**Europeana Foundation
Governing Board Meeting**

11 April 2018
16h30 – 18h00 CET
Teleconference

Current efforts to counter de-indexing of Europeana
Proposed action: For information and discussion

Web traffic to Europeana's end user services (including the Europeana Collections portal, europeana.eu) has been lower than hoped: 3,858,018 sessions logged, compared to a target of 6M in our Business Plan 2017. One of the main reasons is that there have been issues with the visibility of Europeana on web search engines in the past years as our traffic comes mainly from search engines (at least 65% came from organic search in March 2018).

Hypothesis
Our hypothesis is that Europeana, as a publisher of web content, is not search engine friendly for the following reasons:

- As an aggregator, almost all content we publish has already been published elsewhere (on providers' sites)
- Content and metadata quality issues result in some of our pages being not always attractive to users, *e.g.*, some pages show sparse metadata. We also have a lot of (near-)duplication across our pages (when the metadata is very similar, *e.g.* same titles).

We therefore believe that we can gain a lot by enhancing our data and making it work better for SEO (Search Engine Optimization). Our long-term endeavour with regards to SEO has been geared towards offsetting the negative aspects of being an aggregator of data of varied quality levels. We need to differentiate our content better from the one that is published at our partners' site. This strategy aligns well with other long-term data-related efforts that have deployed: enhancing data quality at source, data enrichment by employing Linked Data, connecting objects across collections, *etc.*

Solution areas
In the past months, we have tried to develop a more systematic approach to identify solutions. We have surveyed our past SEO actions, looked at traditional SEO categories. We then inventorized all possible options and prioritized them considering our resources and the product developments already planned (see slides here). We also consulted with Governing Board member Lora Aroyo (see minutes of meeting here).

We have identified the following SEO areas as potentially relevant for Europeana:
- data-based on-site SEO: SEO actions focusing on better exploiting data at europeana.eu - typically, publishing richer metadata

- other on-site SEO: other technical and editorial SEO actions at europeana.eu - for example, enhancing page load speed
- data-based off-site SEO: SEO actions focusing on better exploiting our data outside of europeana.eu - typically, having Europeana data being used (and that use reflected) in other sites
- other off-site SEO: other marketing and editorial actions outside of europeana.eu. For example promoting Europeana content on social media

In addition, some 'meta-level' areas could be improved, in order to guide and prioritize the various options in each of the work areas above:
- evaluation of SEO actions - to more precisely evaluate which actions work well for our case
- logging on Europeana.eu - to get more insight on the concrete content and tasks that are of most interest to our users, and eventually place more relevant material in their path

Preliminary results
In the past months, our first data-related actions have already brought benefits suggesting that our hypotheses are correct. As reported in DSI-3 C.2 (Usage pattern report), we have:
- enhanced the data we publish, thanks to the Entity Collection. Concretely, we provide new, richer pages (for the entities themselves) and we better connect existing pages (for the objects related to these entities),
- filtered the pages that we promote to search engines via our sitemap. We have removed those that have too low a score against a basic metadata quality measure.

Soon after the deployment of these improvements in September 2017, we could observe an improvement of the indexing of our pages - from 7 million to 17 million pages indexed by Google. In the meantime the traffic to Europeana.eu, as shown in the diagram below, has also moderately increased - in the last 6 months only December was below 320,000 sessions per month, a level that had not been since May 2016 (with one exception).
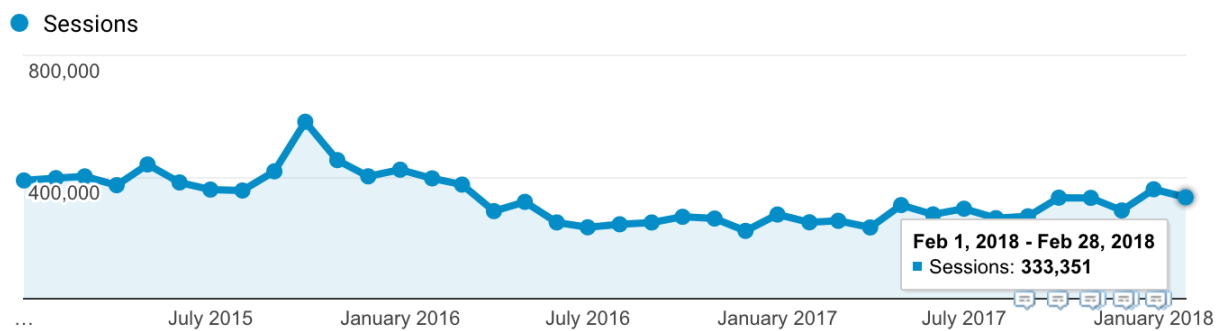
Considering that we met 64% of our traffic target in 2017 and that in the first two months of 2018 we are at 75% of target, we are making good progress. However, traffic is still substantially lower than the 500,000/month that we are striving for. The work on the Entity Collection has just begun and we can pull more data and enrich our objects at a larger scale. We are also only at the beginning of rationalizing our sitemap and better track the effect of these efforts. We have noted that while more pages from our sitemap are indexed now, they still represent a minority of the entire set. Our sitemap contains 10.6 million pages but only 3.9 million are to be found in the total 17 million pages indexed by Google overall. This means that even though Google considers the pages we submit in our sitemap, most pages indexed by Google were found by Google itself (e.g. by following links on already indexed pages), without much guidance on what we consider to be our better content. This raises the issue on whether we can generalize a strategy for further guiding search engines to Europeana's content, including sitemaps but also considering internal links on Europeana

2

Collections' object pages and browse entry points, as well as references provided on other Europeana platforms (Europeana Pro), social media, and the websites of our partners.

Actions
The actions we are going to take in short term are [with the corresponding SEO area]:
1. Include entities from the Entity Collection in the sitemap (they are not included yet), preferably leveraging identifiers from our data.europeana.eu Linked Open Data service. This is a straightforward extension of recent developments of the API and Collections portal - to be ready in Q2 2018 [data-based on-site SEO]
2. Get more backlinks from big Linked Open Data sources (Wikidata and DBpedia). We already have outwards links that result from our metadata enrichment efforts - to be ready in Q2 2018 [data-based off-site SEO]
3. Include 'organizations' in the Entity Collection and organization pages in Europeana.eu - to be ready in Q3 2018 [data-based on-site SEO]
4. Publish Schema.org metadata alongside our pages, so that search engines can 'read' directly structured metadata associated with our content. This is based on earlier R&D work, and is scheduled for implementation - to be ready in Q3 2018 [data-based on-site SEO]
5. Investigate the tuning of search results to reflect the metadata quality. A lot of our pages show results from our own search engine, which currently does not take metadata quality into account. Being able to rank lower bad-quality objects is currently a longer-term target for us, but it has the potential to raise the satisfaction level with our content, for users and search engines alike - Q4 2018 [data-based on-site SEO]

In addition to these initiatives, we are going to continue enhancing interlinking between our spaces, notably Europeana Pro and Europeana Collections: attributions of image used in Pro now results in direct links to Europeana Collections. We will also to continue marketing and communication activities with SEO impact, such as producing high quality curated content on a regular basis (and doing keyword research related to it), and guest blogging and PR activities to get mentioned and linked on influential websites.

Finally, we are going to formalize a team across Europeana (1) to make sure we can seize all opportunities for synergy across the activities mentioned earlier (2) to raise the profile of the SEO issue across teams and at all levels (3) to prioritize other longer-term actions, such as the ones that have been discussed in a recent meeting with Lora Aroyo.

Circulation: Europeana Foundation Governing Board Members, Observers and DCHE

Classification: Public