

B O I E S , S C H I L L E R & F L E X N E R L L P

1999 HARRISON STREET* SUITE 900* OAKLAND, CA 94612* 510-874-1000 FAX 510-874-1460

August 10, 2011

BY ECF

The Honorable Donna M. Ryu
United States Magistrate Judge
Northern District of California
1301 Clay Street, Courtroom 4, 3rd floor
Oakland, California 94612

Re: *Oracle America, Inc. v. Google Inc.*, No. 3:10-CV-03561-WHA (N.D. Cal.)
Joint Letter Regarding Oracle's Request for Non-Mobile Data and Projections

Dear Judge Ryu:

This is further to our joint letter to Your Honor of this past Wednesday, August 3, regarding the non-mobile data issues that were the subject of the hearing before the Court and the parties' subsequent meet and confer last Monday, August 1.

The parties have not been able to reach agreement on Oracle's requests for non-mobile data and projections. We therefore ask the Court to decide these issues.

Oracle's Position:

In an effort to reach a compromise, Oracle has radically reduced its non-mobile data requests, and even arranged to buy some of the necessary non-mobile data from a third party. The data Oracle now seeks could be produced in an Excel spreadsheet and would easily fit on a thumb drive. Google has this data. Nonetheless, Google – which searches the entire Internet in fractions of a second millions of times per day – claims that producing even this small amount of data is too great a burden. Google's unsubstantiated claim of burden, for this small quantity of relevant data, cannot trump Oracle's right to reasonable discovery.

A. Non-Mobile Data

The Court's July 22, 2011, *Daubert* order makes clear that "Google profits from Android indirectly," and the measure of damages in this case must consider the "overall value of Android." (Dkt. 230 at 9:16, 9:21.) Your Honor further held, at the August 1 discovery hearing, that Oracle is entitled to seek reasonable discovery concerning network effects of Android.

Google's business is advertising, and it sells advertising by auctioning "keywords" – the ability to have an advertisement appear alongside search results associated with a particular query. As Professor Cockburn explains in his report, and as is well accepted by both economists

The Honorable Donna M. Ryu HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY
 August 10, 2011
 Page 2 of 11

and courts, advertising is a business in which network effects play a significant role. A recent article in the business section of the New York Times makes that very point about Google:

As the overwhelming search leader, Google has advantages that tend to reinforce one another. It has the most people typing in searches — billions a day — and that generates more data for Google's algorithms to mine to improve its search results. All those users attract advertisers. And there is the huge behavioral advantage: "Google" is synonymous with search, the habitual choice. Once it starts, this cycle of prosperity snowballs — more users, more data, and more ad dollars. Economists call the phenomenon "network effects"; business executives just call it momentum. In search, Google has it in spades, and Microsoft, against the odds, wants to reverse it.¹

Android contributes to this "snowball" effect by increasing the number of users, which results in "more ad dollars" for Google, not just in its mobile business, but in non-mobile as well. As Google admits below, "[t]he inquiry is whether Google's *overall advertising revenues* have been affected by the availability of Android devices and, if so, in what way." Oracle seeks data that would allow its experts to quantify that effect. Specifically, Oracle's experts intend to isolate the effect of Android on Google's non-mobile revenues through a statistical analysis of search volume, click-through rates, cost per click, and revenue for keywords before and after the introduction of Android. This analysis cannot be conducted using combined keyword quarterly averages, as Google proposes, because such averages alone cannot isolate the effect of Android. An economist would not measure the effect of electric cars on the automobile industry by looking only at the average price of all GM cars, from Chevettes to Corvettes.

Oracle has radically reduced its non-mobile data requests to a set of data that Google admits it has, that Google admits it can retrieve, and that is small and easily manageable. Specifically, Oracle seeks production of the following data:

- (1) Oracle will provide to Google a set of 5,000 keywords. Because Google claims it cannot provide such a list, Oracle has made arrangements to buy those keywords – at a cost of \$10,000 – from a third party that maintains certain keyword data. The 5,000 keywords represent an appropriate sample from that third party's keyword data set.
- (2) For each of those 5,000 keywords individually, Google will provide quarterly data for the period 1/1/2004 through 7/31/2011 on the following (limited to the United States):
 - Total search volume (e.g. number of auctions or searches per month per keyword)
 - Average Cost Per Click ("CPC")
 - Average Click Through Rate ("CTR")
 - Total Search Advertising Revenue
- (3) Google will provide the quarterly aggregate data between and including Q1 2004 and Q1 2011 on the following (limited to the United States):
 - Total search volume (e.g. number of auctions or searches per month per keyword)

¹ "Can Microsoft Make You 'Bing'?" *The New York Times*, July 30, 2011, at BU 1, available at http://www.nytimes.com/2011/07/31/technology/with-the-bing-search-engine-microsoft-plays-the-underdog.html?_r=1&sq=google%20bing&st=cse&scp=2&pagewanted=all

The Honorable Donna M. Ryu HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY
August 10, 2011
Page 3 of 11

- Average CPC
- Average CTR
- Total Search Advertising Revenue

Google claims – without any competent evidentiary support – that this narrowed request is still too burdensome. In fact, the requested dataset is so small that it could be produced in an Excel spreadsheet and stored on a thumb drive. Moreover, it is clear that Google tracks this data. Google's Adwords Traffic Estimator² provides the average monthly number of searches per keyword. Accountholders can run multiple queries – at least 50 – simultaneously. Moreover, the Traffic Estimator provides results in a second or less that include, by keyword, “global monthly searches,” “local monthly searches,” and estimated average CPC. Similarly, Google's Keyword Tool³ allows anyone to type in one or more terms, and instantly view up to 100 variations on those terms, with “global monthly searches” and “local monthly searches” for each such variation. The search can also be filtered to limit the data to a specific region, such as the United States. Thus, one can enter “plush dolls” into the search query, and instantly see that “plush dolls” had an average of 49,500 searches per month on desktop and laptop devices globally, and an average of 33,100 per month in the United States. The same data for variations on the term – like “plush Elmo doll” – are also instantly provided. The Google filter further allows the user to break down results in numerous ways, such as searches using “desktops and laptop devices” versus “all mobile devices” versus “mobile WAP devices” versus “mobile devices with full internet browsers.” In addition, Google's “Insights for Search” feature allows a user to obtain statistics about search volume for specific terms in specific geographic regions in any specified time frame from 2004 to present.⁴

Clearly Google has historical data on search per keyword and CPC per keyword, and has developed tools to allow such data to be rapidly searched and retrieved by the public. Google now says this data is limited and inaccessible. But Google itself wrote on its Adwords API Blog, “[a]s you could imagine, Google has *a lot* of historical data about keywords and the traffic they generate.”⁵ (emphasis Google's) Google's blog further states that “AdWords API users can access our trove of historical keyword and bid data via the Traffic Estimator Service, which gives detailed estimates of how much traffic a keyword may generate at various CPC values.”

Google next objects that the requested data could not be useful because – according to Google – a dataset based on 5,000 keywords is arbitrary, and ironically, *too small* to be statistically significant. Google's objection – without having seen the data or the analysis – is unfounded and premature. Because Google claims it can produce neither a list of keywords nor

² This is an online tool provided by Google and located at https://adwords.google.com/o/Targeting/Explorer?__u=1000000000&__c=1000000000&ideaRequestType=KEYWORD_STATS#search.none.

³ This is an online tool provided by Google and located at https://adwords.google.com/o/Targeting/Explorer?__c=1000000000&__u=1000000000&__o=te&ideaRequestType=KEYWORD_IDEAS#search.none.

⁴ <http://www.google.com/insights/search/> (described by Google as “See what the world is searching for”)

⁵ <http://adwordsapi.blogspot.com/2008/07/traffic-estimator-service-explained.html> (This page bears a 2010 Google copyright.)

The Honorable Donna M. Ryu HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY
August 10, 2011
Page 4 of 11

a sample, Oracle has arranged to purchase, for \$10,000, a list of randomly selected keywords from AdGooroo, a third-party vendor that tracks almost half a million keywords. Google's footnoted speculations that Oracle would "select keywords that it would expect to have risen in 'price' since Android devices were introduced" and that Oracle intends to select a "non-representative group of keywords" are simply false. Oracle's proposal is neither arbitrary nor statistically insignificant. To the contrary, it provides a reasonable, useful dataset *and* resolves Google's objections that it would be too burdensome for it to provide the keywords itself.

Rather than providing the data by keyword necessary to Oracle's analysis, Google proposes that Oracle rely on aggregate data alone. However, such data alone cannot demonstrate network effects in this case. Keywords are the economic unit by which Google sells advertisements – for each monetized search, Google holds an "auction" by which it sells advertising space connected to specific keywords.⁶ Android increases the value of keywords generally because it increases the number of searches and, therefore, the volume of keyword sales. Separately, Android increases the utility and therefore price of certain keywords; for example, location-specific keywords become more valuable as they are used more heavily on Android. Searchers on Android devices are more likely to search for location specific keywords (increase search volume) and click through (increase CTR); accordingly, advertisers will pay more (CPC) and generate more revenue for Google (total revenue). These benefits are not observable without examination of keyword specific data. Google's argument that Professor Cockburn's damages report did not mention keywords is nonsense. Professor Cockburn's report states that he did not calculate network effects because Google refused to produce this data.

B. Non-Mobile Projections

Oracle's RFP No. 170 seeks non-mobile projections from 2004 through the present. Google has refused to produce any such projections. Oracle offered, prior to the parties' July 21 letter to the Court, to reduce its request to projections created in the 15-month-period surrounding fall 2008, when Android was first shipped commercially. Judge Alsup has since held that the infringement may have occurred at different points in time for distinct patents and copyrights. Consequently, Oracle is still willing to limit its request to non-mobile projections created during five quarters, but proposes that those quarters be calendar year Q3 2005, Q3 2006, Q3 2007, Q3, 2008, and Q1 2009. If no projections are available from any particular quarter, Oracle would ask that projections instead be produced from the next available quarter.

C. Evidence

The Court made clear that the burdens associated with Oracle's original requests required that Oracle make an evidentiary showing so the Court could perform the balancing necessary under Rule 26. Oracle submits that its much reduced data requests should require less of an evidentiary showing. Moreover, Oracle's ability to demonstrate the fact of Android-related network effects, and Google's own recognition of them, is impaired by Google's failure to

⁶ For any given keyword, Google encourages advertisers to place advertisements on *both* mobile and non-mobile devices through a single AdWords' campaign, through its default setting and Google "smartpricing." <http://adwords.google.com/support/aw/bin/static.py?hl=en&guide=23292&page=guide.cs>

August 10, 2011

Page 5 of 11

produce that information. Your Honor directed Google on August 1 that “any analysis of any kind that Google did that touches on the issue of network effects [of Android] has to be produced” by Google, and Google must certify to Oracle that it has conducted such a search and produced all such documents. Google has failed to produce any such certification, refuses to state when it will do so, and appears to deny that the Court in fact gave that specific instruction.

In any event, the evidence demonstrates the existence of Android’s network effects on Google’s non-mobile business, including revenues, Google’s awareness of those effects, and Google’s efforts to increase those effects. That evidence includes:

[REDACTED]

The Honorable Donna M. Ryu
August 10, 2011
Page 6 of 11

HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY



For all of these reasons, Oracle respectfully requests that the Court order Google to (1) produce the non-mobile data and projections described in this letter, and (2) produce “any analysis of any kind that Google did that touches on the issue of network effects” and certify to Oracle that it has done so, no later than August 15, 2011.

Google’s Position:

At the hearing last Monday, Your Honor made clear that Oracle would be required to show “specific evidence” supporting its requests for the extensive non-mobile data and projections that it is seeking before it would be entitled to that additional data. Oracle has not met its burden, yet it still seeks to impose substantial burdens on Google.

As we discussed with the Court during the hearing, Google has already produced to Oracle significant information regarding Google’s overall, i.e., mobile and non-mobile, advertising revenues. Thus, the information that Oracle requests would be additional information over and above what Google has already provided.

We have carefully reviewed the materials that Oracle has shared with us that it believes support its requests, including the documents cited above by Oracle. We do not believe that those materials justify production of the extensive additional non-mobile data that Oracle is seeking. We believe that this is particularly true with respect to Oracle’s continued requests for extensive data for some number of specific “keywords” for which advertisers pay Google.

The Honorable Donna M. Ryu HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY
August 10, 2011
Page 7 of 11

First, we note in this regard that *none* of the documents identified by Oracle refer to any metrics or data based on keywords, and that the Cockburn damages report that has been excluded by Judge Alsup does not even mention “keywords.” Second, the documents on which Oracle relies at most refer generally to possible effects on *overall Google advertising revenues* as a result of increased mobile searches; the additional data that Google is prepared to provide to Oracle, as described below, would be directly responsive to and would relate to those same general, overall revenues. Third, the documents cited by Oracle are more of what Oracle has previously cited, namely documents reflecting that Android is generally “good for Google” (including Google search and other Google services), as it enhances the mobile search experience – and the Court has already ruled that such general documents are insufficient to justify the additional discovery. Fourth, as is explained more fully below, the gathering and production of the data that Oracle has requested would be a very substantial undertaking that would take significant resources – including both engineering resources, computing resources, and time – to complete.

In view of the above, Google does not believe that Oracle can meet the burden that the Court indicated it must meet before the Court would order any of the additional requested non-mobile discovery, over and above the non-mobile information that Google has already provided to Oracle (which includes Google’s overall, non-mobile revenue numbers and financials).

If the Court nonetheless concludes that Oracle has made a sufficient showing to justify some additional discovery on Google’s non-mobile advertising revenues, Google respectfully submits that the proposal made by Google to Oracle would be more than adequate. More specifically, Google has proposed that – if the Court concludes that Oracle has made the required threshold showing – Google would provide to Oracle aggregate and/or average quarterly data for the past seven years (2004 through the first quarter of 2011) regarding Google’s total number of searches, average “cost per click” for search ads, average “click through rate” for search ads, and total advertising revenues.


This data could be gathered in a reasonable period of time – likely in a three to four week time frame – and, while it would require significant engineering time and resources, it would not require the far greater and extensive amount of engineering time and resources that Oracle’s broader, keyword-based requests would entail. More importantly, this data would also be consistent with – and indeed go well above and beyond – the assertions made in the excluded Cockburn report regarding alleged “network effects,” which relate to Google’s overall advertising revenues (mobile and non-mobile) and do not raise any issues, metrics or data based on keywords. This data would also cover a time period that pre-dates the introduction of Android-based smartphones by several years and would provide Oracle with a pre-Android frame of reference.

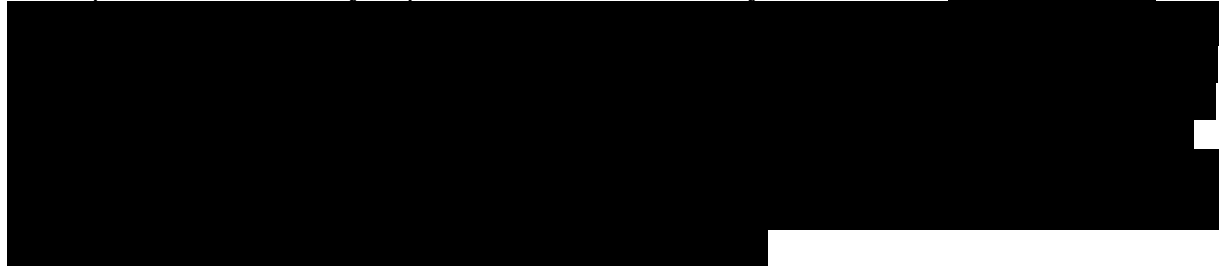
In contrast, Oracle’s proposals for resolution of the non-mobile data issues have been consistently unreasonable. In its July 1 letter to the Court, Oracle originally requested massive amounts of data that would take months to gather if they could be gathered at all, including “web content indexed by Google” and “number of apps for the desktop,” and metrics (such as keyword data for display ads) that make no sense. Whether these requests were due to Oracle’s failure to

The Honorable Donna M. Ryu
August 10, 2011
Page 8 of 11

HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY

diligently investigate through publicly-available materials the advertising services for which it is demanding highly confidential and proprietary information or its continued overreaching in both discovery and its damages theories, they plainly sought to impose an unreasonable burden and significant costs on Google.

Oracle claims to have “radically” reduced its demands, but it has reduced them only from the impossible to the nearly impossible and extraordinarily burdensome. 



Even then, moreover, the resulting dataset would provide information regarding only an extremely small subset of keywords, namely an arbitrary group of 5000 keywords. As we have also advised Oracle, there are literally hundreds of millions of keywords that are included in searches in just a single month. The relatively small group of 5000 keywords would apparently be selected by Oracle or a third party vendor based on unidentified criteria; based on the keywords chosen, the results of any such analysis could be drastically skewed – and, for the reasons stated above, whether they are representative of any larger group of keywords would as a practical matter be impossible to test.⁸ In view of these parameters, the data requested by Oracle (if it could even be gathered) could not in any way be statistically significant, and any analysis based on them would be purely speculative and not grounded in any reliable scientific methodology. The popularity of or metrics for any specific keyword or keywords cannot have a bearing on whether the existence of Android devices has had an effect on Google’s overall advertising revenues.

Importantly, moreover, none of the documents relied on by Oracle suggest any such bearing, and Oracle’s analogy about “GM cars” is inapplicable. If anything, this analogy counsels against Oracle’s requests; an arbitrarily selected group of 5000 keywords out of hundreds of millions used in just a given month would have no more bearing on Google’s overall advertising business than would the price of just a few specific GM models having specific accessories on GM’s overall business. The relevant inquiry, in short, is not whether one or more

⁷ Google can provide further or more detailed information or evidentiary support regarding these facts if the Court requires.

⁸ Oracle has refused to advise Google how the 5000 keywords would be selected, other than to say they would be “keywords that have actually been auctioned by Google at various points in time.” It appears from Oracle’s portion of this letter that Oracle would intend to select keywords that it would expect to have risen in “price” since Android devices were introduced – without consideration of whether such keywords were representative of any larger group, whether increases in their “prices” may be offset by decreases in “prices” for other keywords, or what if any effect Android may have had on Google’s overall revenue. Thus, it appears that Oracle indeed intends to select a non-representative group of keywords that it believes are most likely to support its theories.

The Honorable Donna M. Ryu HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY
August 10, 2011
Page 9 of 11

individual keywords are more “expensive” now than they were prior to the introduction of Android phones and tablets. The inquiry is whether Google’s *overall advertising revenues* have been affected by the availability of Android devices and, if so, in what way.

Given the unreasonable burdens associated with gathering the information requested by Oracle, its statistical insignificance, and the necessarily speculative nature of any conclusions that could be drawn from it, Google submits that the costly and burdensome exercise requested by Oracle would in the end necessarily be unproductive and improper. Oracle has not suggested how analysis of such an extremely limited data set – which would require extensive effort to generate – could be useful in addressing its “network effects” arguments, let alone the causal nexus and other requirements it must meet under Judge Alsup’s Daubert order in support of its revised damages report.

Oracle’s reliance on the limited data available to AdWords customers on Google’s websites is misplaced. The website data to which Oracle refers are only *estimates of future projections* or *approximations of averages*. [REDACTED]

[REDACTED] The estimates themselves can also be manipulated in that the user of the estimator tool can change the input variables. See, e.g., the information at <http://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=27331> and <http://adwords.google.com/support/aw/bin/answer.py?hl=en&answer=8692>. And the data provided by the “Insights for Search” tool that Oracle cites is data relating to queries or searches – not the extensive metrics requested by Oracle.

In contrast to these estimates at a single point in time based on limited recent data, the “by keyword” data that Oracle is requesting is extensive, accurate historical data covering over seven years of time. The availability of the estimates is therefore irrelevant to the far larger and much more extensive data requests that Oracle wrongly contends are “reasonable.”

Similarly, Oracle’s assertion that the information it wants could be provided in a spreadsheet and fit on a thumb drive (and is therefore presumably reasonable) is both facile and misleading. For purposes of the Court’s decision, the issue is what kind of effort must be undertaken to gather the data – not the size of the resulting dataset or how many data points are to be included after that extensive effort is made. It may be a trite example, but the largest known prime number could also fit on a thumb drive, but it required two years and the processing of multiple supercomputers to discover. As outlined above, the effort in this case would be far in excess of any reasonable effort, especially in view of the information Google has already produced.

On the separate issue of “projections,” Oracle has also this week expanded its demands rather than “radically reduce” them. Instead of attempting to compromise its demands, Oracle has most recently requested revenue projections made by Google over a *five-year* period, rather than the *fifteen-month* period it originally requested. Based on the actual historical data Google has already provided and the additional information Google is prepared to provide, however,

The Honorable Donna M. Ryu HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY
August 10, 2011
Page 10 of 11

production of documents showing past projections is unnecessary and is not relevant to any alleged "network effects."

Finally, Oracle suggests that Google has "refused" to comply with the Court's direction regarding the production of any additional documents that may discuss network effects relating to Android. We are in the process of complying with that direction and will do so. We note in this regard that our investigation to date has shown that Google has in fact produced a number of such documents to Oracle, including the Google documents cited above. We are currently determining whether there are any additional such documents that need to be produced in order to supplement Google's already extensive production of documents relating to Android and Google advertising revenue (and not "any analysis of any kind that Google did that touches on the issue of network effects").

Google therefore respectfully requests that the Court deny Oracle's request for additional non-mobile data on the grounds that Oracle has not made the specific showing necessary to justify any such additional discovery. In the alternative, if the Court finds that Oracle has made a sufficient showing, Google requests that the Court limit the data that Google must provide to the aggregate and average data identified above, for the time periods stated.

We are available to discuss these issues further with the Court, by telephone or in person, at the Court's convenience.

Respectfully submitted,

BOIES, SCHILLER AND FLEXNER LLP

KING & SPALDING LLP

/s/ Fred Norton
Fred Norton

/s/ Bruce W. Baber
Bruce W. Baber

The Honorable Donna M. Ryu
August 10, 2011
Page 11 of 11

HIGHLY CONFIDENTIAL -- ATTORNEYS' EYES ONLY

ATTESTATION OF FILER

I, Fred Norton have obtained Mr. Bruce W. Baber's concurrence to file this document on his behalf.

Dated: August 10, 2011

BOIES, SCHILLER & FLEXNER LLP

By: /s/ Fred Norton
Fred Norton

Attorneys for Plaintiff
ORACLE AMERICA, INC.