

Europeana 2012-2013: usage and performance update

CIBER Research Ltd, May 2013

*Growth, change
and impact:
mobile devices,
stickiness,
loyalty, social
media, virtual
exhibitions and
methodology*

Key Findings and recommendations

Methodology: for the first time CIBER have been able to evaluate Europeana usage by all the available quantitative methodologies: deep log analysis, ClickStreamer logs and Google Analytics. In fact we believe this is the first time the methodologies have been compared in regard to usage of a single website.

The ‘ClickStreamer’ log series were used for the first time by CIBER and provided for the period June–December 2012. These include long-awaited cookie data, enabling a much more effective analysis of user engagement and improved definition of core user groups. Google Analytics (GA) data were also used from September 2012 to April 2013 and provide independent verification of the ClickStreamer data set.

We were requested to explore the use of Google Analytics for this project in order to evaluate whether its popularity is matched by its capabilities. It proved an interesting, valuable, but also frustrating experience. The level of access provided did not always enable full use of advanced features (flow analysis in particular). Nevertheless, GA proves a very useful usage tool, albeit one which sometimes underestimates usage, and also one which needs careful calibration and interpretation to obtain full benefits.

Using multiple sources of data also has a downside as it highlights differences and divergences which need to be resolved. Considerable effort has gone into analysing out the resulting confusion caused. If you only have one clock you either trust the time it tells, compensating for known errors, or do without. If you have two clocks that tell different times, you cannot trust either: you know less not more.

Analyses: the analysis provided has a two dimensional matrix. One dimension is derived from the original requirements, to look at six aspects of visits: stickiness, behaviour, loyalty, sharing, location, and flow. The second dimension is a classification of sources of visits: direct/referral, search engine, social media and API.

Overall growth: there has been a steady and impressive growth in the volume of page hits on Europeana.eu since the autumn of 2011. At the close of 2012, the analysis of the various data sources available since October 2010 suggests that projected forward into 2013 annual on-trend growth will be about 25%.

Mobile: use of mobile data and devices is growing rapidly and more rapidly than desktop use—a fivefold increase during 2012—but it has not reached the significant levels forecast in 2011. Tablets dominate the ‘mobile’ space, and most are iPad, but it is probable that a good number are tethered in home or office rather than ‘on the go’. Evidence from usage over the New Year holiday period (2012/13) tends to confirm that tablets are associated with personal rather than corporate use.

Social Media: taking the 'social segment' the overall year-on-year visitor growth is 34%, compared to an overall visitor growth for European of 90%. Exclude blogs and visitor growth falls to 25%. Looking at blogs alone the visitor growth rate is 58%. The social element is a little more significant on the exhibitions site and predictably more significant for blog.europeana. Social media use is a complex area which is bedevilled by problems of identification, definition, novelty and interpretation. Given the importance accredited to it in Europeana planning circles, and the passions typically associated with it, we recommend a detailed investigation to discover why it has driven relatively low volumes of traffic towards Europeana (around 1% of all traffic), why usage is not growing relatively speaking, whether it is generating more 'quality' traffic from users with a greater propensity to share and what significance can be read into use of Europeana data 'offshore', on sites like Facebook. There is a greater propensity for social media to share, but the activity itself is very uncommon.

Loyalty, stickiness and Europeana interface changes: no significant impact could be attributed to the introduction of a new portal in October 2011, an event now consigned to history. Moreover, there have since been several revisions, so a detailed analysis would not be apposite to a home-page that was changed yet again in March 2013. Stickiness and loyalty levels are lower than found elsewhere, say, in scholarly sites but that might be expected of a search engine (or catalogue) that boasts little of its own content. The loyal users Europeana has are the cultural institutions and their members. It is estimated that Europeana's core audience, defined as those people visiting five times or more, is about one-tenth the size of its visitor numbers – about half a million people. Regular users tend to be routine users.

Engagement: most visits are over in the blink of an eye (10 seconds), with just one page viewed. This is probably what you would expect of a discovery site rather than a destination one. The trend appears to be towards a less engaged user, but this needs further investigation as it might be due to other factors.

Virtual exhibitions: an undoubted and a qualified success, which seem highly fit for purpose, popular, sticky, and generate high levels of engagement. They are the elephant in the room.

Click-throughs: surely the most important metric of them all, the gold standard metric, which points to real need and value and of huge significance to partner organisations. Analysing click-through by collection, aggregator etc. shows that click-through rises in line with object views in all segments. It does not matter if a collection is large or small, nor the provider, nor the location: if there are more object views click-through will rise in proportion. Therefore: a) improvements to the Europeana site that encouraged click-through would most likely be effective across the board (e.g. ensuring all objects have thumbnails rather than placeholders); b) from the point of view of an individual provider, improving click-through follows from greater object views (e.g. if the object is more popular it will have more click-throughs). We know from CIBER's earlier work that multimedia content attracts more attention than text, but if you factor that out are Europeana users brand aware? Do they prefer content from particular providers? We don't yet know and clearly should do.

Future analysis: As unofficial curators of Europeana logs and having studied usage in depth now for more than three years (and seen people and ideas come and go) CIBER are well positioned to point to gaps in the analysis of Europeana. What are badly missing are national and institutional impact analyses. Europeana's ambition is to get closer to its stakeholders (and their needs) and such studies would contribute significantly to that.

Table of Contents

1.0 Background	7
2.0 Methodological note and discussion	8
3.0 Results	13
3.1 Mobiles.....	14
3.2 Stickiness and loyalty (in the light of portal changes)	16
3.2.1 Returning Visitors	19
3.2.2 Engagement.....	23
3.3 Information seeking patterns	25
3.3.1 User activity by landing page.....	26
3.3.2 Types of journey:	29
3.4 Social media	30
3.4.1 Definitions.....	30
3.4.2 Size and growth (October 2012 – January 2013).....	33
3.4.2.1 Size and growth update (October 2012 – March 2013)	33
3.4.3 Individual social media	35
3.4.4. Country analysis.....	37
3.4.5 Social actions and social media.....	37
3.4.6 Social media and stickiness.....	39
3.4.7 Loyalty, stickiness and quality of social media traffic.....	40
3.5 Virtual Exhibitions	41
3.6 Collections and Click-Through	42
3.7 Digital visibility.....	47
3.8 API	49
3.9 Overview and update.....	49

List of Figures

Figure 1. Taking the pulse of a website using logs	9
Figure 2. Europeana page views 2009-12	12
Figure 3. Europeana Googlebot.....	13
Figure 4. Europeana Audience Overview Jan-Dec, 2012	17
Figure 5. Visits: August 2011- January 2013 compared to same time previous year	18
Figure 6. Europeana Daily Visits 2010–2013	19
Figure 7. Europeana: return visits	20
Figure 8. Europeana: months between visits	21
Figure 9. Hours between visits	22
Figure 10. Days between visits.....	22
Figure 11. Europeana: duration of visit.....	24
Figure 12. Analytics: Social Flow	36
Figure 13. Social Segment Visitors 2010–2013	40
Figure 14. Exhibitions: visits	41
Figure 15. Europeana Click-through.....	46
Figure 16. Europeana Click-through by dataset size 2012.....	47
Figure 17. Europeana: Indexed documents	48
Figure 18. Europeana: active documents	49
Figure 19. Europeana_Pageview_trend.....	51

List of Tables

Table 1. GA cookies, from ClickStream data.....	10
Table 2. Page views by tablet and other mobile devices: country analysis. June–December 2012, from Clickstream data.....	15
Table 3. Page views by tablet and other mobile devices, key pages	15
Table 4. Pageviews by Tablet and other mobile devices. June–December 2012, from Clickstream data.....	16
Table 5. Duration of Visits, December 2012.....	23
Table 6. Pageviews June–December 2012, from clickstream data	25
Table 7. Page Flow All Users Europeana.eu, from clickstream data	27
Table 8. Summary of overall traffic flow around the site based on ClickStream log data June–December 2012.....	28
Table 9. Summary of overall traffic flow around the site as percentage of visits.	28
Table 10. Internal (within-site) traffic flow June–December 2012	29
Table 11. Social segment definition	30
Table 12. Social segment blogs.....	31
Table 13. Growth rates 2011/ 2012 v 2012 /2013.....	33
Table 14. Social sharing, GA data.....	38
Table 15. Social sharing, GA data.....	39
Table 16. Social sharing, GA segmentation	39
Table 17. Exhibitions, Europeana.eu.....	42
Table 18. Exhibitions, visitors 30 Dec 2012 – 29 Jan 2013	42
Table 19. Click-Through by Dataset, Top performers from 785 datasets, Jun-Dec 2012.....	44
Table 20. Click-through by collection country.....	45
Table 21. Overview based on Google Analytics data for most recent six months (16 September 2012– 16 March 2013).	50

1.0 Background

CIBER have been analysing usage of Europeana since 2009 and have now amassed a three-year long-series of data to evaluate Europeana's growth, changes and innovations. As a consequence we have assembled a large evidence base showing how a whole range of people use cultural collections and artefacts, in a virtual environment. Thus logging is the basis of insight and prediction about the purpose and motive of the millions who use Europeana.

Much work was undertaken during the Europeana Connect development project, 2009–2011, and has been widely communicated¹. Following on from Europeana Connect CIBER were commissioned to undertake a selection of follow-up analyses looking at Europeana after the 'important' portal revision of October 2011:

1. Mobile usage: a 'Culture on the Go' update at +12 months.
2. (Increased) stickiness and loyalty (return visits and engagement) in the light of platform change
3. Information seeking patterns and flow
4. Social media usage (the growing of which has since become a Europeana priority)
5. Virtual Exhibition usage

CIBER were also asked to examine whether Google Analytics could help deliver some of these analyses on a routine *and* sound basis and, as a consequence, methodological evaluation has formed a significant undercurrent to this report and has added a level of complexity to the evaluations and discussions. In particular there are outstanding reconciliation problems that arise when comparing Google Analytics data and the full log data (http-access and Clickstream), which loom large in a future in which a growing reliance is placed on GA for evaluation and impact purposes.

Every exploratory data analysis of usage data presents new problems: an incentive to both revise and devise technique, and the opportunity to discover more. The work described here is no exception. The meeting of the End-user Research Workgroup in January 2013 prompted some new research avenues aligned to its theme of 'Benchmarking Success'. We have thus obtained greater insight into the usage of collections and patterns of referral. Therefore, we have added a sixth element to the research, which covers collections, click-through and digital visibility.

2. http://www.europeanaconnect.eu/documents/D3.1.3_eConnect_LogAnalysisReport_v1.0.pdf;
http://www.europeanaconnect.eu/documents/D3.1.3_Annex_Culture_on_the_go.pdf; Information on the go: A case study of Europeana mobile user <http://dx.doi.org/10.1002/asi.22838>

We have also responded to comments on earlier drafts so that our analysis now has a two dimensional matrix. One dimension is derived from the original requirements, to look at six aspects of visits: stickiness, loyalty, information behaviour, sharing, location, and flow. The second dimension is a classification of sources of visits: direct/referral, search engine, social media and API. All the additional comments and requests for clarification have now been incorporated into the main text, to provide one, single report.

2.0 Methodological note and discussion

For the Europeana Connect project we relied upon server http request logs using CIBER's own 'deep log' methods. However, for this project we were charged to see to what extent Google Analytics (GA) might enable us to undertake the analyses requested; in other words to audit and evaluate GA against our deep log techniques and clickstream logs. This is important given the fact that Europeana, like many organisations, are relying increasingly heavily on GA for all their usage and marketing needs. We have undertaken this but have found GA cannot always supply the data required in a convenient form and have thus supplemented it with our own tried and trusted methods. There is great potential to make better use of GA but it requires considerable investment and effort, not only to interpret the output but in experimental design, preparation and configuration of event tracking code.

The series of analyses presented here is thus a hybrid and experiment, we have made some exploration of the potential of GA but more work will be required to gain significant benefit from it. In particular much closer collaborative working is necessary. For effective advanced analytics today, whatever the method of tracking and logging, an experimental protocol must be designed in advance and suitable tracking code inserted into web-pages before data is gathered. In other words you need to prep the web page to maximise data quality.

Also, that there have been changes in the form of the server logs during the progress of this project. A new series of 'ClickStreamer' logs promises to speed-up and simplify future analyses. A separate series in this format also offers coverage of the Europeana API from January 2012, an aspect of the Europeana project that has not been covered by our previous work. However, we only had access to the ClickStreamer series of Portal logs from June to December 2012 (the minimal time-scale necessary for a robust analysis given the seasonal/diverse nature of usage data). As a result we have used the old series of raw http-request logs for a broader overview and perspective; however, delivery of this old series to CIBER ceased in December 2012 and delivery of ClickStream logs ceased on the 31 December 2012. Our connection to the 'big fat information pipe' has always been fragile, and despite numerous requests and promises, is now unfortunately broken.

So, to provide the best and most comprehensive analysis of usage we have used a variety of data sources. And it is worth pointing out their various strengths and weaknesses. There are, in essence, three points at which we can take the pulse of a website. On receipt of a request by the server; by tapping into the internal process of the site's content management system

(CMS); by causing the browser to send an acknowledgement when content is received. The first of these, monitoring incoming traffic, has been used since the web's inception. It relies on http server request log files originally intended for server management and software maintenance. Not being intended for market research purposes, means the record is not always in the most convenient form. On the other hand it may hold information that would not otherwise be collected because it did not seem relevant at the relevant at the time.

Figure 1 outlines the web-server process and the points at which usage can be measured. For a very simple website with no CMS the URL requested (e.g. a link in the clickstream) maps more or less directly to a web-page file, which is despatched by the server back to the client (browser). In this case the traditional server log is in effect also the CMS log. But today, CMS is the norm and the request no longer maps direct to a file but is interpreted by the CMS. As a result records are retrieved from a database and a web page constructed on demand. The cost of this flexibility and complexity is that the incoming request is no longer a straightforward and reliable indication of what was served in response. Interpretation of request logs becomes a matter of 'reverse engineering' the programming of the CMS. In such cases logging from within the CMS becomes attractive. For some purposes this is obvious and inherent to the application area: an online shop for example will almost certainly be linked to stock control, and accounting records. These can be considered specialised varieties of 'log file'; they can be used for analysis in similar ways to the server log. Or, a specific form of log may be kept for market research and data mining. For any special logging the problem is to specify in advance what needs recording.



Figure 1. Taking the pulse of a website using logs

The difficulty of web-server based logging, wherever the monitoring point, is that it does not record what happens at the user end. A web-page is served but there is no record of its receipt. The solution is to insert scripting into the web-page so that on receipt a secondary request is despatched to report back to a logging system. This is the method employed by Google Analytics and others similar solutions such as the open source Piwik. This can, like CMS logging, resolve the 'reverse engineering' problem, but the task of deciding what to track and of deploying the necessary web tracker 'events' to best effect remains. It also needs to be noted that this approach depends on the end-user accepting and not deleting the tracking cookies and scripts. Our research suggests that for this reason significant traffic, perhaps 10–15%, may be untracked by such browser based methods. This could make a big impact on some analyses, especially those regarding relatively lowly used activities and behaviours.

Taking measurements at various stages of what should, in principle, be a single transaction, raises the problem of reconciling the various accounts. Even if the numbers do not agree we should be able to account for differences. The agreement between http-access log and ClickStream is acceptable: over the period June to November 2012 the http-access log shows a page view count higher by 1%. However agreement between either of these sources and Google Analytics is much harder to establish.

Google Analytics depends on JavaScript being active on the client browser and the acceptance of the Google cookies. Without JavaScript the logging data will not be recorded. Without the cookies it is not possible to identify returning visitors, nor gather reliable information about the sequence and timing of page views. Based on the six-month ClickStream series (Table 1), between 15-30% of visits have a Google cookie set when requesting the landing page, this implies a previous visit to Europeana and retained cookies. For visits comprising more than a single page, the GA cookies are present in 85-90% of page views, thus we think it is highly probable that the remaining 10-15% (possibly one in six visits) have blocked cookies and possibly JavaScript and would not therefore be tracked by Google Analytics.

Table 1. GA cookies, from ClickStream data

Google cookie present	Jun-Dec 2012
On referral from Google Search	32%
On referral from external site (not Google)	24%
On referral with no referrer data	20%
after first page viewed	88%

Unfortunately this estimate of 10-15% untracked visits by Google Analytics does not account for the massive gap between the page views reported by GA and those from the Europeana logs which is 26% in the period June-December 2012. In only one month (September) is the difference (16%) low enough to be plausibly attributable to user blocking of GA. For June the figure is 54%: some further explanation is required.

In the period January-May 2011 (Figure 2) a much greater mismatch of page view counts between Google Analytics and the http-access log was observed: the uncorrected figure exceeding 250%. In that case we introduced the concept of an outlier (used in our previous reports): a series of page requests from a single IP address, often over many days, far too numerous to be the efforts of a single user. Thus the 'visitor' displays all the characteristics of an automated agent or robot bar the user-agent identifier. It could be a cloaked robot. Significantly, such cases tend to go unrecorded by Google Analytics as automated agents retrieve web content but do not run JavaScript. In the early months of 2011 identifying less than a dozen such agents was sufficient to bring the logs and GA into near-enough agreement. A similar process can be applied to the 2012 ClickStream series. For example in August 2012 8.2% of all page-views originated from a single IP address located in Beijing. China has a large population, they may have a considerable interest in European culture, the single IP address could be a proxy for many individual users; on the other hand such heavy and sustained use does not display the irregular pattern of use expected of normal users. If an outlier correction is applied then the difference between GA and the ClickStream data can be coerced into an acceptable error band.

Figure 2 also shows the availability of the API Clickstream series since January 2012 and the Portal Clickstream from June 2012 before correction. The plot of http-access logs shows adjusted values up to September 2011 (our final presentation to Europeana Connect). There after the figures are raw and uncorrected. The information to be gleaned from the API series is limited and is not sufficient to make any inference regarding the divergence of activity levels between API and http-access logs. It may just be a coincidence that the API plot and HTTP plot intersect in January and diverge thereafter. It does however raise the suspicion that the API route to content may be neglected thus leading to significant outlier or cloaked robot activity appearing in the Clickstream series.

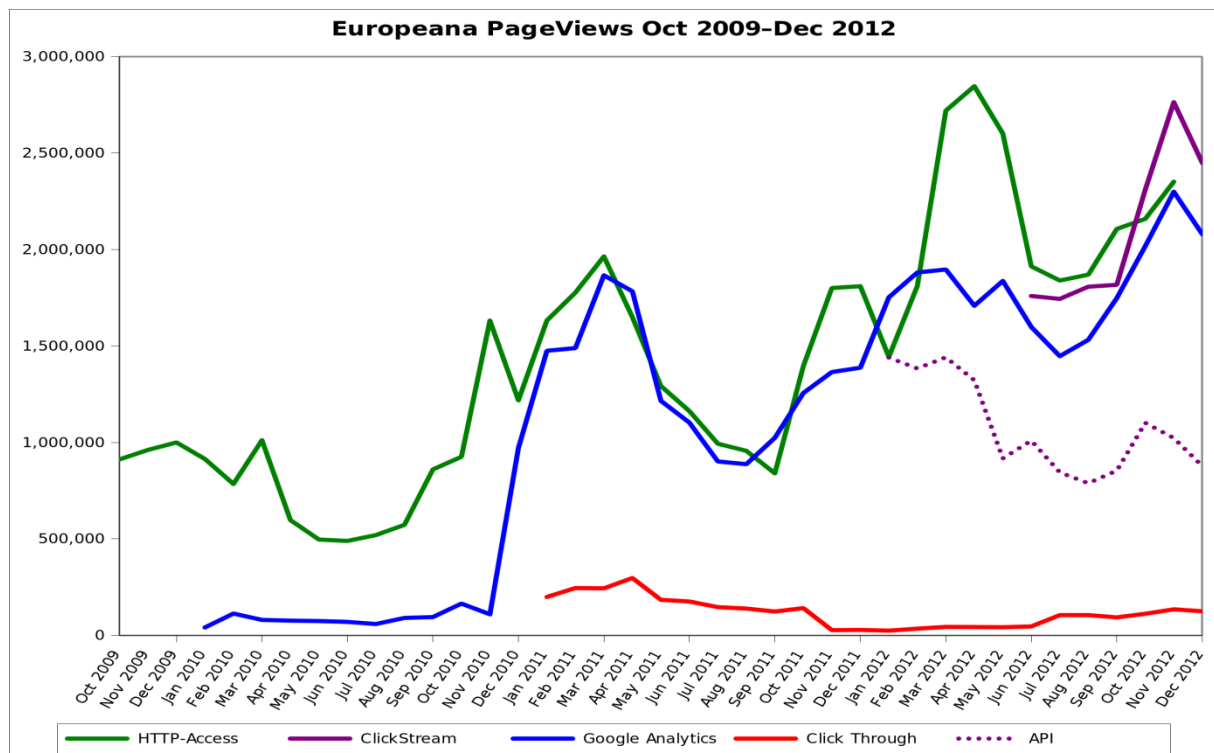


Figure 2. Europeana page views 2009-12

Finally we need to note that:

A 'Click-through' —when a Europeana user requests more detail from a record in Europeana and is redirected to the provider site— will be recorded in Europeana logs but, following usual web-analytics convention, does not count as a hit on the Europeana site. It will therefore not be counted by Google Analytics etc.

- In April-May 2012 the indexing of the site by Googlebot appears to have broken down. It is not clear what impact, this had on the site in the months that followed. But it is probable that this may have exaggerated the seasonal decline we observe during the summer months (compare Figure 3 below and Figure 6).
- Tracing journeys is a complex task, neither logs or Google Analytics provide all the data we need to this reliably. We do not appear to have sufficient access to GA to use the flow graphing feature effectively, and the ClickStream series of logs, though an improvement on the old http-access series still present difficulties when attempting to identify individual sessions.

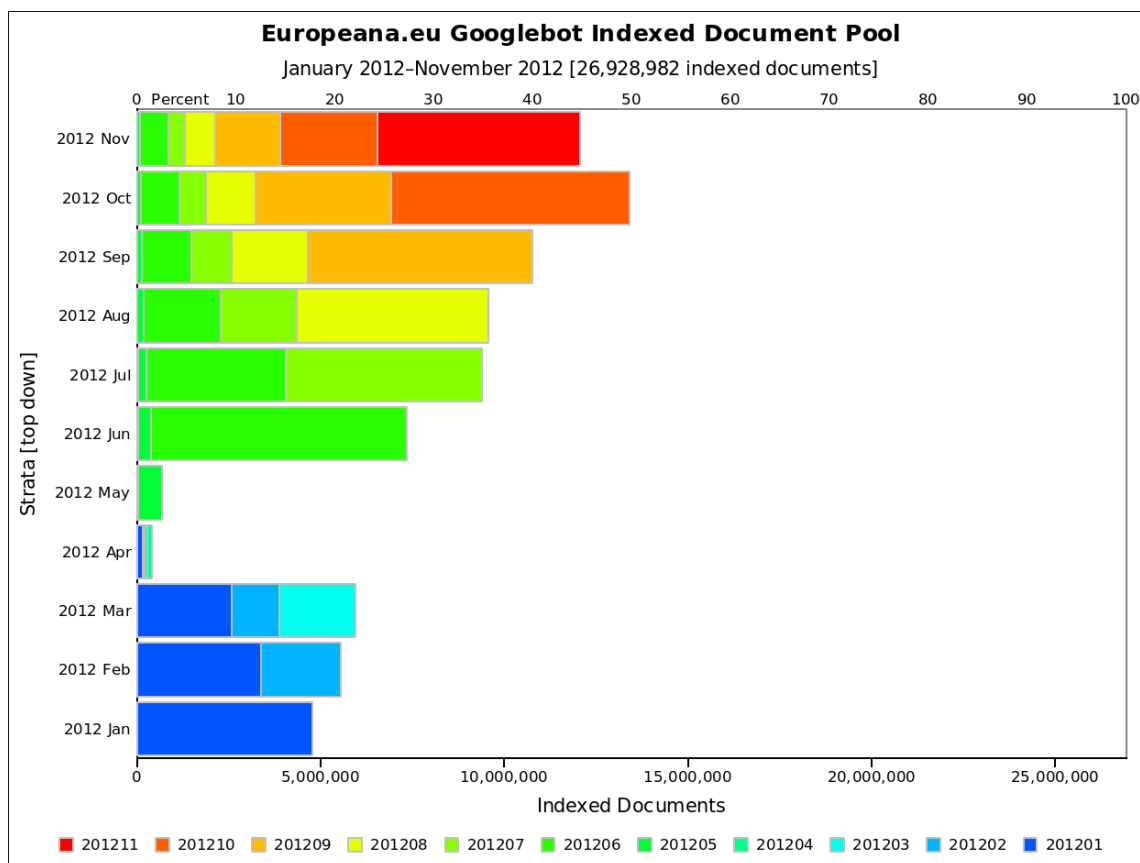


Figure 3. Europeana Googlebot

3.0 Results

This report is a compendium of reports and analysis requested early in 2012; we have revised our coverage of 'Mobiles', 'Engagement' and the related 'stickiness' and loyalty characteristics; we have added new material on social media sharing and virtual exhibitions. As mentioned earlier we were asked to explore how much analysis was possible using Google Analytics data rather than our tried and tested 'deep log' analysis of server log files. There have been many delays in obtaining data and in establishing a correlation between series and sources. Effective access to GA was only achieved in September 2012. The project aims, logistics and methodological considerations have led us to make some observations based upon Google Analytics data. These must be treated with caution until we are satisfied with the correlation with both the long-series of data we have obtained from three years accumulation of http-access logs and the short (six-month) series of ClickStreamer logs. The latter is especially important as it includes the Google Analytics cookies that admit both comparison with GA results and better tracking of single-user multi-page interaction.

In sum: Google Analytics' reliance on cookies and scripting is effective in suppressing the effect of cloaked robots and other automated agents that would distort the profile of a normal sentient user; but the same feature will also miss genuine users who have blocked cookies and JavaScript.

3.1 Mobiles

Shortly after the release of the first draft of this report we were advised that the mobile device usage recorded by Google Analytics may be inaccurate; in particular the usage of smartphones (but not tablets) may have been under-recorded. We have therefore reworked the section placing greater reliance on both the http-access logs and the ClickStream dataset.

CIBER's previous report, 'Culture on the Go', showed rapidly growing use of smartphones and tablets (4 times faster than desktops). At that time (October 2011) our forecast was for Europeana to have a total 3.3 million visitors in 2011, rising to 5.5 million in 2012. The outcome was in fact 3.7 million in 2011 and 5.1 million in 2012; our forecast for 2012 was overoptimistic by some 7%, but not a lot. Our projection for mobile visitors was 134,000 in 2011 rising to 977,000 for 2012. This projection, with mobile visitors accounting for close to 18% of all Europeana visitors in 2012 proved to be too optimistic. Google Analytics did not break down mobile data before October 2011 so we only have a GA figure for 2012: 155,084 visits, slightly more than our forecast for 2011 but only one sixth (16%) of our predicted outcome for 2012.

Why was the forecast so optimistic? Was it because we were dealing with a relatively new phenomenon which was growing fast from a very small base? Perhaps there was confusion between numbers of pages and numbers of visitors: using log rather than GA data, which we have been advised is unreliable during the period of this study, we can see there were 244,000 mobile page views in 2011 and 1.1 million mobile pages viewed in 2012. If so, then a predicted sevenfold increase in mobile activity year-on-year turned out to be nearer a fivefold increase, impressive growth nevertheless. We estimated that 17.9% of Europeana would be mobile by now, in fact in the latter half of 2012 only 4% of visits were from mobile (including tablet) users. The numbers did however reach a record 4.6% in December 2012 and appear to be rising.

There are other variations from that 2011 report. Then we noted that France was the major user of mobile data. Now, according to GA it is the Netherlands, followed by Germany, GB, Norway and Italy. The Clickstream data tells a similar story, but here Germany ranks first (Table 2). However, it should be stressed that what is being detected here is the browser signature of an iPad or similar device not evidence that the user was using a mobile data connection. The numbers for mobiles in all dimensions are small, too easily subject to perturbation by unforeseen or undetected factors. So we need to be cautious in predicting the mobile future.

Table 2. Page views by tablet and other mobile devices: country analysis. June–December 2012, from Clickstream data

	Total	Jun	Jul	Aug	Sep	Oct	Nov	Dec	% of total
Germany	74,036	5,149	6,579	8,230	8,841	11,979	16,652	16,606	9.9
Netherlands	60,320	5,506	5,758	6,633	5,779	9,661	13,252	13,731	8.1
Spain	47,292	3,869	2,689	4,011	6,208	8,444	9,958	12,113	6.4
USA	63,337	5,560	6,966	6,885	9,746	10,630	11,729	11,821	8.5
UK	50,343	4,195	3,784	5,578	7,090	7,635	11,302	10,759	6.8
Norway	42,135	2,287	3,163	4,089	6,021	8,350	8,689	9,536	5.7
France	53,071	5,013	5,990	7,816	7,122	7,628	10,240	9,262	7.1
Italy	39,914	3,514	3,174	4,610	4,829	6,426	8,570	8,791	5.4
Sweden	25,557	1,851	3,477	2,762	3,409	3,724	5,215	5,119	3.4
Denmark	17,009	472	593	1,470	2,234	3,416	3,726	5,098	2.3
Other	271,672	36,830	56,109	35,810	34,792	31,900	39,678	36,553	36.5
All 'Mobile'	744,686	74,246	98,282	87,894	96,071	109,793	139,011	139,389	100
% Mobile	4.18%	2.96%	4.74%	3.84%	4.52%	4.08%	4.32%	4.79%	
All Users	17,819,822	2,505,288	2,074,597	2,290,744	2,126,261	2,691,612	3,220,507	2,910,813	

Table 3 provides an insightful picture of the information seeking behaviour of the mobile user.

Table 3. Page views by tablet and other mobile devices, key pages

	Total	Jun	Jul	Aug	Sep	Oct	Nov	Dec	% of total
Record	466,406	50,374	71,817	59,351	53,783	65,600	83,357	82,124	50.0
Search results	296,262	30,170	28,464	31,839	42,646	44,963	60,496	57,684	31.8
Home page	90,754	6,527	8,835	7,930	14,986	13,890	17,857	20,729	9.7
other page	34,680	3,036	2,829	4,158	5,974	6,185	6,553	5,945	3.7
Click Through	22,638	1,380	2,023	2,911	3,105	3,686	4,766	4,767	2.4
Language Change	21,428	2,102	2,471	2,616	3,081	3,696	3,949	3,513	2.3
All 'Mobile'	932,168	93,589	116,439	108,805	123,575	138,020	176,978	174,762	100
% Mobile	3.10%	2.30%	3.19%	2.86%	3.30%	2.97%	3.27%	3.68%	
All Users	30,068,262	4,068,183	3,650,008	3,804,492	3,739,577	4,652,953	5,406,776	4,746,273	

According to the Clickstream data, 'Mobile' is certainly growing (Table 2). If we bring things right up to date: comparing Nov25–Dec22 2012 to the same four weeks in 2011 we see a fourfold increase in use. But that is still less than 4.5% of all visits, partly because of the general increase in European usage overall. And mobile, at least as far as Europeana is concerned is not smartphone but tablet, and overwhelmingly iPad. The figures from GA on this are unreliable, but our analysis of the more reliable Clickstream data indicates almost half (47%) of all 'mobile' use is from an iPad. See Table 4.

Table 4. Pageviews by Tablet and other mobile devices. June–December 2012, from Clickstream data

	Total	Jun	Jul	Aug	Sep	Oct	Nov	Dec	% of total
iPad	351,855	28,676	31,452	38,682	43,961	56,965	76,244	75,875	47.2
Android	169,018	17,754	18,577	20,103	24,886	25,681	30,152	31,865	22.7
iPhone	161,358	18,928	40,057	20,395	18,100	18,489	22,743	22,646	21.7
Opera Mini	25,844	5,429	4,623	4,008	3,789	2,663	2,854	2,478	3.5
other_mobile	16,730	1,459	1,649	1,710	2,511	2,734	3,447	3,220	2.2
BlackBerry	9,166	938	912	1,255	1,325	1,642	1,830	1,264	1.2
Windows	4,274	551	305	538	447	517	813	1,103	0.6
Symbian	3,727	333	472	913	561	530	468	450	0.5
Opera	2,714	178	235	290	491	572	460	488	0.4
All 'Mobile'	744,686	74,246	98,282	87,894	96,071	109,793	139,011	139,389	100
% Mobile	4.18%	2.96%	4.74%	3.84%	4.52%	4.08%	4.32%	4.79%	
All Users	17,819,809	2,505,288	2,074,597	2,290,744	2,126,260	2,691,605	3,220,505	2,910,810	

We cannot however be sure that all of these devices were truly mobile, many were probably not out-and-about but safely tethered to home or office Wi-Fi. Unfortunately, it is difficult to identify if usage is truly mobile; doubly so if we also take into account the used of 'mobile broadband' as a primary connection. In the long-run perhaps it will not matter, the categories of desktop, tablet, and smartphone being transient and historical. If, cautiously, we believe the GA data, the most commonly reported screen size is 768x1024 (93%); as many pixels on a small screen as would have been found on a full-size desktop monitor ten years ago. The runner-up is even larger 1280x800 (4%), and note the shift in aspect ratio from portrait to landscape. Today, when even a smartphone with a five-inch screen can offer 1080p HD, screen size and viewing distance rather than pixel count or bandwidth is the more important consideration when deciding how Europeana should present content.

Interestingly, when looking at mobile use for the two weeks 24 December 2012 to 6 January 2013, a time when internet use was most likely to occur at home, we see mobile use increasing by 42% over the preceding two weeks (Dec 10–Dec 23) and a 10% drop in non-mobile use. All in all, this is a change in the way people use the 'computer', but this is tablet computing on the sofa, rather than a smartphone in the pocket.

3.2 Stickiness and loyalty (in the light of portal changes)

Stickiness has traditionally been viewed as a measure of engagement, success, satisfaction and loyalty. If someone spends a long time on a visit or repeatedly visits, then the site might be regarded as 'sticky' and that could be considered a good thing. This is especially the case where the site is not engaged in direct selling; if the value of the site cannot be measured by the revenue it generates then perhaps the value may be measured by the users it detains and retains. In the context of Europeana, however, we need to tread more carefully as it is more of a gateway, portal or search engine than a destination site, and it could be argued that

Europeana's main task is to pass on visitors to the original version of the digital object at a provider site, at a healthy rate of knots.

Our task was to evaluate the 'stickiness' of the europeana.eu portal before and after a new release in October 2011. Originally this was to have covered a period of 9 months either side of this event. However, we have chosen to consider a longer period, 12 months. Firstly, because there are several periods of anomalous (unexplained and unusual) activity, in particular the heavy burst of 'outlier' activity in April 2011 which makes comparison difficult. Secondly, seasonal activity with an academic pattern (created largely by high levels of French school use in 2011), though not so notable in 2012 as in previous years, nonetheless hinders correlation between Jan-Oct 2011 and Oct-Jul 2012. Finally, the delay in obtaining data and in preparing this report may be somewhat assuaged by making the presentation rather more up-to-date than originally envisaged.

On examining the data it soon became clear that no significant change can be seen in the log record around October 2011. So, rather than look at 12 months either side of October 2011, which would have taken us back to October 2010, we have settled on a 'long view' of two calendar years 2011 and 2012. This longer time span is not without its own problems: it covers much that is now of only historical interest and may mask emerging trends. Therefore, for some analyses, we have also considered a year-on-year comparison of September–December, 2011–2012. This in effect covers the period immediately around the launch of the new portal in October 2011 with the state of things twelve-months later.

First of all, what can we glean from Google Analytics about general changes in activity? The figures over the full twelve months (Figure 4) show healthy growth in all standard usage/visitor measures other than 'New Visits', however as mentioned below we believe all measures are unduly influenced by anomalous activity in the first half of 2011 so comparative statements about growth and stickiness need to be treated with caution.



Figure 4. Europeana Audience Overview Jan-Dec, 2012

As noted in our previous reports for Europeana Connect there was a very strong surge in site activity between February and April 2011; though we cannot fully explain the causes there was clear evidence of 'outlier' activity— a very small number of visitors accessing an unreasonably high number of pages. This could not be attributed to known API users and our best guess is that this was some form of cloaked robot or possibly an attempted denial of service attack. In addition, this time of the year (spring) had previously been associated with raised activity levels common to sites with high levels of educational use. Now, comparing two years of visits matched week to week the highly irregular pattern in the first half of 2011 is clear. Figure 4 also shows that there was a spike in activity in the week following the launch of the new portal in October 2011 and thereafter a steady weekly cycle resumed. There is no sign that the new portal had an immediate impact on visitor numbers after the first week. The story in the year since has been one of steady growth. From July 2012 onward, comparing year-on-year weekly visitor numbers have more than doubled. So visitor numbers have certainly increased since October 2011 but there is no evidence of a correlation with the changes to the portal.

Comparing a shorter, more regular and settled period: autumn 2011 (Aug–Jan) with autumn 2012 (Aug–Jan) (Figure 5) a clear picture emerges with visitor numbers growing by 120%. The numbers have been growing steadily since July 2012, but the gain 2012 over 2011 was most marked in November. The peak of activity on weekdays compared to weekends is greater, and there is a more pronounced fall-off in activity toward the year-end. The rate of growth has increased compared to a year before.

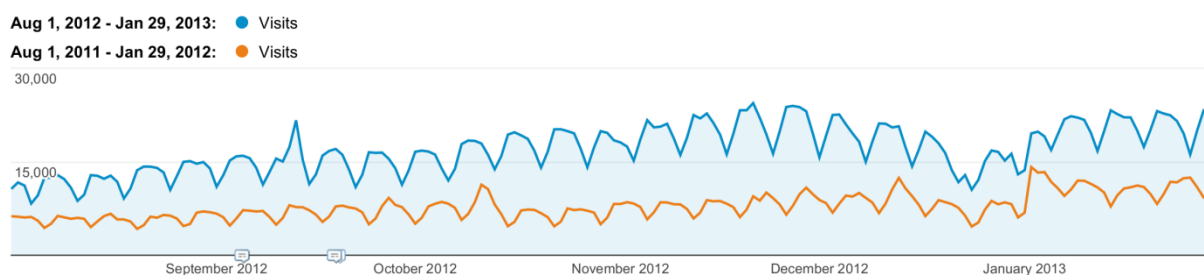


Figure 5. Visits: August 2011- January 2013 compared to same time previous year

Figure 6 charts the daily visitor count 2010–2013. Note the seasonal pattern which follows the rhythms of the school and academic calendar, the drop each weekend and holiday, and — despite perturbation— the steady spiral of growth.

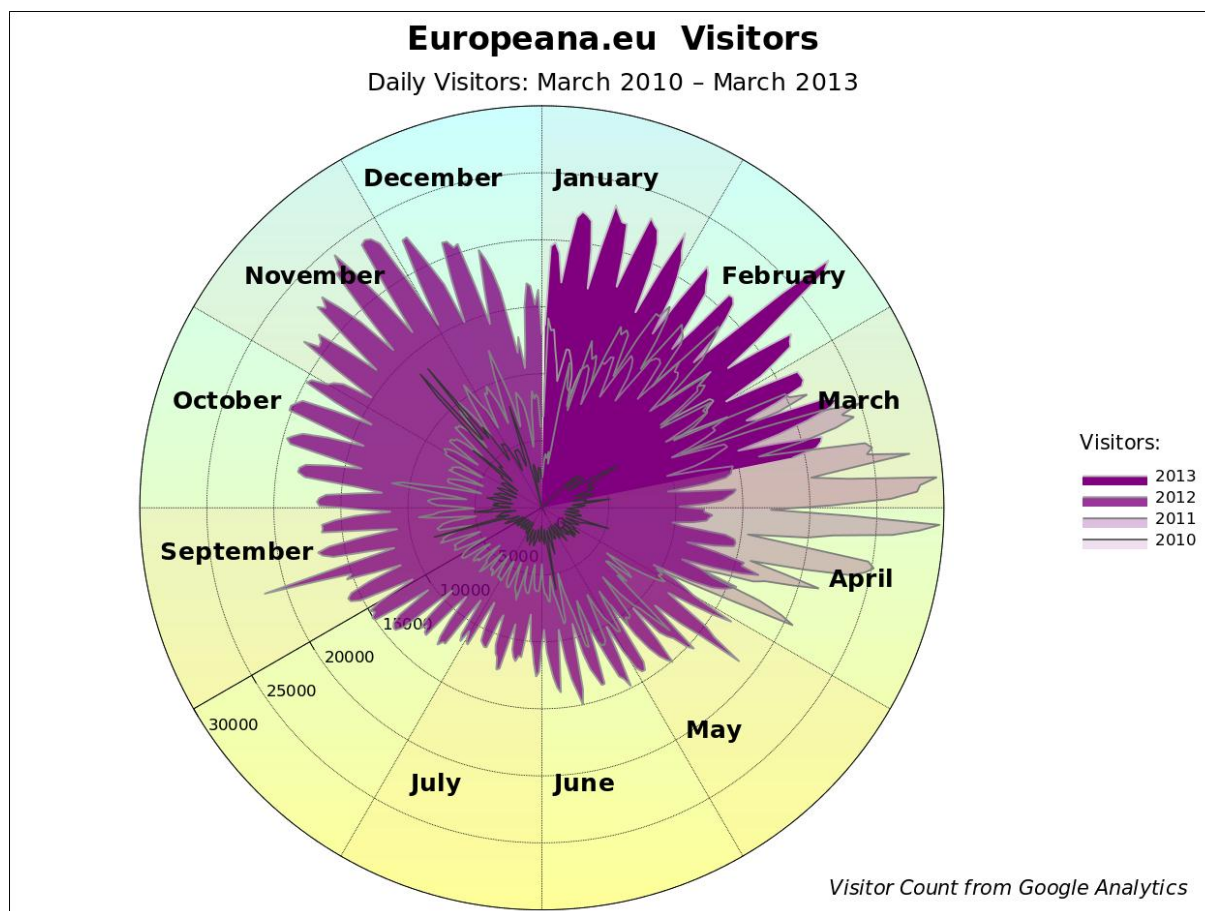


Figure 6. Europeana Daily Visits 2010–2013

3.2.1 Returning Visitors

Stickiness has most often been associated with site loyalty and the propensity of people to revisit. Returnees, unlike dwell time, are definitely a quality metric. We have not been able to undertake this analysis before on Europeana because of an absence of cookies in the raw logs (the surest method for identifying revisits). These cookies are available in the 'ClickStream' series but only from June 2012, so we are really limited to GA data. As mentioned earlier cookie-based visitor identification is not 100% reliable: cookies may be deleted; the same person may access the site from more than one browser. It is therefore probable that there is a systemic overstatement of 'New' and 'Unique Visitors' and a corresponding under recording of returning visits. But we do not know the extent and given the relative importance of this metric, far more meaningful than a Facebook 'like', for instance, Europeana should do more research to establish its real significance, by triangulating the data with demographic, survey or qualitative data.

Comparing the 12 months either side of October 2011, in the year after the portal launch a quarter of visitors were returning visitors, up from 20% in the previous year, which on paper is really an excellent performance and further research is needed to establish the reason for the lift. However, that 20% is probably unrepresentative: comparison of the more recent dataset, September–December 2011 and 2012 shows a stable returning visitor rate of 25%. Thus only one in four visitors return, as compared to two out of five for a typical publisher website. This says something about dependency. Within that 25%, 10% return only once, 4% make three visits, 2% four. Nine per cent of visitors returned five times or more. GA may understate the return rate a little but the distribution follows is a typical 'power law' (Figure 7). This suggests that Europeana's core audience, defined as those people visiting five times or more, is about one-tenth the size of its visitor numbers – about 500,000. Core audience size and growth should be regularly monitored; this tells you more about the success of the service and its long-term viability than anything else.

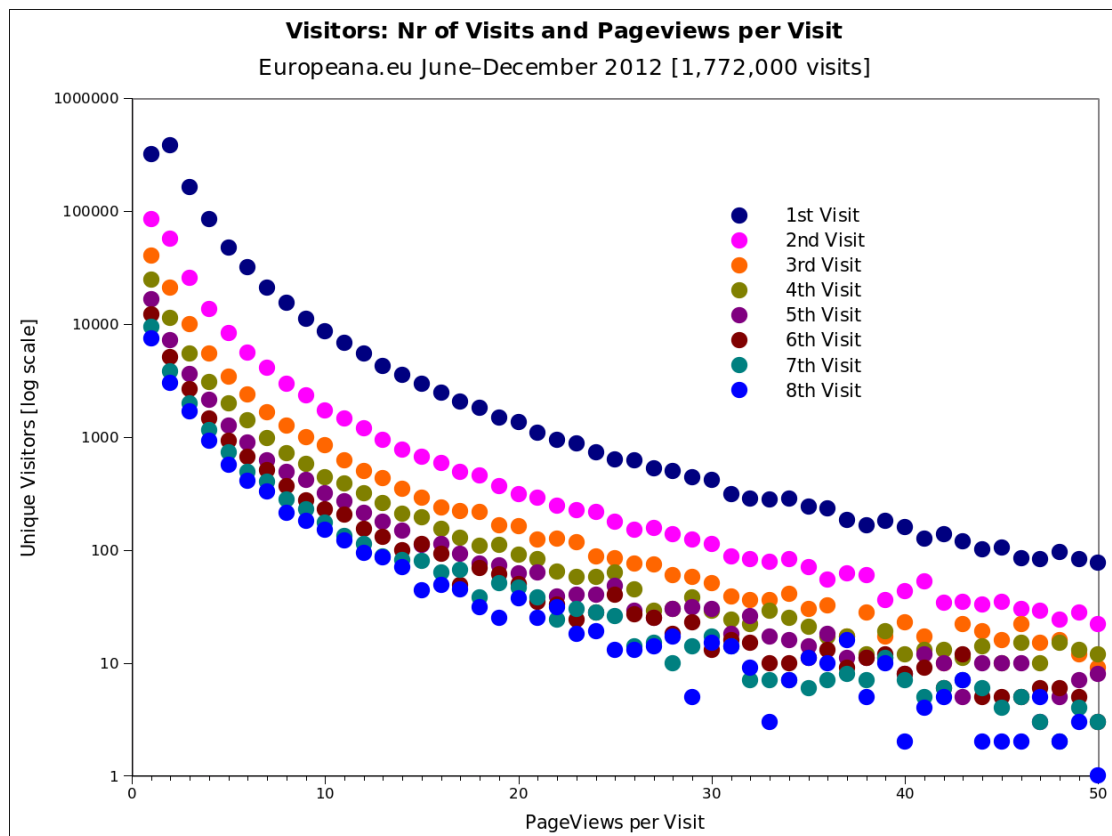


Figure 7. Europeana: return visits

When looking at returning visitors it is well to remember that most visits are very fleeting; even when bouncers are ignored many returning visits are measured in seconds rather than days. So strong is this phenomenon that it is difficult to convey on a single chart. The

following three charts (Figures 8, 9 and 10) are derived from a single dataset, a sample of 600,000 visits made between June and December 2012. These are visits selected because the Google cookies were present and contained timing data for a previous visit. The Google cookie expires 2 years after the last visit so first we look at a timescale of 24 months. In many cases the cookie will have been deleted earlier so the evidence of long-term use of Europeana will be understated. Nonetheless we do see evidence of users who first used Europeana over two years ago, and even a few who have recorded no other visit in the intervening period. But these are counted in single figures compared to the thousands who return within a month (Figure 8).

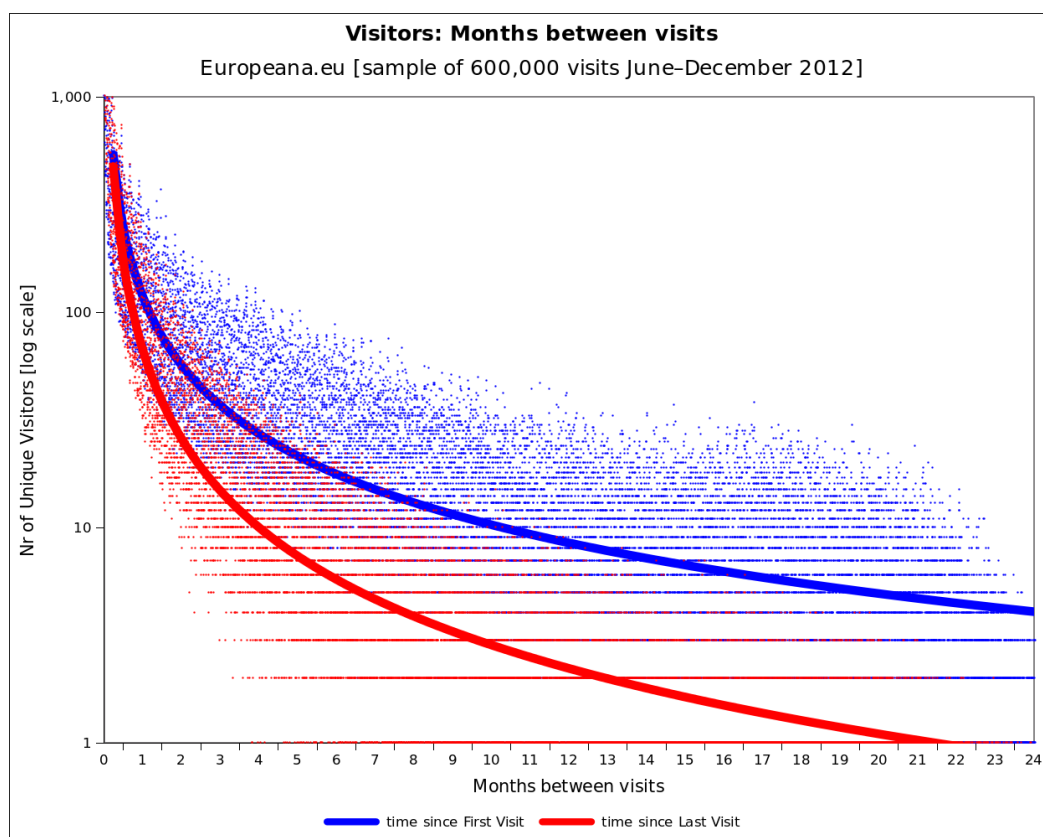


Figure 8. Europeana: months between visits

When we look even closer (Figure 9), at those visitors who return within three days rather than months an interesting pattern can be seen. Regular users appear to have a daily routine; there is a distinct series of peaks in the graph at 24, 48, and 72 hours. In fact, equipped with this insight, we can turn to a daily plot (Figure 10) and see the same daily routine persists through a whole month. It is also possible to see traces of a weekly cycle: the daily peak is a little higher at 7, 14, 21 and 28 days.

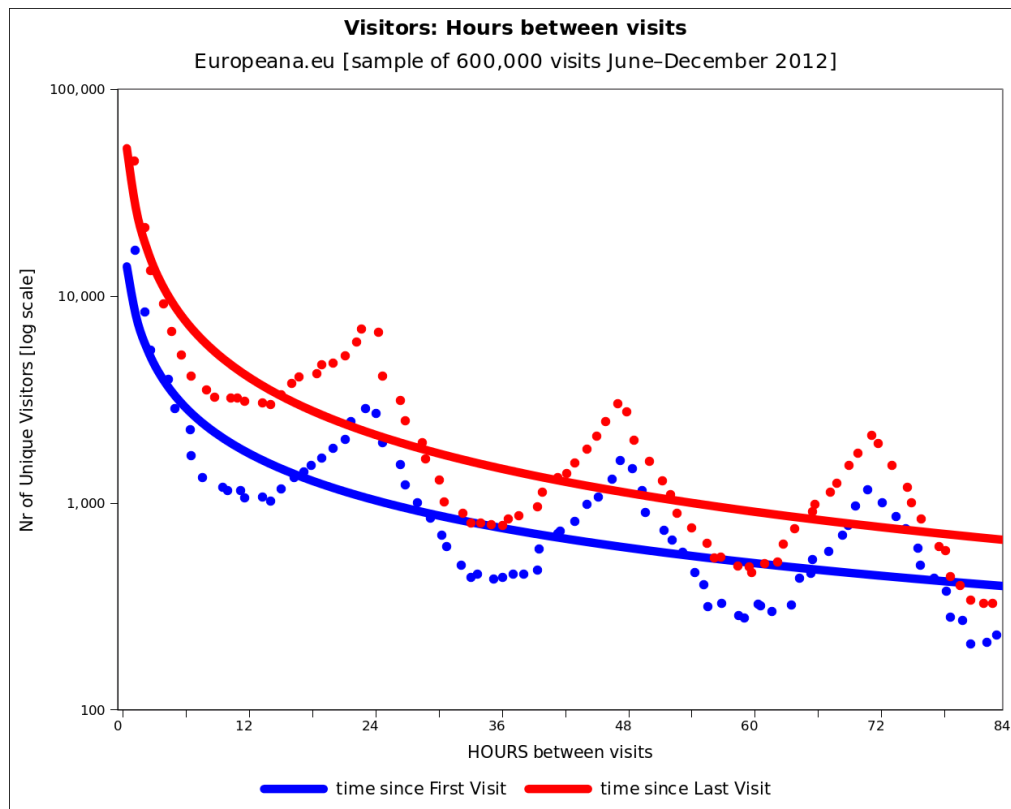


Figure 9. Hours between visits

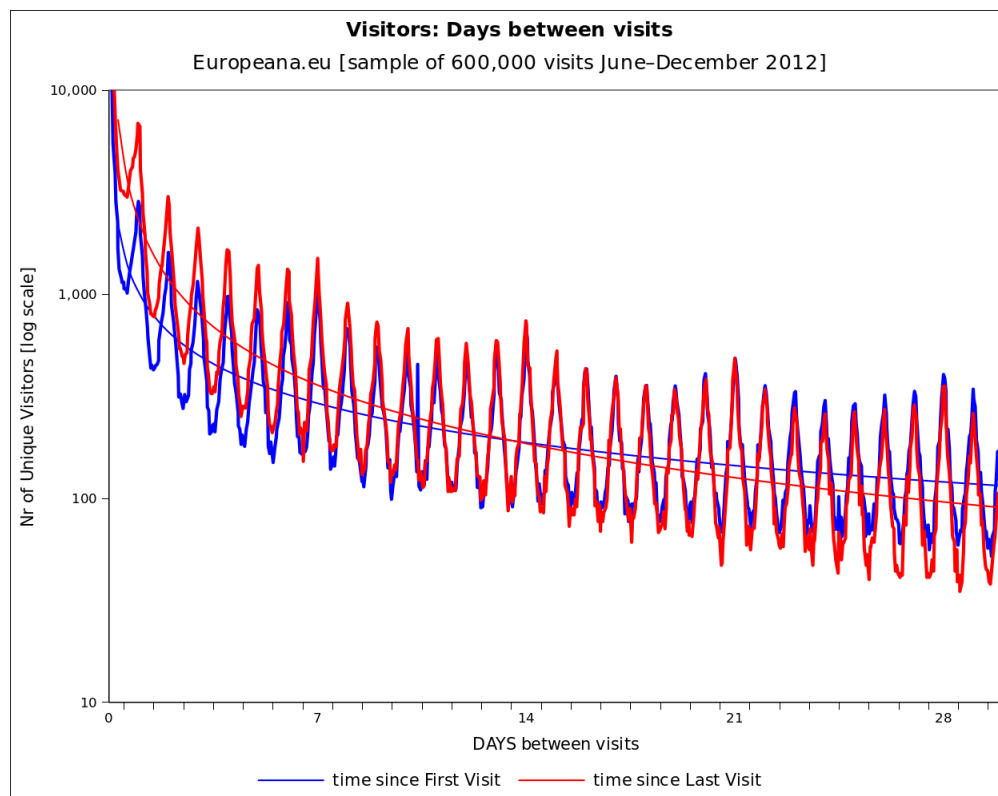


Figure 10. Days between visits

One explanation for this phenomenon is that a significant part of Europeana use takes place within institutions using browsers set up in kiosk mode. However, even when the data is reprocessed with a filter to remove the most obvious heavy institutions referrals the daily pattern persists.

3.2.2 Engagement

We can calculate engagement by considering both: a) duration of a visit; b) numbers pages viewed during a visit. The most recent data shows that 60% of visits are very short (<10secs); and less than 2% are recorded by GA as exceeding 30 minutes (the normal cookie timeout for a visit). Most visits are over in the blink of an eye. This is probably what we would expect of a discovery site rather than a destination site, where the times are much higher. In terms of page views 58% looked at just one page, less than 5% view more than 16 pages. Of course this comes with short visits. The site's character is of course changing with the introduction of virtual exhibitions and when we look at just virtual exhibition use we see people dwelling longer and examining more pages (see Table 13).

When looking at figures for duration of visit it is important to note the highly skewed distribution: most visits are very short, a table with ranges of values can be misleading, as is any 'average' figure. Figure 11 and Tables 5 show visit times for December 2012. The average visit duration is 2 minutes and 19 seconds, it varies little depending on what time span is analysed, whereas the chart reveals the full picture: there is much larger range, a few visits are very much longer, but most are extremely short. In December 2012 fifty-eight per cent of visits were timed at less than ten seconds; only ten per cent of visits fall broadly (1–3 minutes) into the 'average' category band.

Table 5. Duration of Visits, December 2012

Visit Duration	Visits
0-10 seconds	311,193
11-30 seconds	57,548
31-60 seconds	39,689
61-180 seconds	52,233
181-600 seconds	39,245
601-1800 seconds	23,132
1801+ seconds	7,707

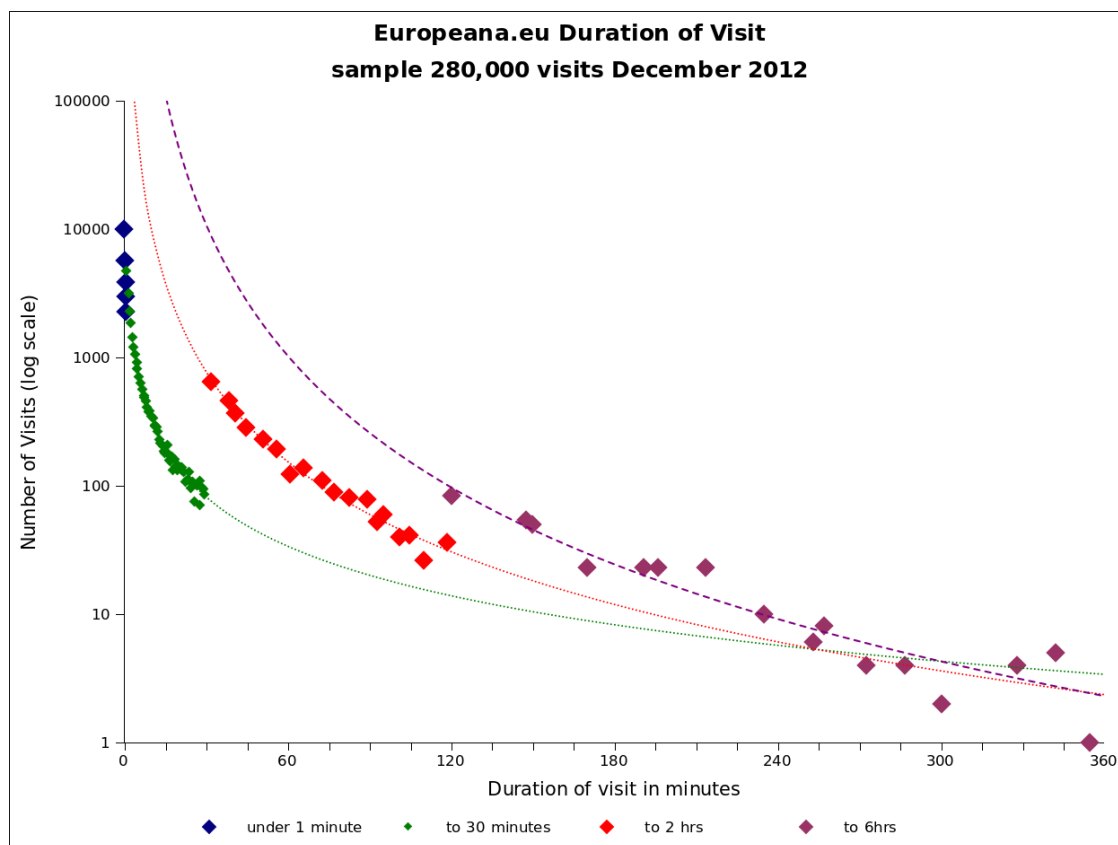


Figure 11. Europeana: duration of visit

The story on 'engagement' is interesting and merits further investigation: page views have not kept pace with overall growth rates having grown just over 60% from Autumn to Autumn and with a huge fall (nearly 30%) being recorded in the number of pages viewed per visit (was previously 5.4 and now 3.8) and a smaller, but still large fall (nearly 17%) in the duration of visits. 'Average' is a very poor measure of visit duration so not much can be read into a decline in this figure from 2:46 seconds, to 2:18 seconds. Especially as the Bounce Rate has fallen (from 54% to 50%), and we might have expected this to go up in the circumstances. So, it is probable that 'stickiness' has increased (fewer bouncers), but is partly masked by a corresponding reduction in the number of 'unreal users' consuming many pages in long sessions.

These 'unreal users' are not search-engine spiders which are already excluded from the analysis. Nor do we mean 'outliers' which are cases where we have come to a firm conclusion that the activity is that of a cloaked bot. Once we have discounted these we are still left with patterns of activity that are implausible, such as sessions that never time out or appear to view an unreasonable pages etc. In some cases that can be explained by kiosk applications in libraries, API usage, or by developer testing. (During the Europeana Connect project CIBER's own usage of Europeana put us into the top 1,500 of Europeana.eu users and yet our usage was not very extensive, nor typical.) Essentially 'unreal users' are that portion of the recorded

usage which we find 'not proven'. There is insufficient evidence to classify as robot or outlier, but the suspicion remains that it would be unwise to fully trust any inference from this data.

3.3 Information seeking patterns

The top page in Table 6 is what might be expected. That the 'record' is the most heavily used item is not surprising, it is the essential core of Europeana content, search (browse thumbnails) is the essential route to this content (or a surrogate for it); the homepage is where most users start. The last two raise interesting questions. Click-through pages are links that lead to the provider site; they are therefore not page-views that are attributed to Europeana but to the provider site. And yet they are perhaps the essence of the function and purpose of Europeana.eu. And 'language change'? This is a JavaScript function that changes the language version of the page, but the 'page' —its content and purpose, remain: should we count this as a new page or just a page refresh? This all has ramifications for counting use and determining performance.

Table 6. Pageviews June–December 2012, from clickstream data

	Total	Jun	Jul	Aug	Sep	Oct	Nov	Dec	% total
Record	9,601,835	1,401,403	1,090,328	1,327,643	1,055,039	1,457,960	1,763,442	1,506,020	55.2%
Search results	4,844,240	623,354	637,496	593,880	617,831	740,380	866,344	764,955	27.9%
Home page	1,576,561	170,274	156,963	185,180	233,521	255,774	297,690	277,159	9.1%
Click Through	720,555	46,204	104,844	104,962	92,998	111,886	134,857	124,804	4.1%
other page	391,207	100,202	28,446	26,640	55,196	43,019	57,795	79,909	2.2%
Language Change	253,348	28,981	27,405	27,517	34,729	43,681	50,202	40,833	1.5%
All Pages	17,387,746	2,370,418	2,045,482	2,265,822	2,089,314	2,652,700	3,170,330	2,793,680	100.0 %

In considering user behaviour this matter of what should be counted, pages or clicks, is important. Pages may not be viewed but merely scanned in the search for something else. This is attested by the large number of page views that can be recorded during very short visits: five or six seconds per page is common. It is rare for the dwell time on any page to exceed five minutes. If there is not a transition to a new page within ten minutes then the visit has probably ended. On the other hand clicking to change language may be a better sign of engagement than the number of pages consumed.

Table 7 shows the traffic flow around Europeana.eu: each row represents a link, each column a landing page. The table is based on traffic recorded by ClickStreamer logs for December 2012. All figures are percentages of the total valid clicks including language changes and the 'show SIWA menu' action. For clarity percentages less than 0.01 are omitted. The SIWA action displays a menu "Click to search for this term in External services", and redirects the user to external search engines. However this action, unlike the similar 'click-through' to a

provider site is not logged, so we can only say the drop down menu was displayed not that the link was followed. Thus though we can record inward traffic (either 'Google' or 'external'), apart from the 4% click-through to provider sites the exit path is unknown.

3.3.1 User activity by landing page

Table 7 examines page flow over one month. Below, tables 8 and 9 use a similar technique but cover a longer period June–December 2012 and concentrate on the key pages of the canonical visitor flow: Homepage, Search, Record, Redirect. On the referral side we can reliably identify four sources of inbound traffic: Google search, Social referrals, API referral, and other Referral including direct traffic. Adding these together will give a total that represents 'visits'. The 'in-site' category covers all referrals from pages within the site thus tables 8 and 9 summarise the detail presented in table 7. Overall there are around 2.1 pageviews per visit, thus these internal referrals constitute 107% of all pages viewed.

Some external referrals are to the redirect: the user is passed on direct to the provider, probably unaware that the link they followed has bounced off the Europeana site. This is least likely to occur with API and Google referrals, most often with 'social' where it accounts for 2% of referrals. Overall such pass-throughs only account for 1% of all visits. This is not enough to support any significant interpretation but might be the starting point for a future investigation of the 'dark social' phenomenon: the followed link has most likely been cut and pasted from some informal source.

Table 7. Page Flow All Users Europeana.eu, from clickstream data

From Page/Site	Total	Record	Search	Home Page	Click-Through	OTHER	LANGUAGE_CHANGE	SHOW_SIWA_MENU	BROWSE_AL L:landing	TIMELINE	SITE_MAP:providers	MAPVIEW	LOGIN:myeu ropeana	BROWSE_AL L:index	MY EUROPE ANA	SAVE_ITEM	REGISTER
SEARCH	26.9	7.23	19.33	0.15		0.05	0.11			0.04	0.01	0.01	0.01				
EXTERNAL (not Google)	25.6	11.63	3.23	7.06	0.25	1.25	0.78	0.53	0.39	0.14	0.14	0.14	0.03	0.02			0.01
GOOGLE	22.8	22.00	0.05	0.66	0.01	0.02	0.03										
RECORD	18.2	10.73	2.89	0.17	3.77	0.35	0.07	0.19		0.01			0.01			0.01	
HOME PAGE	3.0	0.16	1.37	0.67		0.23	0.38	0.03		0.02	0.04	0.02	0.03				
browse-all	2.4	2.36							0.06								
OTHER	0.5	0.03	0.18	0.06		0.12	0.02			0.01	0.03	0.02		0.01			
timeline	0.1	0.02	0.01	0.02		0.01				0.03		0.01					
newcontent	0.1		0.07			0.01					0.01						
usingeuropeana_search	0.1		0.04	0.00		0.02											
login	0.1			0.02		0.01							0.03				
index	0.1		0.03	0.01			0.01										
usingeuropeana	0.04		0.01			0.03											
aboutus	0.04		0.01	0.01		0.02											
usingeuropeana_explore	0.03		0.00			0.02											
rr-f	0.02		0.01														
usingeuropeana_myeuropeana	0.02					0.01											
aboutus_faqs	0.02					0.01											
register	0.01			0.01													
myeuropeana	0.01																
sitemap	0.01					0.01											
contact	0.01																
usingeuropeana_results	0.01					0.01											
aboutus_enevs	0.01																
Total	100.0	54.16	27.24	8.84	4.04	2.19	1.42	0.75	0.45	0.27	0.23	0.21	0.12	0.03	0.01	0.01	0.01

It is important to keep it mind the big picture: search engines send users direct to objects, other referrals are more likely to send users to the home page. Given the way search results are generated and presented providing a link to a search page is not a very reliable way of pointing friends or associates to the very same content that you currently see in your browser.)

Table 8. Summary of overall traffic flow around the site based on ClickStream log data June–December 2012.

Referral from	Total referra	Landing page					
		SEARCH	BROWSE_ALL	RECORD	HOMEPAGE	REDIRECT	OTHER
In-site	8,903,515	4,236,721	9,990	3,654,179	189,928	652,471	160,226
Google	4,050,969	8,736	1	3,917,187	118,415	2,550	4,080
Referral	3,849,722	477,449	72,618	1,931,115	1,104,190	37,702	226,648
Social	332,856	86,623	0	88,251	149,998	7,105	879
API	66,174	11,317	0	51,579	3,188	23	67
Total Pages	17,203,236	4,820,846	82,609	9,642,311	1,565,719	699,851	391,900
implied 'visit'	8,299,721	584,125	72,619	5,988,132	1,375,791	47,380	231,674

Table 9. Summary of overall traffic flow around the site as percentage of visits.

FROM/To	Total referral	Landing Pages					
		SEARCH	BROWSE_AL	RECORD	HOMEPAGE	REDIRECT	OTHER
In-site	107%	51%		44%	2%	8%	2%
Google	49%			47%	1%		
Referral	46%	6%	1%	23%	13%		3%
Social	4%	1%		1%	2%		
API	1%			1%			
Pages per Visit	2.1	0.6		1.2	0.2		
VISITS	100%	7%	1%	72%	17%	1%	3%

The majority of visitors land first on a record (object) page, almost all if they have arrived via Google, half if direct or referral; only social referrals are more likely to begin at the home page. The Visitor Flow report in GA suggests that the bounce rate is higher from object pages. E.g. up to 85% drop off from /portal/record against 50% from /portal/. As most Google Searches land direct on an object page it is highly likely that most visitors instantly decide they have not found what they are looking for and go away. In contrast the most consistently used search terms in Google searches are 'Europeana' or variants thereof: such visitors are using Google as a safer way to type a URL, they probably already know Europeana and are more likely to be intending to find site content rather than direct access to objects.

3.3.2 Types of journey:

This follows from the previous section. Tracing journeys is a complex task, neither logs or Google Analytics provides all the data we need to this reliably. We do not have sufficient access to GA to use the flow graphing feature effectively, and the ClickStream series of logs, though an improvement on the old http-access series still present difficulties when attempting to identify individual sessions. So what follows is the best answer we can give using what we have. We can reliably identify four sources of inbound traffic: Google search, Social referrals, API referral, and other Referral including direct traffic. Adding these together will give a total that represents 'visits'. (This definition of 'visits' will overstate the number relative to more conventional methods but for the purpose of tracing traffic flow the difference can be discounted.) In addition traffic will flow from page to page internally within the site. Table 10 represents the figures as a percentage of the total number of visits as defined above.

From Table 9 it can be seen that 72% of all visits begin at a Record (object) page, 17% at the homepage; 7% with a search. Social sources account for 4% of visits, API users 1%. Note that as this is based on our own analysis of the clickstream, we are using a broader classification of Social referral than Europeana's segmentation of GA. In particular we identify all referrals from blogs as 'Social'; there are consequently over 2,800 sites within our 'social' category.

Table 10 looks at the internal flow in more detail but still simplified to remove the clutter of many minor pages.

Table 10. Internal (within-site) traffic flow June–December 2012

From/To	Total Referral	SEARCH	BROWSE_ ALL	RECORD	HOMEPAGE	REDIRECT	OTHER
Search	53.56	38.68		14.46	0.30		0.12
Record	35.62	5.78		21.48	0.33	7.32	0.71
Homepage	5.04	2.81		0.33	1.36		0.54
Browse_all	4.83		0.11	4.72			
newcontent	0.16	0.13			0.01		0.02
Timeline	0.13	0.03		0.05	0.03		0.02
Login	0.11	0.01			0.03		0.08
using europeana	0.32	0.11			0.02		0.19
aboutus	0.10	0.02			0.02		0.06
Rr	0.03	0.01			0.01		0.01
myeuropeana	0.02	0.01		0.01			0.01
Total Internal	100.0	47.6	0.1	41.0	2.1	7.3	1.8

In Table 10 we are looking only at non-bounce visits; hence referrals from the homepage are low, the bouncers have departed; the homepage leads (3%) to the search page. The key feature is as might be expected given the site design a loop-around the search pages: 39% of search

page landings lead to another search page, 14% exit the loop by going to a Record. From the Record 6% go back to a search, 21% another record, 7% take the redirect to the provider site.

3.4 Social media

3.4.1 Definitions

With so much Europeana planning resting on the social media use of Europeana it is worth first pointing out that there are substantial problems in defining 'Social Media' which need to be clarified in order to make a fair and accurate evaluations and comparison of growth rates and contribution to overall traffic. Even in only to ensure we are all on the same page!

1. The Google Analytics 'advanced segment' for social media, as defined and used by Europeana, contains 20 sources (referrer domains), some of which have registered insignificant or even no traffic at all during the last six months (October 2012–March 2013). The major sources of social traffic are Facebook, and Wikipedia; there is also significant traffic from WordPress, Blogspot, twitter and, a considerable way behind, Pinterest.

Table 11. Social segment definition

Social	Europeana	CIBER A
facebook.com	16928	16928
tweet	0	0
LinkedIn	361	361
YouTube	42	42
reddit	167	167
digg	14	14
delicious	83	83
stumbleupon	0	0
Flickr	222	222
MySpace	0	0
hootsuite	28	28
retronaut	167	167
Wikipedia	10882	10882
bit.ly	0	
tinyurl	0	
t.co	9993	2335
wp.me	0	
blogspot	4044	4044
wordpress	4253	4253
Pinterest	1265	1265
sum	48449	40791
intersect	44085	40791
Oct-Mar	3473308	3473308

2. Twitter. [See Table 11]. The Twitter traffic is identified by "include Source containing 't.co'". Patently, this is too loose a definition as it will not only pick up 't.co', but any domain containing that sequence of characters e.g. search.bt.com. The result is that the number of visits captured by this method is at 9,993 (for the most recent six months) four times greater than the actual number of visits from t.co (Twitter). The true total of social sources (40,791) is inflated by 19% (48,449). The overall effect on the visit count for the social segment is to some extent mitigated by the fortunate chance that the loose 't.co' rule will pick up blogspot.com which is already included by its own rule. You can fix this problem by replacing the rule "include Source containing 't.co'" with "include Source Exactly matching 't.co'" or with "include Source Matching RegEx ^t\\.co\$".

Table 12. Social segment blogs

	Oct 1, 2011–Mar 31, 2012	Oct 1, 2012–Mar 31, 2013
agioritikesmnimes.pblogs.gr	64	0
albufeirasempre.blogs.sapo.pt	21	0
alea-blog.blogspot.com	21	0
athos.weblog.nl	0	21
bazoga.over-blog.com	43	0
bgpw.blog.pl	21	21
blog.apahau.org	0	21
blog.biodiversitylibrary.org	0	21
blog.bnf.fr	21	21
blog.crdp-versailles.fr	43	0
blog.daum.net	21	43
blog.eduzones.com	0	21
blog.euscreen.eu	0	64
blog.hslu.ch	0	21
blog.kairaven.de	21	0
blog.le-miklos.eu	21	0
blog.myheritage.com	0	21
blog.myheritage.fr	21	0
blog.renren.com	0	21
blogs.ec.europa.eu	21	21
blogs.elpais.com	0	21
blogs.epb.uni-hamburg.de	0	21
blogs.helsinki.fi	21	0
blogs.hoy.es	0	21
blog.sina.com.cn	0	64
blogs.kb.dk	0	21
blogs.law.harvard.edu	64	0
blog.slub-dresden.de	21	0
blogs.sch.gr	43	43
blog.uchceu.es	0	21
boblog.corrieredibologna.corriere.it	0	21
boqo.over-blog.com	21	0
cblog.culture.fr	43	0
christypato.blog.br	21	0
cyber-base.lgsl.over-blog.com	21	0

deal.blog.kazeo.com	0	128
deal.blog.mongenie.com	213	0
digiblog.hu	21	0
elishean.unblog.fr	0	21
enfinlivre.blog.lemonde.fr	43	0
estudamais8.blogs.sapo.pt	21	0
eueublog.wordpress.com	0	64
fablog.iransalamat.com	43	0
filmstillsresources.blog.com	0	21
formacion.universiablblogs.net	43	0
googleblog.blogspot.com	21	0
historypcs.edublogs.org	0	21
kluwercopyrightblog.com	21	0
konzervativci.blog.com.mk	21	0
leblog-ffg.over-blog.org	43	21
lehavreblog.blogspot.fr	0	21
libblog.ucy.ac.cy	21	0
mostrefestivaleventi.over-blog.it	0	21
photoblog.alonsorobisco.es	21	0
pisani.blog.lemonde.fr	43	43
poetaille.over-blog.fr	0	21
riposte.sefarade.over-blog.com	0	21
sog.blog.so-net.ne.jp	0	21
somewhereinblog.net	43	43
thehistoryblog.com	0	21
tyukudvar.blog.hu	0	21
yahoo-curious.blog.so-net.ne.jp	0	21

3. Blogs [See Table 12]. The social segment includes blogs but only those from WordPress and Blogger. There are many other blogs hosted elsewhere that are not included. On the other hand treating all referrals originating from a WordPress or Blogger domain may be too broad a definition of a blog. WordPress in particular is a popular hosting platform for photographers' and artists' galleries. No method of classification will be entirely satisfactory but on balance we think the 'social' classification should be broadened to include any domain containing the subdomain 'blog.' or 'blogs.', but excluding blog.europeana.eu. The result is that another 1,085 visits can be added to the social segment.

4. Google Analytics provides under "Traffic Sources" a "Social" analysis. Looking at the "Network Referrals" section of this report it is clear that the GA definition of 'social' is again far broader than either Europeana's own 'Advanced Segment' definition or the corrected and extended version used by CIBER. How many networks are included depends on the period of the report: for March-April 2013 it includes 48, Jan-May 2012/13 includes 78 etc. The definition is as long as a piece of string.

To conclude there are three 'social' definitions at work here: Google, Europeana's social segment, which we were requested to use, and CIBER's own expanded version based on a corrected version of the Europeana social segment. It clearly would be a good thing if we

could "align ourselves with how Google defines this traffic group via their 'social' category within Analytics to avoid confusion"; but that depends on the Google definition being available, as a specification we can apply to our own analysis and as well as a dimension available in GA custom reports.

3.4.2 Size and growth (October 2012 – January 2013)

To place social media referrals in context it is worth first looking at all referrals. Seventy per cent of the 4.5M visits to Europeana in the past year were search referrals, nearly all (97%) from Google. By contrast, runner-up Bing accounts for just 0.5%. Eighteen per cent of visits originate as links from other sites, 11% are direct —typed-in or bookmarked— and campaigns (newsletters etc.) contribute a little over one per cent.

Google Analytics was not reporting social referral before Oct 2011, so we have a limited time series, which we can to some extent enhance with log data. The limited data we have show that there was a slight peak in social referrals around the time of the portal launch in October 2011, but after that it settles down to around 1,000 per week; since August 2012 there has been some irregular growth and the base-rate is now nearing 1500 per week. In April 2013 Social Referrals only accounted for one per cent of all visits to the site, a bare 0.02% higher than a year previous. It could be that Europeana's social media activity takes place solely within the context of these sites and entirely by-passes Europeana.eu. In which case, for the purposes of this study we have very little data to measure, no benchmarks to evaluate. In such a context we cannot refute claims for the efficacy of 'social media', nor can we support them. In the context of the Europeana.eu web-site however social referral is not at present significant and is not growing above the trend for the site as a whole. So the action has to be happening elsewhere.

3.4.2.1 Size and growth update (October 2012 – March 2013)

Table 13. Growth rates 2011/ 2012 v 2012 /2013

		Visitors	Visits	Pageviews	P/V	Duration	Bounce	New Visit
Europeana.eu	2012/13	2696246	3473308	13288682	3.83	00:02:19	49.8	74.9
	2011/12	1413735	1842392	9537422	5.18	00:02:37	55.8	74.4
	All	91%	89%	39%	-26%	-11%	-11%	1%
Europeana.eu	2012/13	33067	39961	213233	5.34	00:03:04	41.43	74.4
	2011/12	24704	30641	140843	4.6	00:02:58	54.72	72.1
	Social Segment (original)	34%	30%	51%	16%	3%	-24%	3%
Europeana.eu	2012/13	25577	32024	174932	5.5	00:03:17	41.1	70.5
	2011/12	20385	26024	121735	4.7	00:03:06	55.2	70.3
	Social A [excluding Blogs]	25%	23%	44%	17%	6%	-26%	0%
Europeana.eu	2012/13	8448	8937	42749	4.8	00:02:10	44.5	90.2
	2011/12	5362	5788	25279	4.4	00:02:25	50.7	81.6

	Social B [Blogs]	58%	54%	69%	9%	-10%	-12%	11%
Europeana.eu	2012/13	34025	40961	217681				
	2011/12	25747	31812	147014				
	Social A+B	32%	29%	48%				
Europeana.eu	2012/13		33961					
	2011/12		26476					
	GA Social		28%					
Exhibitions	2012/13	58201	73341	876943	12	00:02:56	0.62	77.9
	2011/12	44507	52854	347051	6.6	00:02:58	31.9	82.6
	All	31%	39%	153%	82%	-1%	-98%	-6%
Exhibitions	2012/13	2532	3273	35025	10.7	00:02:42	0.76	74.6
	2011/12	1418	1840	6585	3.6	00:01:40	45.2	71.9
	Social A [excluding Blogs]	79%	78%	432%	197%	62%	-98%	4%
Exhibitions	2012/13	404	461	8517	18.48	00:05:04	0.87	82.4
	2011/12	157	174	1118	6.43	00:02:30	28.16	79.9
	Social B [Blogs]	157%	165%	662%	187%	103%	-97%	3%
Exhibitions	2012/13	2936	3734	43542				
	2011/12	1575	2014	7703				
	Social A+B	86%	85%	465%				
Exhibitions	2012/13		3314					
	2011/12		1970					
	GA Social		68%					
Blog.europeana	2012/13	45027	52073	79115	1.52	00:01:06	74.3	84.9
	2011/12	24480	32359	54842	1.69	00:01:38	68.3	74.2
	All	84%	61%	44%	-10%	-33%	9%	14%
blog.europeana	2012/13	7027	9134	13701	1.5	00:01:33	76	71.4
	2011/12	5120	7361	10895	1.48	00:01:39	73.5	63.8
	Social A [excluding Blogs]	37%	24%	26%	1%	-6%	3%	12%
blog.europeana	2012/13	454	539	807	1.5	00:01:16	77.9	81.6
	2011/12	134	173	284	1.64	00:01:24	70.5	68.8
	Social B [Blogs]	239%	212%	184%	-9%	-10%	10%	19%
blog.europeana	2012/13	7481	9673	14508				
	2011/12	5254	7534	11179				
	Social A+B	42%	28%	30%				
blog.europeana	2012/13		9423					
	2011/12		7793					
	GA Social		21%					

Delays in the production of the report, fortuitously, made it possible to extend the period of analysis until April 2013 and thus analyse a six-month period October to March and compare it with the same period of the previous year. This comparison of 2011/12 with 2012/13 has particular advantages for the Europeana data as it concentrates on the months which are most active and omits the April–May period of 2011 and 2012 which were affected both by a drop in Googlebot activity and an anomalous and unexplained rise in server hits.

So what have we learnt about social growth between Oct-March 2011/12 and 2012/13? [See Table 13]. The overall year-on-year visitor growth is 90%. However if we look at the 'social segment' the visitor growth is 34%. Exclude blogs [Social A] and visitor growth falls to 25%. Looking at blogs alone [Social B] the visitor growth rate is 58%. The social element is a little more significant on the exhibitions site and predictably significant for blog.europeana.

3.4.3 Individual social media

The dominant network is Facebook with nearly 30,000 referrals in the year since the new portal launch. The 'average visit duration' of these Facebook sourced visitors is, according to Google Analytics just over 3 minutes. Although 'average' is a poor single metric to use in this context —the distribution being log-normal— the duration is slightly higher than the 2.5 minute average for all visitors. So more dwell time for social media users, but not really sufficient to build a strong case for more committed users, and anyway see our earlier comments about the problems of using dwell time in isolation as a metric.

Facebook was followed by WordPress in popularity, nearly 9000 referrals, Blogger (over 4200), Twitter (nearly 3300) and Netvibes (just over 2000).

When we consider and compare only the relatively stable Autumn months (Sept–Dec, 2011 and 2012) the overall doubling of traffic on the site is not matched by a corresponding growth in social referrals year on year: Facebook (nearly 10000 referrals, 2012) and Twitter (1650 referrals) traffic in particular shows only a 12% increase in visits. Only WordPress, with only a third of the Facebook traffic (3037 referrals in 2012; 162% year-on-year growth) has kept pace with the overall pace of the site. However, Twitter is an interesting case because while there is little growth in referrals, dwell time has in fact doubled. The average for Twitter was 2.5 minutes in autumn 2011, 5 minutes in 2012. Pinterest, Europeana's latest social media venture, a content sharing service that allows members to "pin" images videos and other objects to their pin board, currently featured on the Europeana homepage (and so attracting considerable publicity), surprisingly perhaps comes in at 6th in the social media ranking, with a light traffic flow (681 visits Sep–Dec 2012). The high number of page views per visit from Pinterest (average 12) and very long dwell time (12 minutes) suggest 'unreal user' activity, something odd is happening here. We suspect, as this feature on the home page is quite recent, that this may be internal development or testing activity. This should be checked, otherwise a false impression might be provided.

We can contrast the traffic flow for the site as a whole given in Tables 7–10 with the flow of social media visits presented by Google Analytics (Figure 12). For the site as a whole most inbound traffic goes direct to a record (about half of all non-search engine referrals) and twelve per cent to 'search'. Interestingly, for social referrals half the inbound traffic goes to the homepage and around seventeen per cent to 'search'.

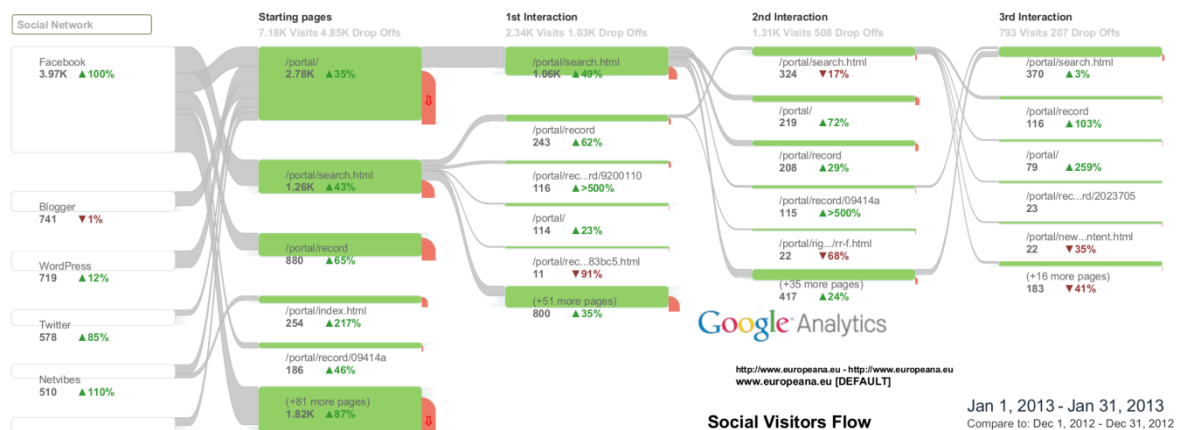


Figure 12. Analytics: Social Flow

An informal analysis of 'trackbacks' provided by Google Analytics suggests that much of the social traffic may be by people involved in development or research in digital humanities and related fields, not a very representative group: insiders. During the period, 30 Dec 2012–29 Jan 2013, when there were 6,628 social referrals, blog.europeana.eu had 8,000 visitors. It is probable the blog users are already familiar with Europeana in which case it is probably not bringing in many new users.

'My Europeana' may also be considered to belong in the 'social' category. Between June and December 2012 a total of 1400 users were recorded as having logged in with a userid, less than 300 even in the busiest month of November. Though a few users appear to login and view many hundreds of pages in a month there is little evidence of regular and sustained use of 'My Europeana', less than 50, a tiny amount, have used the feature for three or more months in seven. The majority appear in the record one month only during which they view less than forty pages. Overall logged-in users account for one-half per cent of all page views. Given the high hopes once vested in My Europeana, and the lessons that can be learnt from its relative fall from grace, further research is fully merited.

Note also the use of the 'save social tag' action in the ClickStream log. This has only been used a total of 189 times between June and December 2012.

Social media, then, is not driving Europeana growth, and unlikely to do from the evidence we have to hand. The example of Pinterest is illustrative. Consider the featuring of a link to Pinterest on the Europeana homepage: this would appear to be of net benefit to Pinterest. Europeana has over 2,000 'followers' on Pinterest and over 600 'pins', but referrals back to Europeana during the last four months of 2012 amount to 680. The big question is what, in the context of Europeana, is social media for? Should we expect it to drive traffic to Europeana, or is Europeana the glue layer that enables Pinterest to be a showcase for Europeana's provider institutions? There is scope for a more comprehensive research

programme in this area, linking together the traffic analysis of the Europeana web-presence (including blog, exhibitions, API) with similar data drawn from Europeana's providers.

Of course, all these social media initiatives are insignificant compared to 'the bread and butter' search engine referral; and not just via Google: pionier.net.pl the Polish aggregator site brought in 37,000 (5%) referrals, compared to facebook.com 29,000 (<4%), with all its millions of members. A study by Europeana (<http://pro.europeana.eu/pro-blog/-/blogs/1660413>) shows that API use by the Polish partners is proving very successful in sending traffic to Europeana.

Most of the social media (narrowly defined) traffic appears to flow into the home page rather than to specific items. This is in marked contrast to referrals from blogs which are more often to a specific page.

3.4.4. Country analysis.

First a note of caution: determining the users location is only approximate and, particularly when looking at the standard Google Analytics report, language choice and country are not the same: 'Language: en-us' is not the same as 'Country/Territory: United States'. The language indication is merely the default setting of the browser and cannot be relied upon. Location, which is based on IP address allocation, can also cross borders.

Taking this into account it is still somewhat surprising to find that the most active country for social media traffic to Europeana.eu is Spain. In the most recent six-months Spain accounts for 8.8% of social media traffic as defined by Europeana's own Social Media advanced segment, the USA is second (7.1%). Taking into account the much larger population of the USA and the mature state of social media uptake there this is unexpected. However, as we have already observed Social Media accounts only for 1 per cent of visits and visitors, so the statistics are likely to be unstable and be perturbed by factors which can be difficult to identify.

3.4.5 Social actions and social media

In order to find out whether users coming from social media are more likely to share you first have to define 'likely to share'. The clickstream logs show negligible use of the 'SAVE_SOCIAL_TAG' action. For the period June–December 2012 (the only period for which we have clickstream logs) the action occurred 189 times. Set against 9.6million accesses to object pages (FULL_RESULT_HTML) and 4.8million presentations of search results (BRIEF_RESULT:search), and 1.6 views of the homepage (INDEXPAGE) it is clear that not much sharing goes on; so insignificant that we need to look for another definition of 'social media sharing'.

If we turn to the Google Analytics 'Social Plugins', the numbers are still low, but better: September 2012–March 2013, 3,945 'Unique Social Actions'. Set against the 3.4million visits in that period a social sharing action occurs at a rate of one per 866 visits (0.12%). When that report is restricted using Europeana's own 'Social Media advanced segment' the number is reduced to 291 'Unique Social Actions'. There is indeed a greater propensity to share by visitors coming from social media: a rate of one per 146 visits (0.68%). But the actual numbers are very small, in fact of the 142 Social sharing sources used by all visitors only three —Facebook, Google+, and Twitter— appear when the report is restricted to 'social segment' referrals. One reason for this may be that the 'advanced segment' has been defined too narrowly— inputs should match outputs; all the social sites recorded by Google Analytics as 'social sources' should be included in the segment. The alternative is to restrict the Social Plugins report to match the advanced segment. In that case the 'all users' figure declines to one in 1,104 (0.09%) [the Social segment is, of course, unchanged at 0.68%]. So, users coming from social media are more likely to share. However there might be a strong element of auto-correlation here, a tautology: social media users share because that is what social media is about.

Finally, (Tables 14, 15, and 16) when social media sharing is compared with the equivalent September to March period of a year earlier —a periodisation that conveniently spans the most active seasons of use and avoids the erratic usage statistics that were observed in April–May of 2011 and the absence of the Google search indexing detected in April-May 2012— then social share actions from social media defined by the social segment declined from 704 to 291 (-59%) while the all users total showed an increase from 3414 to 3945 (+16%). This again suggests that the 'Social media traffic' advanced segment has been too narrowly defined.

Table 14. Social sharing, GA data

ALL Visits	Sep 16, 2012 - Mar 16, 2013		Sep 16, 2011 - Mar 16, 2012		
facebook : like	2060	0.4534	1720	0.4768	20%
twitter : share	500	0.1122	503	0.1326	-1%
facebook : share	346	0.0849	292	0.0732	18%
Google : +1	191	0.0424	421	0.1817	-55%
print : share	162	0.0503	103	0.0406	57%
gmail : share	76	0.0154	58	0.0161	31%
email : share	45	0.0104	33	0.0112	36%
blogger : share	33	0.0087	25	0.0062	32%
stumbleupon : share	28	0.0054	40	0.0094	-30%
favorites : share	23	0.0048	27	0.0064	-15%
+132 others					
Total (ALL users)	3945		3414		16%

Table 15. Social sharing, GA data

Social segment Visits	Sep 16, 2012 - Mar 16, 2013		Sep 16, 2011 - Mar 16, 2012		
facebook : like	228	0.7835	373	0.5298	-39%
facebook : share	21	0.0722	0	0	
Google : +1	21	0.0722	290	0.4119	-93%
twitter : share	21	0.0722	41	0.0582	-49%
Total (Social Segment)	291		704		-59%

Table 16. Social sharing, GA segmentation

All Visits compared to Visits referred from Social segment		
	ALL Users	Social Segment
facebook : like	2060	228
twitter : share	500	21
facebook : share	346	21
Google : +1	191	21
Total Social Share Actions	3097	291
Total Visits	3,417,817	42,615
Visits per Social Share	1,104	146
shares as % visits	0.09%	0.68%

3.4.6 Social media and stickiness

Table 12 & 13 refer. If we just count number of referrals then Sept 2012–March 2013 saw a top-ten with 'Facebook likes' at the top with 2060 and 'favorites: share' at number ten with 23 referrals. Factoring in 'stickiness' may change the ranking but that means we are then attaching importance to 'stickiness' in the case of some very insignificant sources.

Figure 6 has shown the number of Europeana visits over three years. From the first we can see both the pattern of growth year-on-year despite various perturbations and also the seasonal nature of usage. Figure 13 uses the same plot to show only the Social Segment. Note the very different scale of this plot, the peak value of the all users plot (April/May 2011) approaches 30,000 visits per day, for the Social Segment the peak in August 2012 is 1,100 visits per day. Compare the erratic spikes of social segment usage with the stable weekly rhythm of overall usage. Note also the lack of a significant growth pattern.

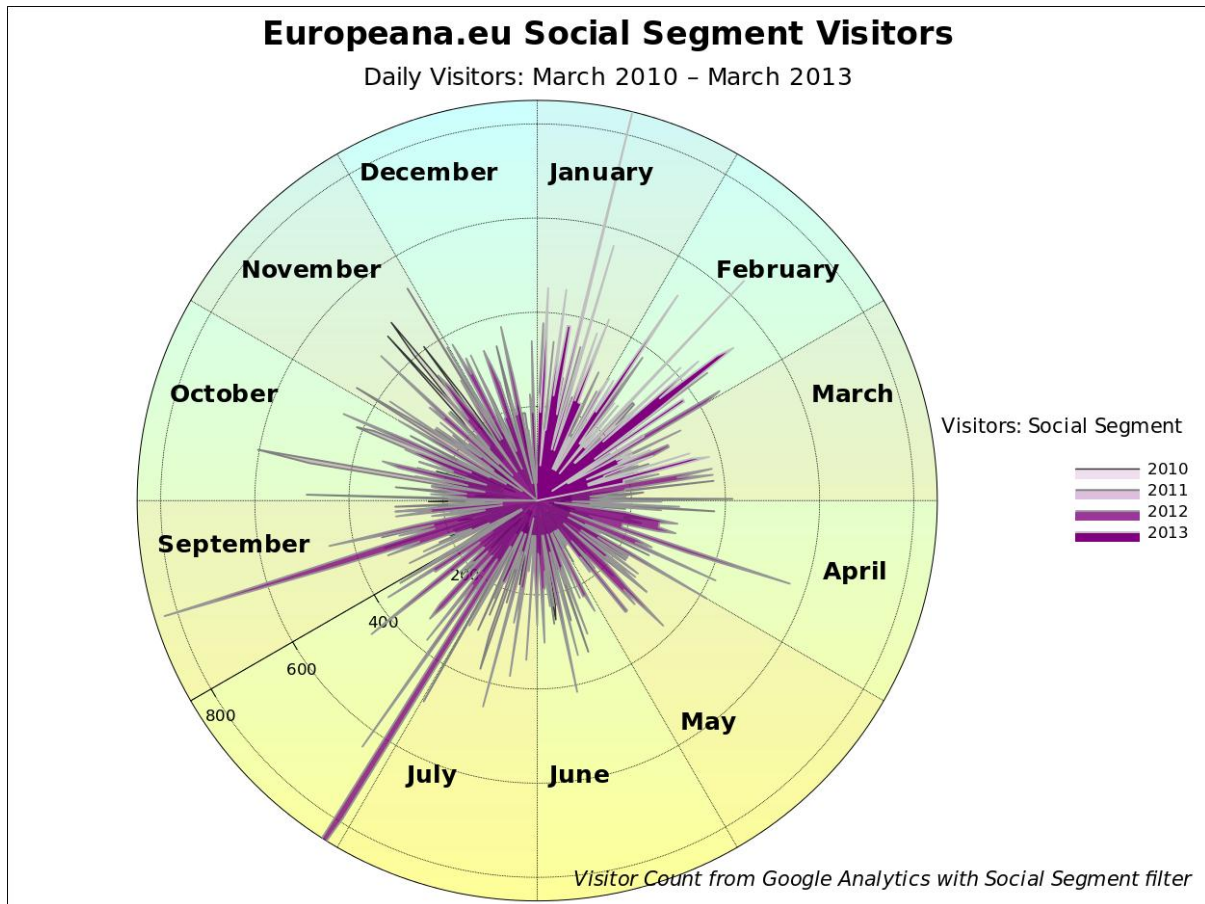


Figure 13. Social Segment Visitors 2010–2013

3.4.7 Loyalty, stickiness and quality of social media traffic

Even the most optimistic reading of the Europeana metrics we possess does not support the contention that social media generates more quality traffic (in terms of dwell time, page numbers etc.), the difference is not significant and the overall populations are low. CIBER do not have access to Europeana's social media metrics (on Facebook, Twitter etc.) and really can only comment on the Portal (+ Exhibitions and blog.europeana). However, it does appear that metrics such as the number of 'likes' are not as robust as those metrics we can derive from Europeana's own logs, and for this reason should not be equated. Thus fifteen-thousand Facebook 'likes' might mean nothing more than you have 15,000 addresses on what is effectively a mailing list. Whereas fifteen-thousand returnees or click throughs says something much more considered and pro-active.

3.5 Virtual Exhibitions

Exhibitions were only just featuring towards the end of or Europeana Connect work so CIBER came to this topic fresh and very interested. It looked like a break-through and here surely is something that would capture the interest of the digital information consumer and armchair tourist. It could 'speak' to a lot of people. Certainly given that the homepage seems to have become increasingly a promotional tool and virtual exhibitions are clearly thought to have a major role here, in promoting, highlighting and sampling Europeana, as there is a prominent carousel from which you can choose an exhibition to visit.

The amount of space allocated to comment and feedback on exhibits suggests a degree of interactivity is expected; furthermore exhibitions are by their nature places to view and browse and therefore we should expect that people spend greater amounts of time here than elsewhere on the Europeana site. Dwell time is more meaningful a metric here.

We have to rely solely on GA for this evaluation (Figure 14) as we do not have raw log files for the 'exhibitions' site. Sept-Dec, 2011 and 2012 data shows that there has been a 50% increase in visitors and 'pages per visit' has increased from 7 to 12 pages, the bounce rate is very low (0%) compared to the main site, so people appear to be dwelling; and we might have, at long last, that much sort after stickiness. About 10% of exhibition visitors appear to be using a mobile (tablet) platform, which is relatively high.

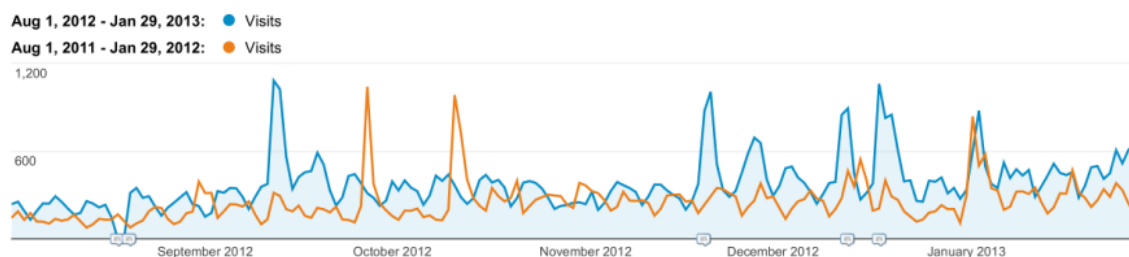


Figure 14. Exhibitions: visits

The most recent figures (Tables 14 & 15) show that the overall number of exhibition visits (less than 50,000 Sept-Dec 2012) is still relatively low relative to the visits to the main site (1.6 million). That is, just over 3% of all visitors find their way to an exhibition. But that is perhaps an unreasonable comparison; they are, after all, a relatively novel feature and fifty-thousand visits are significant when placed in contrast to the traffic flows associated with social media.

Table 17. Exhibitions, Europeana.eu

	Sep–Dec 2012	Sep–Dec 2011	
Visits:	47,078	30,826	52%
Visitors	38,573	26,245	47%
Pageviews:	550,807	210,396	162%
Pages per visit	11.7	6.8	71%
Duration of visit	00:02:44	00:02:58	-8%
Bounce rate	0%	32%	-99%
New Visits	80%	83%	-3%

Table 18. Exhibitions, visitors 30 Dec 2012 – 29 Jan 2013

	Visits	Pages/Visit	Bounce Rate
royal book collections	2,158	11.4	0.2%
1914-1918	1,435	13.1	0.3%
european sports	754	11.9	0.0%
Total visits	14,078		

Thirty-per-cent of visits to exhibitions come from the carousel on the main site homepage (11,881 visits), so homepage promotion appears to be successful. In fact nothing else is really very successful (e.g. newsletters). In contrast to the main site Search traffic is far less significant (less than a quarter) as a source of visitors. Whilst referral traffic tends to be directed to the main page, direct traffic lands on specific exhibitions, notably '1914-1918' with 6,540 visits (13% of total) September–December 2012. There is a strong flow from one exhibit (record) to another which suggests visitors are following the exhibition sequence. In conclusion: exhibitions are sticky and successful but interest (as is the nature of the exhibition trade) is volatile.

3.6 Collections and Click-Through

We have calculated the click-through rate as a ratio of page views: click-through (redirect) divided by object (record) pages. We count all object pages viewed not just those where there was a click-through. For example, if during an extended visit, the homepage, a search result, then several objects were viewed, before finally a click-through, then the rate would be calculated taking into account all those objects. This would be different from a calculation based on the observed frequency of page transits from the object page. Table 7 sets out these frequencies: this shows a click-through action as only 4% of all user actions but, if a visitor has reached a record page, there is a 20% probability that if the session continues then the next action will be a click-through (3.8% Record to Click-through divided by the 18.2% total of actions from Record pages.). This we consider to be the maximum figure by any plausible definition of click-through. By our preferred and recommended measure overall click-through is 7.32%.

In November 2012 there were over 800 datasets in the Europeana database, containing a total 20,163,672 records: 54% were images, 41% text, 2% sound, and 1% video. The supplement spread sheet lists the most popular collections (destinations) June–December 2012 as measured by the volume of click-throughs to the provider site. As noted in Table 6, click-through (redirect) is about 4% of all traffic through the site.

A direct count of Object views and redirects gives a click through figure of 7.26% overall. There is obviously great variation when we consider particular aggregators and collections and would expect even more if the analysis is taken down to particular objects. But even so the results clearly cluster at a figure of less than one in ten.

The full analysis of click through rates for each aggregator and dataset is given in the spreadsheet Europeana_CLICKTHROUGH.xlsx Tables 16 and 17 summarises these results by looking at the top ten in terms of Object views, click-through numbers and percentage rate.

Table 19. Click-Through by Dataset, Top performers from 785 datasets, Jun-Dec 2012

Dataset	Object	ClickThrough	CT rate
01004_L_DE_slub_DeutscheFotothek__fotos_ese	372,228	67,201	18.1%
01202_L_FI_SwedishLiteratureSociety_edelfeltmdtsam	867	578	66.7%
03486_L_DE_BSBMunich	289,008	22,373	7.7%
08504_Ag_EU_ATHENA_Ministry_for_Cultural_Heritage_	547,651	2,160	0.4%
08523_Ag_EU_ATHENA_Carinthian_Regional_Museum	2,205	1,033	46.8%
08535_Ag_EU_ATHENA_Institute_for_Museum_Research	365,800	37,328	10.2%
08602_Ag_EU_EFG_InstitutLuce	308,320	36,298	11.8%
09102_Ag_EU_MIMO_ESE	114,423	17,144	15.0%
09401p_Ag_RO_Elocal_fotograf	2,357	1,202	51.0%
09404_Ag_PL_ELocal.xml	297,888	18,899	6.3%
09407a_Ag_ES_ELocal_esegen	298,420	11,831	4.0%
09418e_Ag_FR_ELocal_CNTGFNDG	26	12	46.2%
09418o_Ag_FR_ELocal_MTRSCHIN	2	1	50.0%
09418p_Ag_FR_ELocal_MTRSHUCN	7	4	57.1%
09418q_Ag_FR_ELocal_MTRSJULD	16	8	50.0%
09418s_Ag_FR_ELocal_MTRSMSMA	2	1	0.5
10106_Ag_EU_STERNA_48	59,267	39,436	67%
11614_Ag_EU_OpenUp!_RBGK	112,635	20,123	17.9%
9200103_Ag_EU_TEL_Gallica_a0142	455,485	12,962	2.8%
92013_Ag_EU_TEL_a0398_TEL_NBKM_Albymika1	105	54	51.4%
92037_Ag_EU_TEL_a0444_BritishLibrary	262,752	38,414	14.6%
92106_A_NL_NatArch_database	308,310	4,953	1.6%
All Objects and Click-through	9,642,311	699,851	7.3%

Table 20. Click-through by collection country

Dataset Country	Object	ClickThrough	CT rate
EU	5,297,043	386,979	7.3%
n/a	1,210,978	112,434	9.3%
DE	792,648	96,331	12.2%
ES	423,867	15,776	3.7%
PL	297,888	18,899	6.3%
UK	300,329	8,786	2.9%
SE	206,300	6,226	3.0%
IE	156,952	3,586	2.3%
NL	147,318	4,289	2.9%
FR	128,656	11,604	9.0%
NO	114,015	1,676	1.5%
AT	78,454	2,258	2.9%
IT	67,771	2,275	3.4%
DK	57,504	2,230	3.9%
SI	46,883	4,139	8.8%
LT	42,521	3,404	8.0%
GR	42,025	955	2.3%
RO	37,740	4,785	12.7%
PT	32,130	1,899	5.9%
SK	27,920	4,119	14.8%
BG	26,179	2,600	9.9%
BE	25,776	624	2.4%
FI	20,895	1,107	5.3%
EE	18,166	343	1.9%
LU	11,052	252	2.3%
HU	9,534	80	0.8%
OTHER	6,253	268	4.3%
CH	5,213	1,168	22.4%
CY	3,633	179	4.9%
LV	2,458	251	10.2%
MT	2,358	133	5.6%
CZ	1,852	173	9.3%
All Objects and Click-through	9,642,311	699,851	7.3%

Figure 15 plots Click-through and Object Views for all datasets, aggregators, and countries. From this we can see that despite the variability of individual datasets the overall picture is consistent: more object views are matched by more click-throughs. The graphic is consistent with the 7.3% aggregate value for click-through.

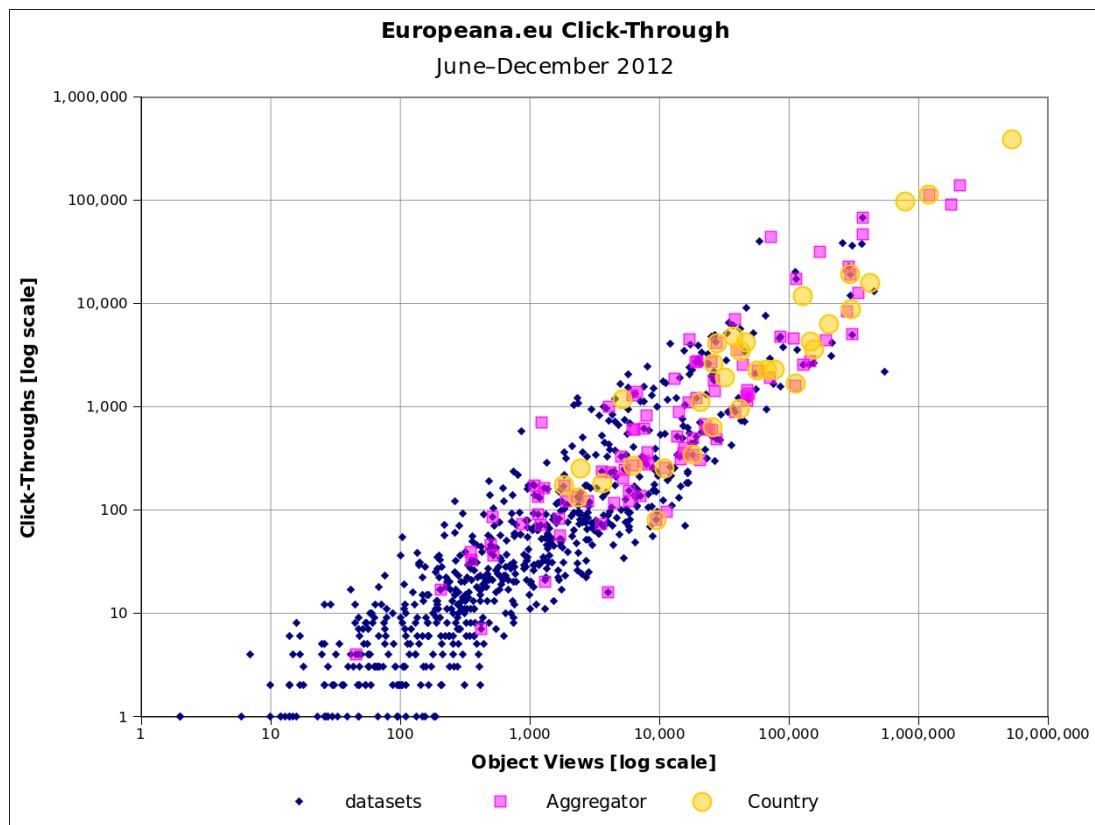


Figure 15. Europeana Click-through

Figure 16 shows the same dataset plot but in this version the datasets points are against the number of items in the dataset (collection): note how the points are far more scattered.

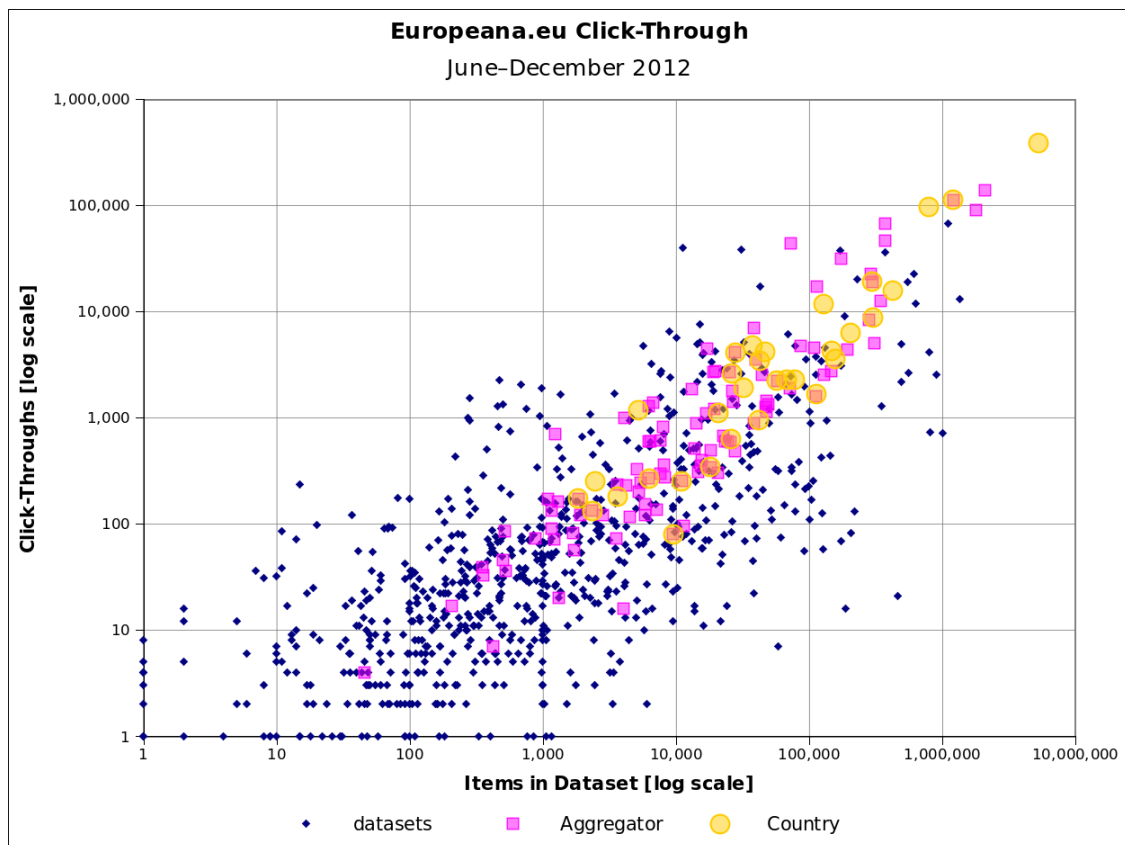


Figure 16. Europeana Click-through by dataset size 2012

There is a much stronger correlation between object views and click-through than there is between the number of objects in a dataset and click-through: click-through rises in line with object views in all segments. It does not matter if a collection is large or small, nor its provider, nor location: if it gets more object views click-through will rise in proportion. Therefore: a) improvements to the Europeana site that encouraged click-through would most likely be effective across the board (e.g. ensuring all objects have thumbnails rather than placeholders); b) from the point of view of an individual provider, improving click-through follows from greater object views (e.g. if the object is more popular it will have more click-through). We know from earlier work that multimedia content attracts more attention than text, but if we factor that out are Europeana users brand aware? Do they prefer content from particular providers? We don't yet know.

3.7 Digital visibility

Click-through is a measure of the outward flow, but what can we say about the inbound side? How much of Europeana's content is visible, how much is used? Clearly Google, which accounts for two-thirds of all inbound traffic, demands attention. The next largest category is unknown: the referrer of a quarter of inbound traffic has either been withheld or, in order of

decreasing probability was followed from a non-browser source (e.g. a link in a document or email) or was typed in.

For digital visibility purposes, we need to consider how effectively the site is crawled and indexed by search engines, in practice that means the Googlebot. We can then compare that with the content that is accessed by users.

In November 2012 the Googlebot indexed 12,073,076 records (Figure 17). Since June 2012 it appears to have crawled, on average, about 50% of the site each month, with some content being crawled every month. It also appears that some content has been removed or relocated since the total of unique document IDs located by Google is, at 26M, greater than the total document collection. Hardly anything was crawled by Google between April–May 2012 and content indexed before that date appears not to have been revisited.

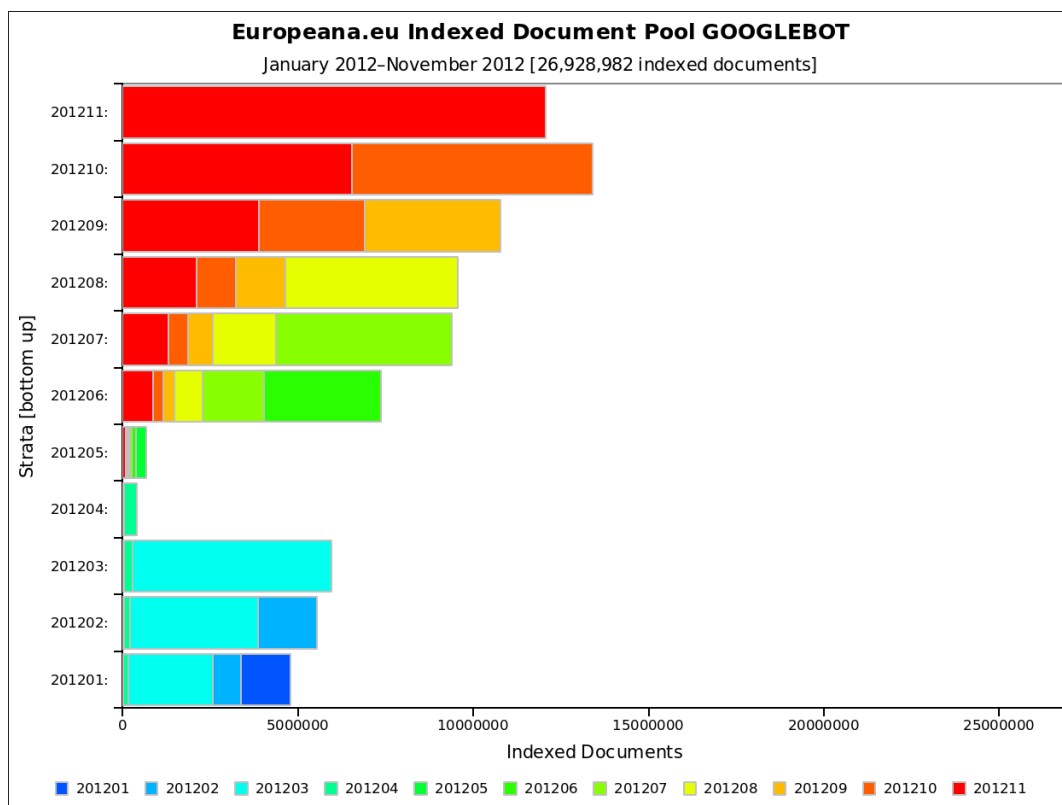


Figure 17. Europeana: Indexed documents

We can compare this picture of the site indexing by Google with its corollary, a picture of how much of the content was consulted by real users (Figure 18).

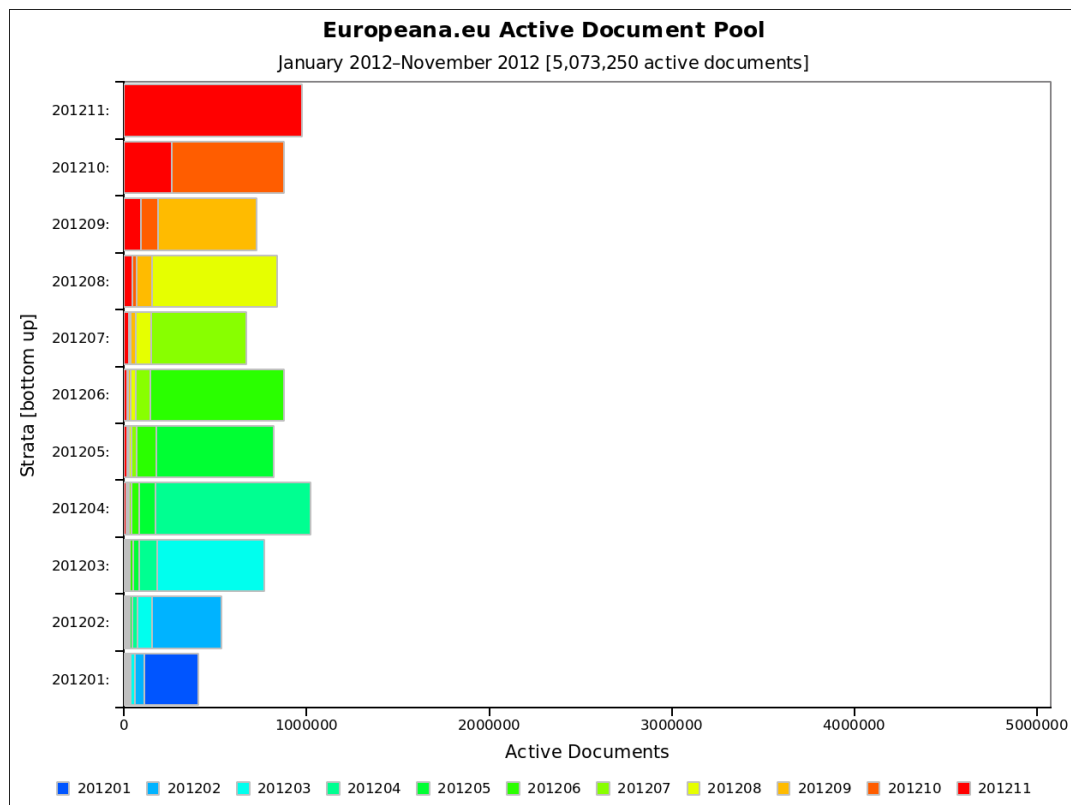


Figure 18. Europeana: active documents

The 'pool' of active documents —at least one page-view of a 'record' by at least one user over the period January–November 2012— is 5,073,250. That is about one-quarter of all documents currently in Europeana. In the most recent month for which we have a complete record, November 2012, just under a million (973,927) records were accessed by users. It can also be shown that each month 70–80% of material consulted is 'new' in that it has not been accessed in the immediately preceding month. Thus although much of Europeana's riches remain unexplored, and some heavily featured items will attract a lot of hits from many users, the use of the collection is not confined to a small core of 'bestsellers'.

3.8 API

CIBER obtained a series of API logs from January 2012, which are in the clickstream format, but they contain very little information that can be analysed. Looking at Figure 2 it can be seen that this API log-series appears to show declining use. Thus it would seem that API usage is effectively main-site use that originates with a referral from one of the designated API domains. It is possible to analyse the click-stream logs for this segment, but we need to start with a clear definition of what constitutes API usage.

3.9 Overview and update

The following Table 18 provides an overview of the key usage metrics and also breaks down the data by visit types (Direct, Search Engine, Social Media, API, and Referral). Note in Table 17 the significantly above average page-per-visit number for direct visits. To a lesser extent this also applies to Social and API visitors. However API visitors are less likely to be new visitors and have longer visits. Given that both Social and API visits are a subset of Referrals both visit more pages and stay longer than the norm for referrals, Social visitors are however typical of other referral in rate of new visits and bouncers.

Table 21. Overview based on Google Analytics data for most recent six months (16 September 2012– 16 March 2013).

Aspect of visit	ALL	Direct	Search	Social	API	Referral
Visits	3,417,817	349,232	2,455,144	42,615	45,596	594,718
Page views	13,216,262	2,060,315	8,302,613	233,601	256,153	2,773,933
Unique Visitors	2,650,851	286,309	1,943,864	35,421	32,029	496,155
Pages per Visit	3.9	5.9	3.4	5.5	5.6	4.7
Average visit (minutes).	2.3	3.6	2.0	3.2	4.8	2.8
New Visits 75%	75%	79%	75%	76%	59%	77%
Bouncers 49%	49%	49%	51%	43%	31%	42%
Country						
Germany	14%	13%	14%	7%	4%	12%
France	9%	8%	10%	6%		8%
Spain	7%	6%	7%	9%	9%	9%
Italy	7%	8%	7%	5%	1%	5%
Netherlands	7%	8%	7%	6%	3%	5%
Poland	6%	4%	5%	4%	51%	8%
USA	5%	9%	4%	7%	3%	8%
Great Britain	4%	4%	5%	4%	3%	3%
Belgium	3%	3%	3%			2%
Sweden	3%		3%			
Eire					13%	
Ukraine					3%	
Portugal				6%		
Norway				4%		
Hungary						4%
Per cent						
% of Visits	100%	10%	72%	1%	1%	17%
% of Page views	100%	16%	63%	2%	2%	21%
% of Visitors	100%	11%	73%	1%	1%	19%

Long term overall growth: a note. In calculating overall site growth we have combined the various measures of pageview and visitor numbers from http-access logs, clickstream and Google Analytics. This enables a long-run of data from October 2010 that closes the various lacunae and perturbations of the individual series. From this we can calculate the slope of a linear regression line; that is a smoothed 'average' of the daily values. Projecting this trend of past performance forward we predict on-trend growth in 2013 of about 25%.

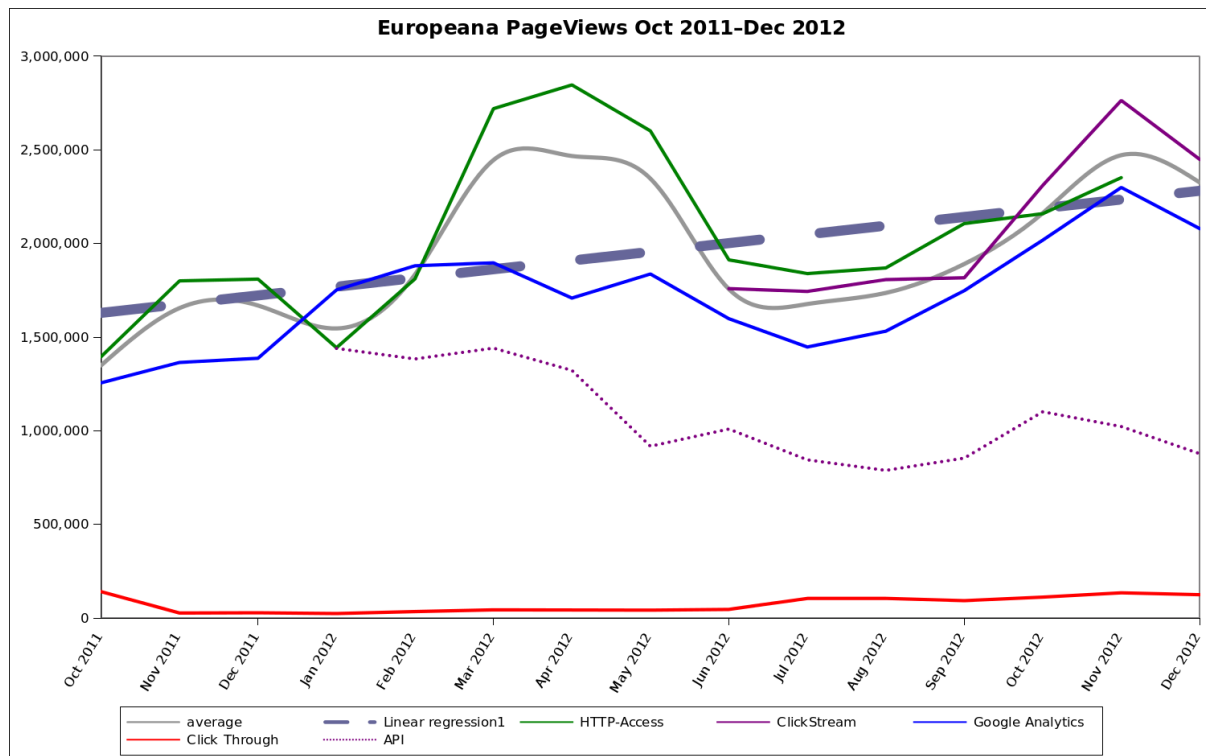


Figure 19. Europeana_Pageview_trend