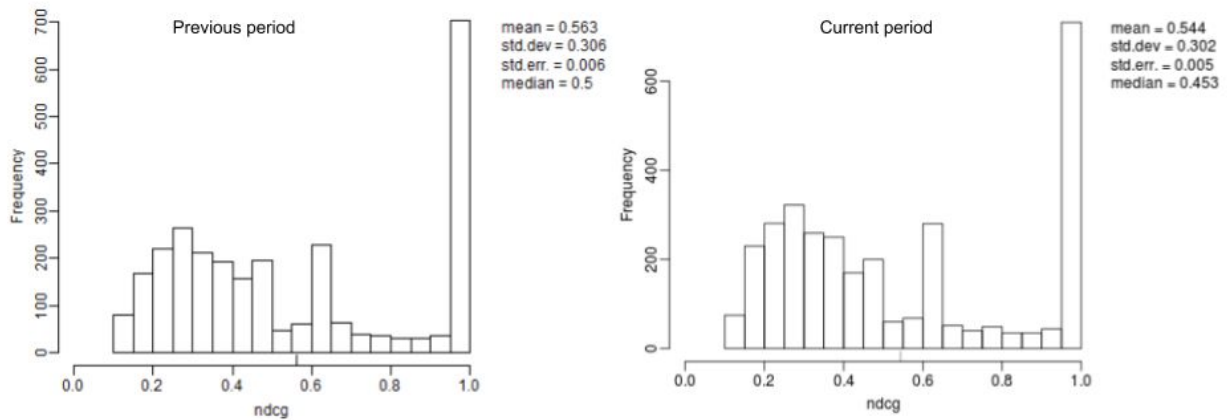
Comparison of nDCG distribution for the previous (left) and current period (right)

---

[23] https://github.com/usnistgov/trec_eval

Comparison of Precision at 10 distributions for the previous (left) and current period (right)

Comparison of Reciprocal Rank distributions for the previous (left) and current period (right)

From the table and figures above we can see that the results are very similar statistically, with just a slight decrease in effectiveness for the current period. Additionally the distributions of Reciprocal Rank and Precision show that the effectiveness is very variable, with queries where users click only once in the first page of results, usually in the first or second position, and queries where the clicks start after the first 10 results.

Additional information: filters

The use of filters can help users to formulate queries. In the table below we can see the type of filters most used. Most of them are filters that are displayed by default for any query in the left side of the search page. The type of the material is the most used filter, probably because its visibility is very high: it can be selected before the query is launched, as the options are displayed in the very search button.

Filter	Previous	Current
TYPE/MIME_TYPE	33.54%	45.05%

DATA_PROVIDER	13.05%	11.23%
PROVIDER	4.72%	6.84%
REUSABILITY/RIGHTS	7.63%	6.97%
MEDIA (URL to the original media file)	8.33%	6.42%
LANGUAGE	4.06%	4.54%
COUNTRY	4.73%	4.28%
COLOURPALETTE	8.69%	3.74%
THUMBNAIL	3.45%	2.81%

Type of predefined filters in the portal most used in the queries

Additional information: Entity Collection

We can see in the table below the the number of records in our collection that contain links to specific entities from the Entity Collection, for the previous and the current period analyzed. We can see that the amount of links in our collection is similar respect to the previous period, with more than 40% of our records enriched.

Entity type	Previous	Current
Agent	2.72%	2.5%
Place	26.59%	24.1%
Concept	25.8%	26.95%
Any	47.23%	44.98%

Records with links to entities in our Entity Collection