

FEC Legal Resources Analytics

Preliminary Read-out

Updated, 1/05/2015 with [search analysis](#)
[Draft]

This preliminary read-out explores the existing FEC.gov site analytics, specifically around legal resources, to get a glimpse of existing user behavior and help brainstorm research questions we want to answer during the research with site users. This draft relies heavily on Google Analytics for the site, but also examines Foresee survey results and incorporates search logs to get clues at how the site is being used. These hints will help drive hypothesis generation and lead to better research questions for the second phase of the legal research when we engage, observe and individuals using these resources.

Important Limitations & Notes:

Existing site statistics are an imperfect but useful gauge of what content is most interesting, and users flows through the site. The usefulness of the analytics is limited by some important factors:

- Google analytics does not track all portions of the site equally well. For instance, the site search and other database driven applications on the site are not always well tracked. Importantly, data about advisory opinions (AO's) are not available in Analytics; it appears that even the AO home page is not included in our analytics account.
- Google analytics is notoriously difficult to use well (in part due to it's relatively complex interface), and this read-out writer's prowess at pulling insight out of it is to some degree limited.
- For the purposes of this report, all analytics data is based on a 13 month sample starting in November 2014 and running until December 2015, unless otherwise noted.
- The analysis is limited to the following site sections:
 - /law
 - /feca (statutes)
 - /cfr (regulations)
 - /em (enforcement matters)
 - /audit
 - /policy
 - /litigation (court cases)

- The read-out is currently limited by not yet having completed an analysis of the site server search logs.

Findings:

The findings below are preliminary, and presented without an overarching narrative. With each, a possible research implication is included. These findings, and this document as a whole is meant as the start of an exploration and discussion; while we need not (and indeed, can not) completely understand why the analytics look the way they do, they can help us hypothesize about some ways the site is being used at current, and may help us to identify patterns to watch out for when we engage with users.

1. Roughly, legal research represents 8% of all site traffic, attorneys represent 4% of user base

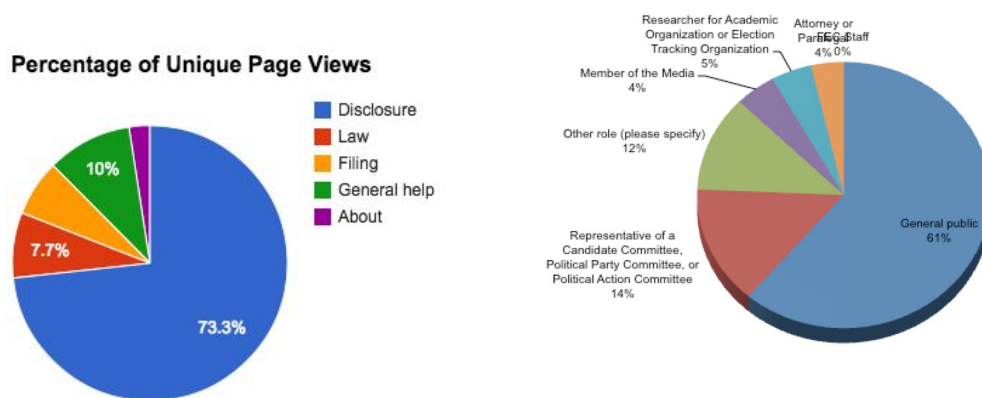


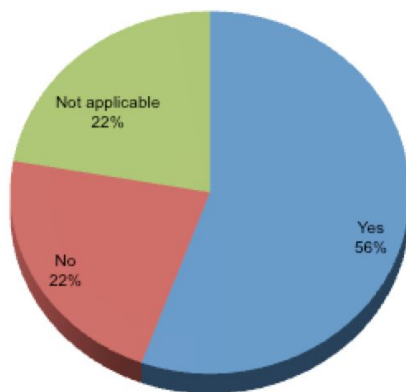
fig. 1 & 2. unique page views and traffic analysis source: analytics notes doc, march; foresee survey results; "what was your role in visiting the site", october (attorneys in orange)

While the legal resources are by no means the largest percentage of the traffic or site content, they represent a small but important segment and area of importance, as they represent a kind of power user.

2. Users of legal resources are somewhat mixed on their usefulness.

While most users agreed that the site provided them with sufficient information to interpret the legal and compliance information (56% yes, 22% no), they were almost evenly split on whether the legal and compliance information was provided in a manner that was easy to use (34% yes, 33% no), and only somewhat more agreed that it was provided in a manner that was easy to comprehend (45% yes, 33% no). These findings are limited (only 9 users responded) but they provide good starting points for deeper

Did the site provide enough information to help you interpret the legal and compliance information?
FEC Web Browse
October 01, 2015 - October 31, 2015
N: 9



research exploration.

fig. 3, 4, 5: Foresee October report

3. The most often accessed legal sections are litigation and statutes


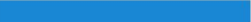




























1.	 /law.shtml	61,665	 27.51%
2.	 /litigation/	44,462	 19.84%
3.	 /feca/	29,472	 13.15%
4.	 /cfr/	16,492	 7.36%
5.	 /litigation_CCA_B.shtml	14,797	 6.60%
6.	 /litigation_CCA_C.shtml	10,027	 4.47%
7.	 /recentdevelopments.shtml	8,179	 3.65%
8.	 /procedural_materials.shtml	5,804	 2.59%
9.	 /policy.shtml	4,073	 1.82%
10.	 /litigation_CCA_Alpha.shtml	4,056	 1.81%
11.	 /litigationrecent.shtml	3,486	 1.56%
12.	 /litigation.shtml	3,399	 1.52%
13.	 /litigationmajor.shtml	2,685	 1.20%
14.	 /lobbybundlingfaq.shtml	2,463	 1.10%
15.	 /litigation_CCA_M.shtml	2,444	 1.09%

fig. 6 unique pageviews under "law" section receiving over 1% of the traffic there

1.	 /em.shtml	7,952	 52.23%
2.	 /mur.shtml	4,532	 29.77%
3.	 /adr.shtml	1,452	 9.54%
4.	 /enfpro/	1,289	 8.47%

fig. 7: unique pageviews under "em" section receiving over 1% of the traffic there

















1.	 /audit_reports.shtml	1,860	 19.00%
2.	 /2008/	1,803	 18.41%
3.	 /2012/	1,708	 17.44%
4.	 /2010/	1,624	 16.58%
5.	 /audit_reports_unauth.shtml	1,402	 14.32%
6.	 /audit_reports_pres.shtml	583	 5.95%
7.	 /audit_reports_auth.shtml	565	 5.77%
8.	 /2006/	120	 1.23%

fig. 8: unique pageviews under “audit” section receiving over 1% of the traffic there

sources: [content drilldown](#), law; [content drilldown, em](#); [content drilldown, audits](#)

A quick ranking shows that litigation is by far the biggest traffic driver, followed by the statutes and the code of regulations, then by the matters under review, and policy statements, with audits receiving the least. Notably, because the paths are different, the bars are not perfectly comparable between path sections. (the raw number of unique pageviews are a better gauge).

This implies that while gaining a holistic view of how each of the kinds of content fits together in the user’s mental model and should be given some degree of emphasis in the research, understanding how users engage with the litigation section (and other popular sections) of the site should perhaps be given more immediate attention.

3. Users often visit the Legal Home multiple times per session



fig. 9: page view overview, Google Analytics

The legal home’s pageviews are almost 4 times the rate of it’s unique pageviews. Because pageviews include every time a user loads the page in one session (whereas unique pageviews do not), this indicates that an individual user is hitting this page multiple times during a session. This could indicate that users are referencing multiple items from this

legal home page, or it could also indicate that users are browsing for the desired content from this page, but not finding it. In order to find out which is more common, our research will have legal resource users demonstrate how they use the site at current, with specific emphasis on how they use the resources linked from this page.

3. Users often go back and forth between resources

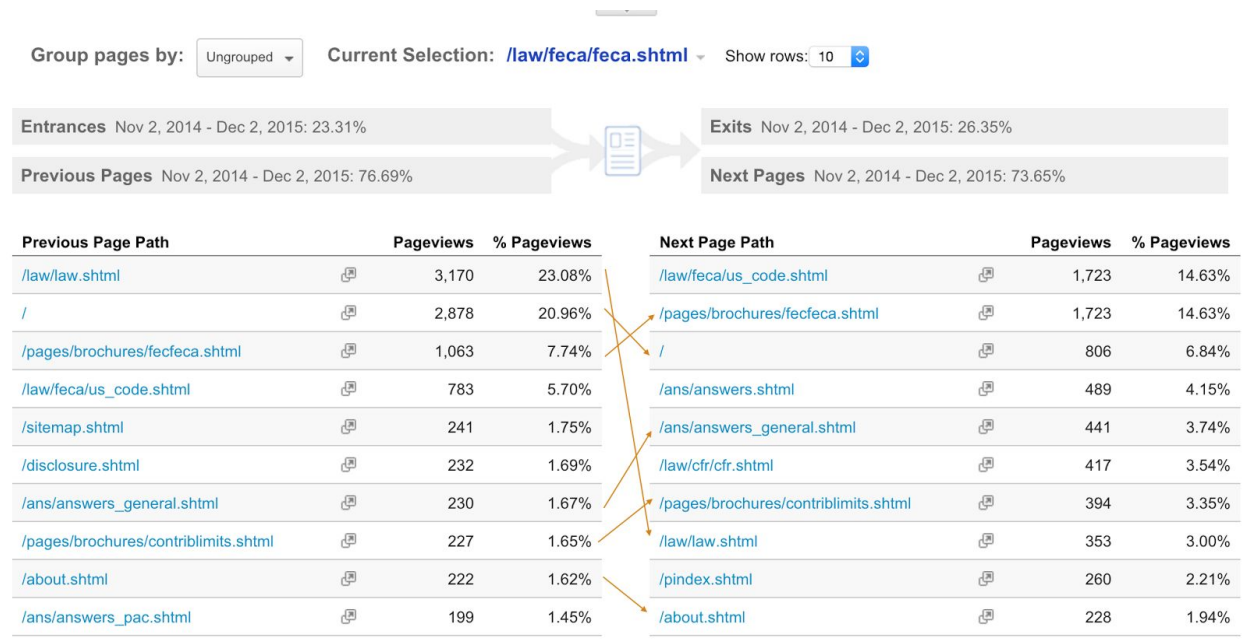


fig. 10, next page path for feca.shtml, Google Analytics. Reference: [Interactions Report from /law/](#)

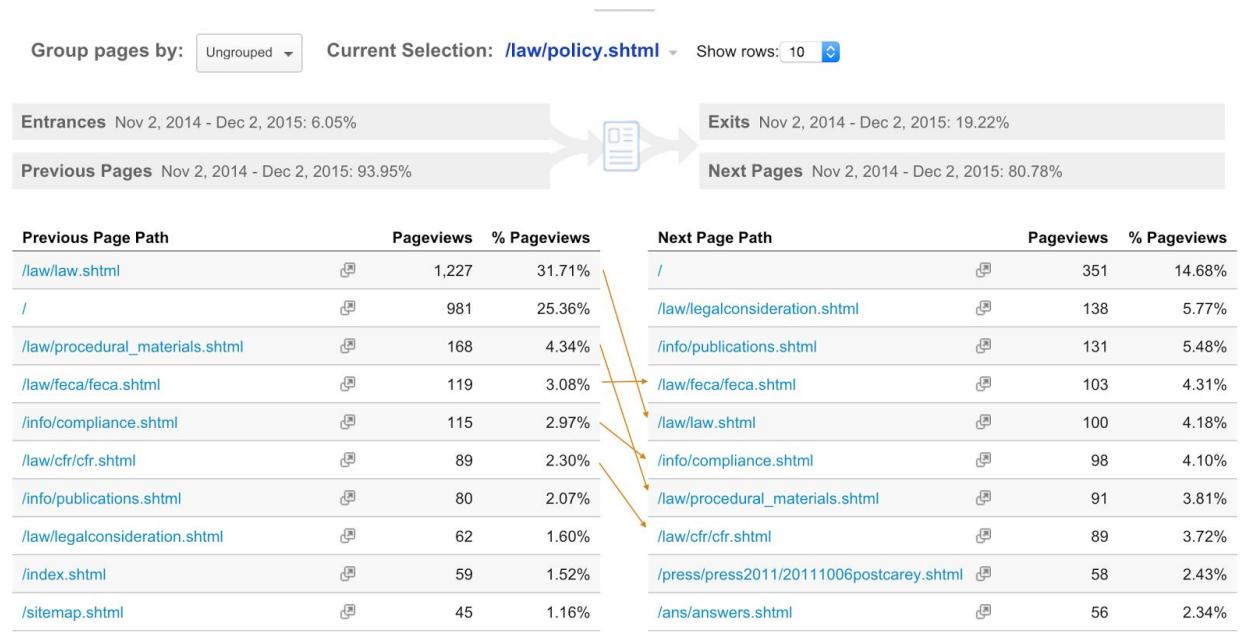


fig. 11: next page path for policy.shtml, Google Analytics

Related to the previous finding that users often view the legal home page multiple times, in analyzing the last and next pages a user visits before they visit a given page, we see multiple occasions where the user goes back to the previous resource they were viewing, after viewing it. In some cases, this is the previous page in the hierarchy (such as the legal home page), which could indicate a user going back either having found or not found the information they desired. However, in other cases, the resource is linked but not in the direct navigation flow of the page in question. This indicates that users are searching for a pattern of content about a given topic that resides across the different site sections. While it's unclear why precisely this is happening, the research should attempt to uncover how the different content types are used in conjunction with each other to solve users' information needs.

4. Search Analysis

For the search analysis, we received search log files for 10/01/2015 to 11/25/2015.


```

1 165.193.100.197 - - [26/Nov/2015:12:54:19 -0500] "GET /vivisimo/cgi-bin/query-
meta?&query=jarl%20mohn&v:project=fec_search_02_prj&x=0&y=0&binning-state=source%3D%3DReceipts%0Aemployers%3
D%3DJarl%20Mohn%0Arcpt_amt%3D1000%3A2000%0Aname%3D%3DWarner%20For%20Congress%28%20Mohn,%20Jarl%29%0Arptyr%3D
%3D2008%0Areported_by%3D%3DWarner%20For%20Congress%20%28C00416040%29%0A HTTP/1.1" 200 119381 "-" "Mozilla/5.
0 (compatible; AhrefsBot/5.0; +http://ahrefs.com/robot/)"
2 54.210.37.165 - - [26/Nov/2015:12:54:33 -0500] "GET /vivisimo/cgi-bin/query-
meta?&query=jarl%20mohn&v:project=fec_search_02_prj&x=0&y=0&binning-state=source%3D%3DReceipts%0Aname%3D%3DC
ommittee%20For%20Change%28%20Mohn,%20Jarl%29%0Aemployers%3D%3DSelf%0Aoccupations%3D%3DInvestor%0Atran_tp%3D%
3D15-Contribution%0Arptyr%3D%3D2008%0Arcpt_amt%3D28000%3A29000 HTTP/1.1" 200 119372 "-" "Mozilla/5.0 (
compatible; AhrefsBot/5.0; +http://ahrefs.com/robot/)"
3 165.193.100.197 - - [26/Nov/2015:12:54:35 -0500] "GET /vivisimo/cgi-bin/query-
meta?&query=jarl%20mohn&v:project=fec_search_02_prj&x=0&y=0&binning-state=source%3D%3DReceipts%0Aemployers%3
D%3DN%2FA%0Arcpt_amt%3D2000%3A3000%0Aname%3D%3DBerman%20For%20Congress%28%20Mohn,%20Jarl%29%0Atran_tp%3D%3DU
ncategorized%20Receipt HTTP/1.1" 200 119291 "-" "Mozilla/5.0 (compatible; AhrefsBot/5.0; +http://ahrefs.
com/robot/)"

```

fig. 12: portion of results (IP addresses changed to protect user privacy)

We parsed these using a perl script to extract the relevant parts, namely the search queries, and counted the occurrence of each distinct query. This is an imperfect process, but the results produced were spot checked and seem reasonable.

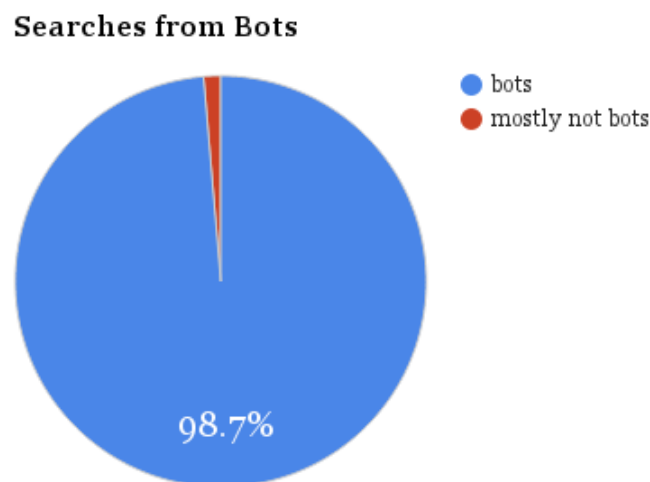


fig. 13: searches originated from acknowledged bots

We found that the vast majority, at least 98% of the searches in October and November originated from bots, software programs meant to scrape the search results for use in other programs. Bots often leave a specific signature (see above, AhrefsBot/5.0). After we

removed the largest offenders of the bot traffic, we coded the rest of the results as pertaining either to campaign finance data searches (e.g., information about a specific candidate or committee), or as that which could conceivably be a search for guidance. Again, this process is imperfect, as it relies on inferring the searchers intent from the term. Often, this was relatively clear cut: people searching for “terminating a committee” are very likely looking for guidance on how to do so. Sometimes it was less so: someone looking for “Super PAC” could be looking for a specific Super PAC or they could be looking for information on what makes a Super PAC. In order to keep the data in the analysis, we tended to categorize guidance relatively broadly (keeping ambiguous searches). Of the traffic not coming from bots, about 81% of it sought campaign data, 19% conceivably searching for guidance.

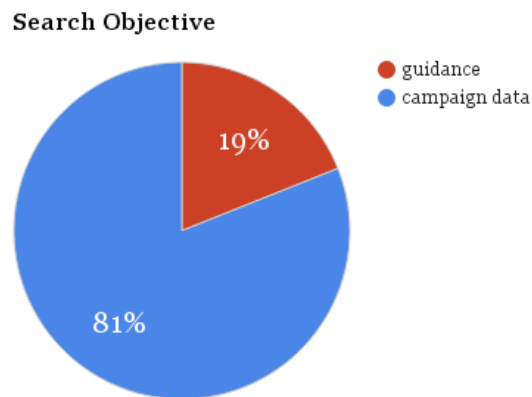


fig. 14: searches by major site section objective

From here, the analysis becomes even more interpretative and open to debate. We coded all guidance related search responses receiving more than 2 occurrences among the following categories:

- how to do a thing (e.g., “how to become a senator”)
- rules around a thing (e.g., “foreign currency”)
- where is this category of thing (e.g., “forms”)
- where is this specific thing (e.g., “advisory opinion 1977-42”)
- what is this thing (e.g., “current rulemaking”)
- and one additional category for search queries that were ambiguous and could clearly be either guidance or campaign finance data related (e.g., “contributions to candidates”)

		mixed result of guidance and data	how to do a thing	rules around a thing	where is this category	what is this thing	where is this specific thing
super pac	118						
bitcoin	100						
soft money	59						
advisory opinions	58						
donors	44						
earmark	43						
winning percent	40						
contribution limits	39						
donations	37						
advisory opinion	37						
hard money	35						
civil penalties	29						
ao request	28						
change in officers	28						
funding a bill	26						
limits	26						
terminating committee	24						

fig. 15: portion of table used for categorizing search terms

Each term could be assigned to multiple categories. The occurrence of each search term was then summed for their categories. Of the categories, two, “what is this thing” and “mixed results” are most open to debate. These represent ambiguous searches, where the term could be interpreted as a user seeking a definition or, alternatively, to look for instances of the search term. For instance, for the search term “Super PAC”, a user may be searching to learn more about how super pacs function, or they could be searching for instances of Super PACs.

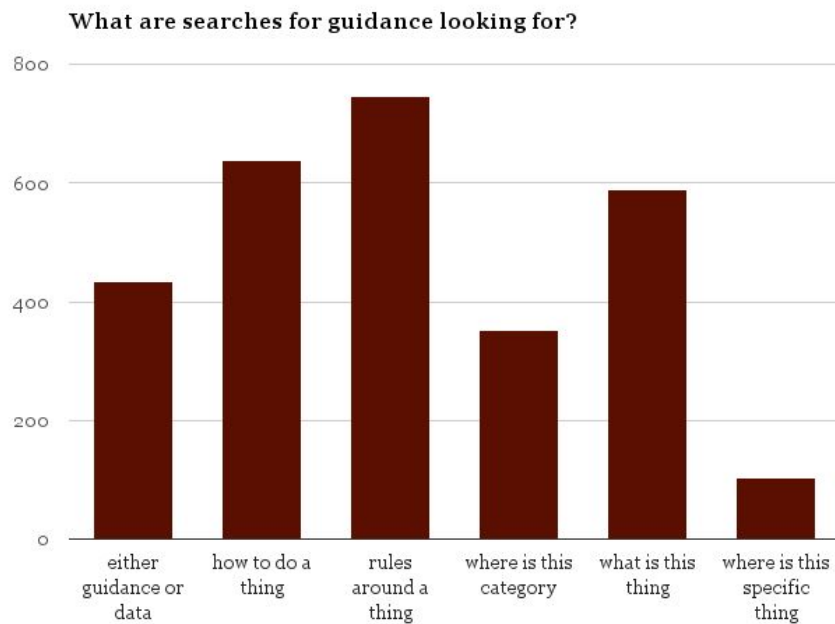


fig. 16: search term category occurrences

For these categories, by far the least common is a user seeking out a specific, known advisory opinion, matter under review, or regulation. Most common were searches seemingly looking to accomplish a task or to understand the rules around a task.

Limitations

First, it's worth restating that these results are preliminary; results are based on interpretations and best educated guesses, inferring as best as possible. That said, while any given term may be misinterpreted, unless there is a particular systemic bias at play, the misinterpretations may in fact net out. Second, it's worth stating that the advanced search data was interrogated less thoroughly. The advanced search functionality allows searching without including a search term in a pattern detected by the script we used. If time allows and further analysis is sought, this may be a great next step.

Interpreting the search analysis results

Probably the most reliable finding is that most people are not trying to find specific MURs and AOs using the main search box. Whether they know to go to the specific search tools for those is unclear. A perhaps more compelling finding is probably the relative

frequency that individuals seem to seeking to find guidance related to an action they are interested in. This may indicate that individuals start with general query and try to dig their way deeper and deeper. If that's true, the legislation, AOs and MURs may represent various strata of bedrock; it's unclear as of yet how often they reach it.