

Information Seeking Behaviour and Usage on a Multi-media Platform: Case Study Europeana

David Nicholas and David Clark

Abstract The article examines the usage of the Europeana.eu website, a multi-lingual online collection of millions of digitized items from European museums, libraries, archives and multi-media collections. It concentrates on three aspects of use: (1) stickiness and site loyalty; (2) social media referrals; (3) that associated with virtual exhibitions. Three methods for obtaining the data are examined: Google analytics, ClickStream logs and http server logs. The analyses produced by Google Analytics are highlighted in the article.

Keywords Information seeking · Google analytics · Log analysis · Multimedia website · Europeana

David Nicholas is a Director and founder of the CIBER research group (<http://ciber-research.eu>). The group is perhaps best known for monitoring behaviours in the virtual space, especially in regard to the virtual scholar and the Google Generation, which has been featured widely in the media, including on BBC TV and Australian TV (ABC). David is also a professor at the College of Communication and Information Studies, University of Tennessee and at the iSchool at Northumbria University. Previously David was Head of the Department of Information Studies at University College London for seven years (2004–2011) and previous to that was Head of the Department of Information Science at City University for a similar period of time. Email: Dave.Nicholas@ciber-research.eu.

David Clark has forty years experience of data processing and information management in publishing and related industries. He returned to full-time education in 1995 to take a masters degree in knowledge engineering. He has a PhD in Computer Science, from the University of Warwick, where he taught and researched from 1997–2001. He then moved to University College London, leaving in 2011 to form CIBER Research Ltd. He is a Director of the CIBER Research Ltd.

D. Nicholas (✉)

College of Communication and Information Studies, University of Tennessee, Knoxville, USA
e-mail: Dave.Nicholas@ciber-research.eu

D. Clark · D. Nicholas

CIBER Research Ltd., Newbury, UK
e-mail: David.Clark@ciber-research.eu

Background

Transactional log studies are becoming more and more popular ways of evaluating websites (Davis 2004; Davis and Solla 2003). In particular, Http server logs and clickstream logs have proven to be very effective ways of analysing behavior on textual websites (Nicholas et al. 2006). The primary goal of many log or usage data studies is to find out about use rather than users. In terms of usage studies, previous log studies have led to different conclusions about the success or otherwise of the Big Deal and consortium subscriptions to journals. Davis (Davis 2002) challenged the composition of geographic based consortia. He recommended libraries create consortia based on homogeneous membership. On the other hand, Gargiulo (2003) analysed logs of an Italian consortium and strongly recommended Big Deal subscriptions. Essentially one of the limitations of basic log analysis is the fact that there is little possibility to link use data with user data, hence a vague and general picture of users' information seeking behaviour is obtained. This technical restriction makes it difficult to use the demographic data of users for finding out about differences in information seeking activities of users in regard to different tasks, statuses, genders and so on. In this paper, Google analytics data will be investigated to explore new ways of evaluating behavior on the web.

Europeana, launched in 2008 as a prototype and operating as a full service since 2010, is a gateway, portal or search engine to the digital resources of Europe's museums, art galleries, libraries, archives and audio-visual collections (Fig. 1). Europeana is regarded as trusted (curated) source connecting users directly to authentic and curated material. It provides multilingual access to 26 million European cultural objects in 2200 institutions from 34 countries. Books and manuscripts, photos and paintings, television and film, sculpture and crafts, diaries and maps, sheet music and recordings, they're all there. Europeana claim that there is no longer the need to travel the continent, either physically or virtually. If you find what you like you can download it, print it, use it, save it, or share it¹.

While Europeana is essentially a portal it also has aspirations well beyond that; it believes it can help stimulate the European digital economy; it also mounts online exhibitions and takes part in crowd sourcing experiments (World War 1 is currently the subject of such an experiment). Europeana is also working with other digital channels to distribute their content, most notably Google, Wikipedia and Facebook.

It is a site that currently attracts around five million visitors and is used heavily by humanities scholars, heritage professionals and even tourists. 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 we use logging as the basis of insight and prediction about the purpose and motive of the millions who use Europeana.

¹ <http://www.europeana.eu/>

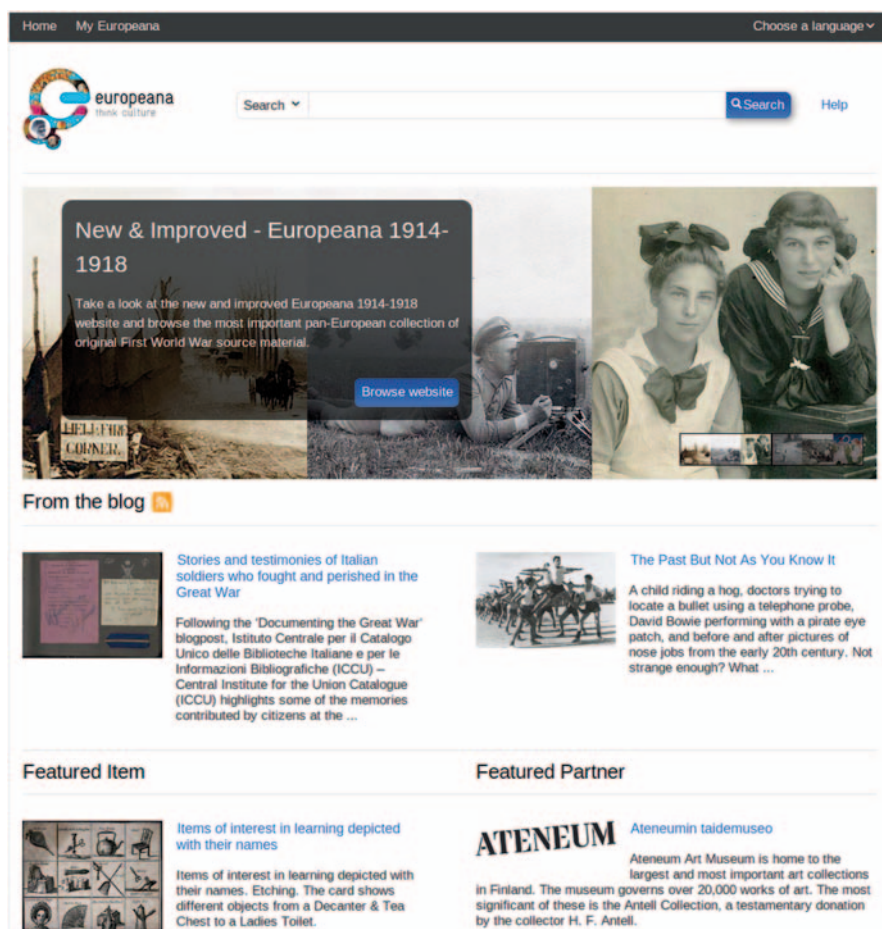


Fig. 1 Europeana home page

Aims and Objectives

The study reported here features our latest research which focuses on three types of digital behavior prevailing in Europeana which we regard to be particularly significant and strategic forms of digital behaviour not only for Europeana, but for all information providers on the Web. These are:

1. **Stickiness and user loyalty.** *Stickiness* is anything about a Web site that encourages a visitor to stay longer (engagement) and visit more often (returnees). All information providers are interested in what constitutes stickiness and how they can make their sites stickier.
2. **Social media referral.** Volume and characteristics of the traffic coming from Facebook, twitter and the like, which could potentially drive a lot of traffic to Europeana and encourage much re-use of Europeana content.

3. **Virtual exhibition usage.** Virtual exhibitions are a recent innovation in which Europeana sets much store. Clearly these exhibitions provide a lot of added value for a site which essentially functions as a search engine. Exhibitions could capture the interest of the digital information consumer and armchair tourist. They could ‘speak’ to a lot of people.

A prime objective of the study is to see what Google Analytics could provide in regard to robust and precise usage data and how it compared with our traditional usage sources—http server logs and ClickStream logs.

While textual websites, like those of scholarly publishers and libraries, have been well researched from a digital usage point of view (Julien and Williamson 2010), very few studies of multi-media platforms have been undertaken, and so in the way the paper is quite unique.

Methodology

For CIBER’s earlier Europeana work we relied upon server http request logs using CIBER’s own ‘deep log’ methods (Nicholas et al. 2013). However, for the study reported here we wanted also to see what extent, the now ubiquitous, Google Analytics (GA) data could undertake key information seeking analyses more cheaply and effectively. This is important given the fact that Europeana, like many organisations, are relying increasingly heavily on GA for all their usage and marketing needs. While we have utilised GA heavily in this paper as we will learn GA cannot always supply the data required in a convenient form and have thus supplemented it with our own tried and trusted deep log 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, and this is generally not undertaken by institutions and analysts.

In addition, to the http logs and GA data, we also had access to a series of ClickStream logs which had become recently available. However, we only had access to the ClickStream 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 sometimes used the old series of raw http-request logs for a broader overview and perspective.

Thus, to provide the best and most comprehensive analysis of Europeana 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 informa-



Fig. 2 Taking the pulse of a website using logs

tion that would not otherwise be collected because it did not seem relevant at the relevant at the time.

Figure 2 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.

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, as we shall learn in greater detail, 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 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.

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 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: 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.

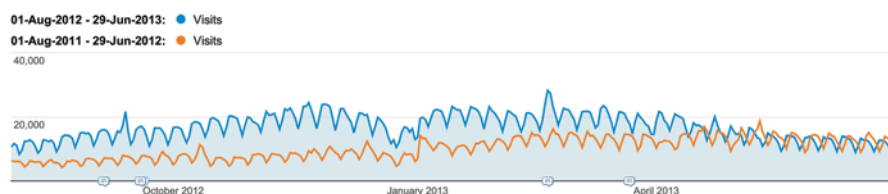


Fig. 3 Visits: August 2011–January 2013 compared to the same time the previous year. (Source: GA)

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.

Results

Stickiness and Loyalty

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.

First, let us provide the necessary general usage data as a context to the stickiness investigation. How is overall usage going over time and what patterns can we see? Comparing the more regular and settled periods: autumn 2011 (Aug–Jan) with autumn 2012 (Aug–Jan) (Fig. 3) a clear picture emerges with visitor numbers growing healthily 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, so it appears to be accelerating.

Figure 4 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.

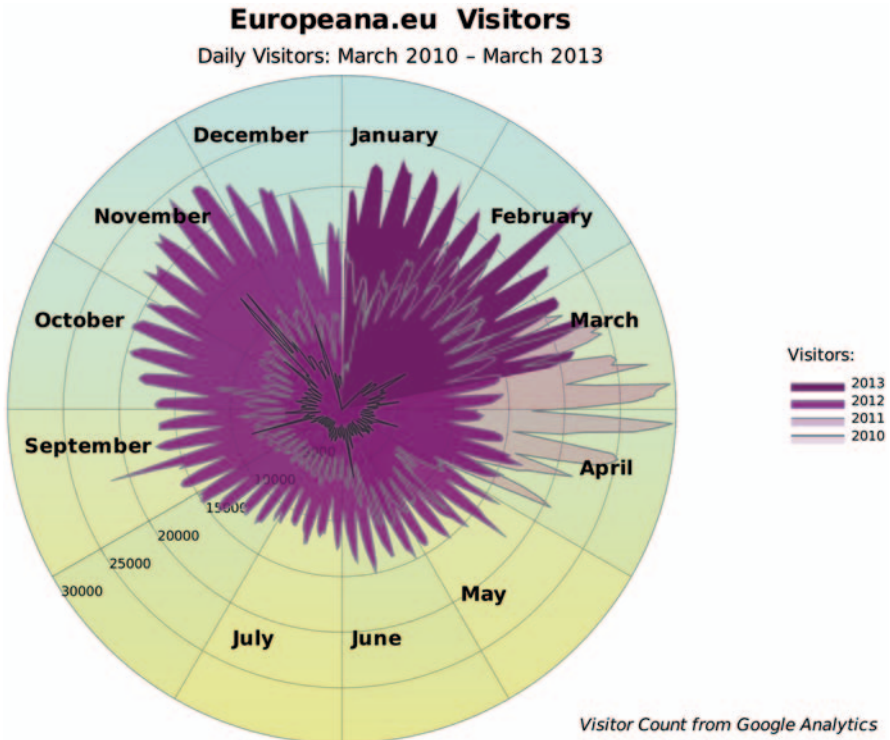


Fig. 4 Europeana daily visits 2010–2013

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 of this and given the relative importance of this metric, far more meaningful than a Facebook ‘like’, for instance, Europeana hopes to do more research to establish its real significance, by triangulating the data with demographic, survey or qualitative data.

Only one in four visitors return to Europeana, 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 visi-

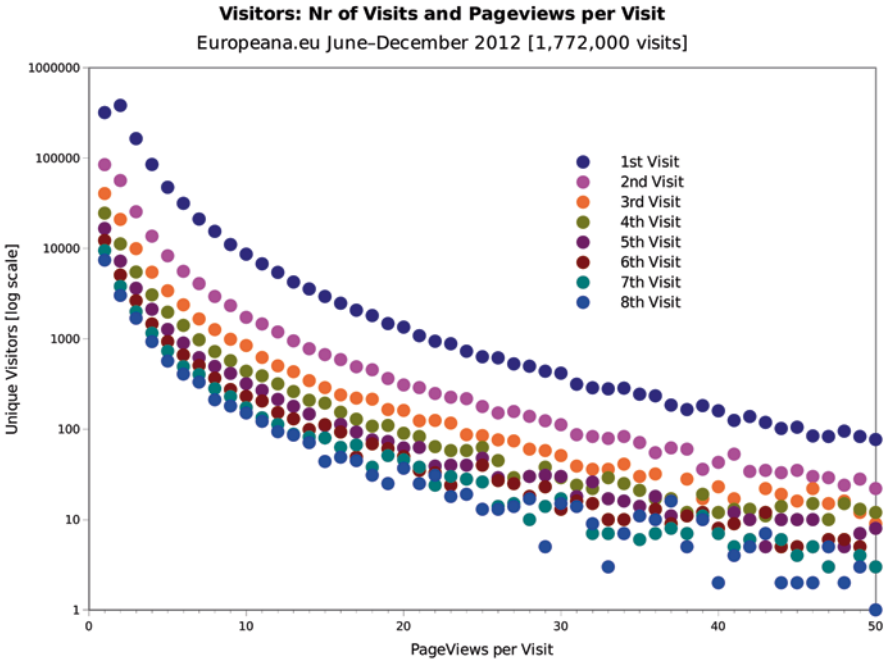


Fig. 5 Europeana: return visits. (Source: GA)

tors returned five times or more. GA may understate the return rate a little but the distribution follows a typical ‘power law’ (Fig. 5). 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.

When looking at returning visitors it is well to remember that most visits are very fleeting; even when bouncers (one visit, on view) 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 (Figs. 6–8) 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 (Fig. 6).

When we look even closer (Fig. 8), 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 h. In fact, equipped with this insight, we can turn to a daily plot (Fig. 8) and see the same

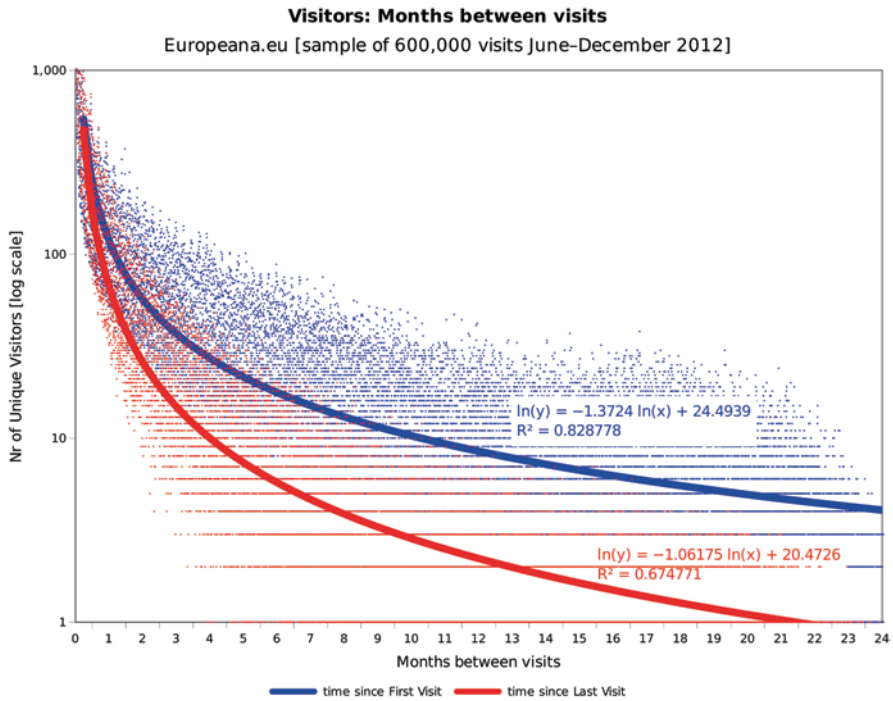


Fig. 6 Europeana: months between visits. (Source: GA)

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.

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.

Engagement

We can calculate levels of 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 (<10 s); and less than 2% are recorded by GA as exceeding 30 min (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 come to the virtual exhibition section we can see people dwelling longer and examining more pages.

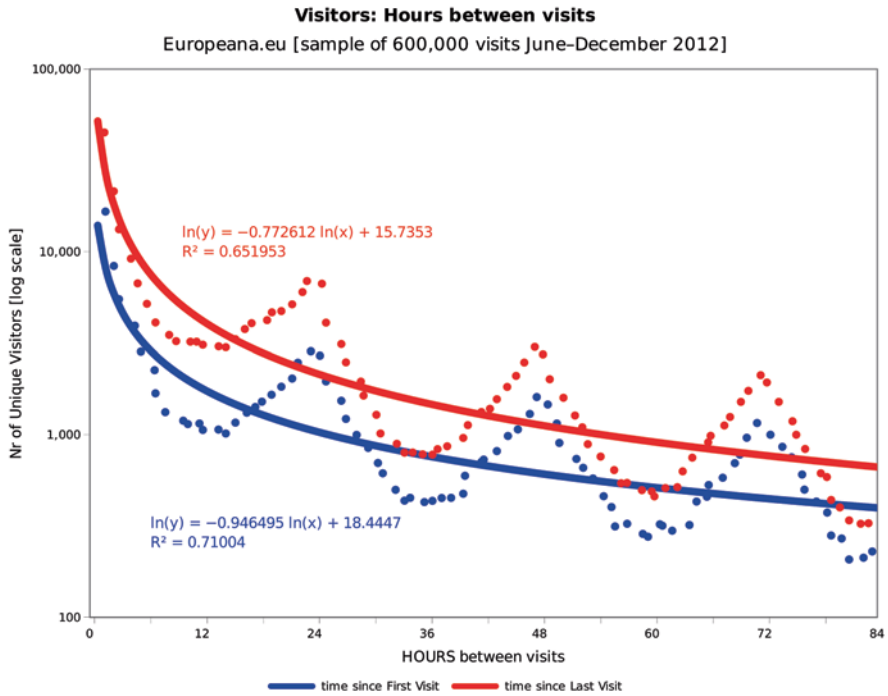


Fig. 7 Hours between visits. (Source: GA)

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. Table 1 and Fig. 9 and show visit times for December 2012. The average visit duration is 2 min and 19 s, 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 58% of visits were timed at less than 10 s; only 10% of visits fall broadly (1–3 min) into the ‘average’ category band.

The story on ‘engagement’ is an interesting one: usage (page views) has 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 s, to 2:18 s. 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

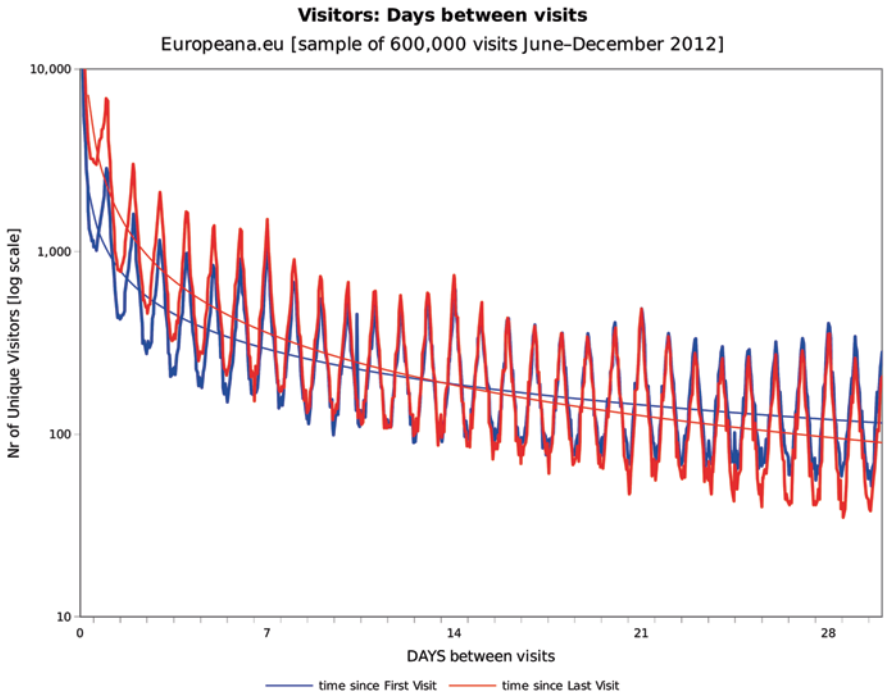


Fig. 8 Days between visits. (Source: GA)

Table 1 Duration of visits, December 2012. (Source: GA)

Visit duration (s)	Visits
0–10	311,193
11–30	57,548
31–60	39,689
61–180	52,233
181–600	39,245
601–1800	23,132
1801+	7,707

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. 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.

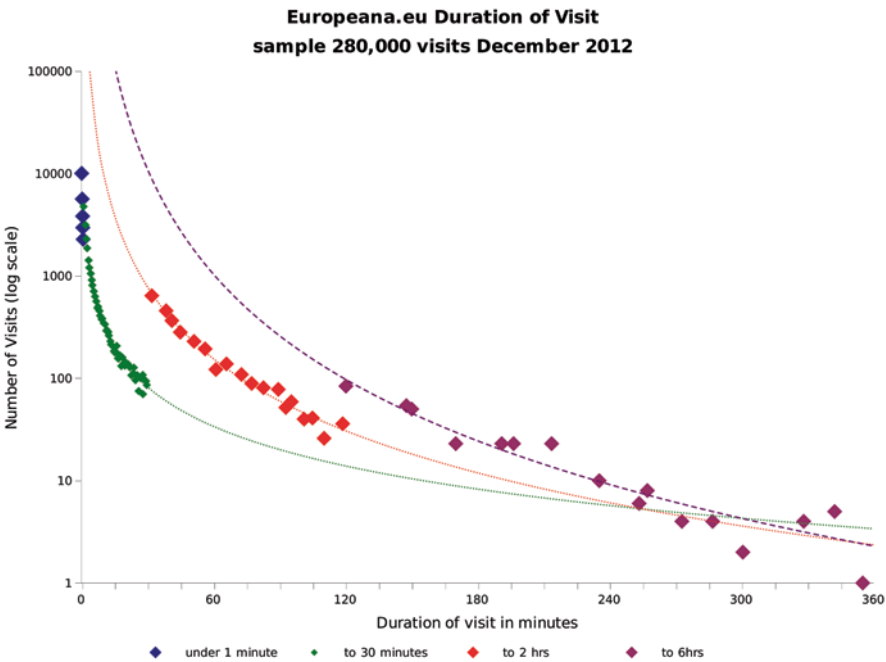


Fig. 9 Europeana: duration of visit. (Source: GA)

Social Media

With so much Europeana (and scholarly publisher) planning (and hopes) resting on social media use for growth and re-use 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.

The Google Analytics ‘advanced segment’ for social media, as personally 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) See Table 2. 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, the latter being publicised on the Europeana homepage for many weeks during 2013. We shall return to individual performance later in this section, here we shall confine ourselves to the problems of definition.

An interesting definitional case is Twitter. 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 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

Table 2 Social segment definition (GA)

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

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. The problem can be fixed by replacing the rule “include Source containing ‘t.co’” with “include Source Exactly matching ‘t.co’” or with “include Source Matching RegEx $^t.co$ ”.

Blogs pose definitional problems too (Table 3). 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.

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 2012/2013 includes 78 etc. The definition is as long as a piece of string and makes social network behaviour very difficult to delineate.

Table 3 Social segment blogs (selection only). (Source: GA)

	Oct 1, 2011–Mar 31, 2012	Oct 1, 2012–Mar 31, 2013
agioritikesmnames.pblogs.gr	64	0
bazoga.over-blog.com	43	0
bgpw.blog.pl	21	21
blog.bnf.fr	21	21
blog.crdp-versailles.fr	43	0
blog.daum.net	21	43
blog.euscreen.eu	0	64
blogs.ec.europa.eu	21	21
blogs.helsinki.fi	21	0
blog.sina.com.cn	0	64
blogs.law.harvard.edu	64	0
blog.slub-dresden.de	21	0
blogs.sch.gr	43	43
boqo.over-blog.com	21	0
cblog.culture.fr	43	0
christypato.blog.br	21	0
deal.blog.kazeo.com	0	128
deal.blog.mongenie.com	213	0
digiblog.hu	21	0
enfinlivre.blog.lemonde.fr	43	0
estudamais8.blogs.sapo.pt	21	0
eueublog.wordpress.com	0	64
fablog.iransalamat.com	43	0
formacion.universiablogs.net	43	0
googleblog.blogspot.com	21	0
kluwercopyrightblog.com	21	0
konzervativci.blog.com.mk	21	0
leblog-ffg.over-blog.org	43	21
libblog.ucy.ac.cy	21	0
pisani.blog.lemonde.fr	43	43
sog.blog.so-net.ne.jp	0	21
somewhereinblog.net	43	43

To conclude there are three ‘social’ definitions at work here: Google’s, Europeana’s social segment and CIBER’s own expanded version based on a corrected version of the Europeana social segment and used for the following analyses.

Size and Growth in Traffic

To place social media referrals in context it is worth first looking at all referrals. Seventy per cent of the 4.5 M visits to Europeana in the past year (2012) 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 1 %.

Google Analytics was not reporting social referral before Oct 2011, so there is 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 a new portal launch in October 2011 (thanks to associated publicity one presumes), 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. Between Oct–March 2011/2012 and 2012/2013 the overall year-on-year general visitor growth is 90 %. However if we look at the ‘social segment’ the visitor growth is 34 %. 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 significant for blog.europeana.

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 bypasses Europeana.eu. 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 website 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, on the social media sites themselves.

Individual Social Media

The dominant social media network is Facebook with nearly 30,000 referrals in the year since the new portal launch (Oct 2011). The ‘average visit duration’ of these Facebook sourced visitors is, according to Google Analytics just over 3 min. 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 min 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 min in autumn 2011, 5 min 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 min) 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 with the flow of social media visits using Google Analytics. 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’. 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’, a personal/customised facility, 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.

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.

Country Analysis

First a note of caution: determining the user's 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.

Social Actions and Social Media

In order to find out whether users coming from social media are more likely to share (thought to be a positive by information providers) 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 9.6 million accesses to object pages (FULL_RESULT_HTML) and 4.8 million 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 equivalent, 'Social Plugins', the numbers are still low, but better: September 2012–March 2013, 3,945 'Unique Social Actions'. Set against the 3.4 million 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

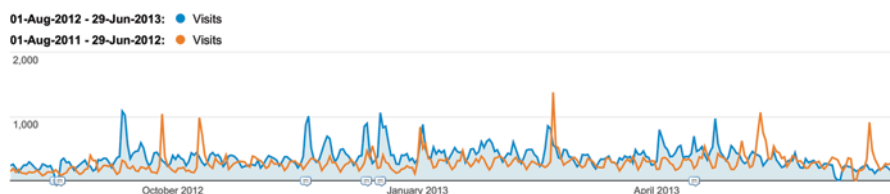


Fig. 10 Exhibitions: visits. (Source: GA)

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.

Virtual Exhibitions

Exhibitions were only just featuring towards the end of or Europeana Connect work in 2011 so CIBER came to this topic fresh and very interested in looking at the impact it has had. It looked like a break-through in Europeana thinking and here surely is something that could capture the interest of the digital information consumer and armchair tourist, strategic markets for Europeana. It could ‘speak’ to a lot of people. Certainly so as 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; 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 a more meaningful a metric here.

We have to rely solely on GA for this evaluation (Fig. 10) 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 also relatively high.

The most recent figures (Tables 4 and 5) show that the overall number of exhibition visits (less than 50,000 Sept–Dec 2012) is still relatively low relative compared 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

Table 4 Exhibitions, Europeana.eu. (Source: GA)

	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 5 Exhibitions, visitors 30 Dec 2012–29 Jan 2013. (Source: GA)

	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		

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.

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.

Conclusions

For the very first time CIBER has 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 three methodologies have been employed in regard to usage of a single website. We were especially interested to find out whether Google Analytics’ popularity is matched by its capabilities and this article produces many useful GA derived analyses. GA proved to be a very useful usage tool, albeit one which sometimes underestimates usage, and also one which needs careful calibration and interpretation to obtain full benefits.

Of course, using multiple sources of data has a downside as it highlights differences and divergences which need to be resolved. Considerable effort has gone into ironing 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.

In respect to the results of the analyses:

- a. 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. In regard to engagement: most visits are over in the blink of an eye (10 s), 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.
- b. Social Media: taking Europeana's definition ('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 %. 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, there is a need for 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.
- c. Virtual exhibitions are an undoubted and a qualified success, which seem highly fit for purpose: for viewing rather than reading. They are popular, sticky, and generate high levels of engagement. They are the elephant in the room.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Davis PM (2002) Patterns in electronic journal usage: challenging the composition of geographic consortia. *Coll Res Libr* 63(6):484–497
- Davis PM (2004) Information-seeking behavior of chemists: a transaction log analysis of referral URLs. *J Am Soc Inf Sci Technol* 55(4):326–332
- Davis PM, Solla LR (2003) An IP-level analysis of usage statistics for electronic journals in chemistry: making inferences about user behavior. *J Am Soc Inf Sci Technol* 54(11):1062–1068
- Gargiulo P (2003) Electronic journals and users: the CIBER experience in Italy. *J Ser Community* 16(3):293–298

- Julien H, Williamson K (2010) Discourse and practice in information literacy and information seeking: gaps and opportunities. *Inf Res Int Electro J* 15(1):n1
- Nicholas D, Clark D, Rowlands I, Jamali HR (2013) Information on the go: a case study of European mobile users. *J Am Soc Inf Sci Technol* 64(7):1311–1322
- Nicholas D, Huntington P, Jamali HR, Watkinson A (2006) The information seeking behaviour of the users of digital scholarly journals. *Inf Process Manag* 42(5):1345–1365