

DAVIS WRIGHT TREMAINE LLP

1 THOMAS R. BURKE (CA State Bar No. 141930)
 DAVIS WRIGHT TREMAINE LLP
 2 505 Montgomery Street, Suite 800
 San Francisco, California 94111
 3 Telephone: (415) 276-6500
 Facsimile: (415) 276-6599
 4 Email: thomasburke@dwt.com

5 RONALD G. LONDON (Pro Hac Vice)
 DAVIS WRIGHT TREMAINE LLP
 6 1919 Pennsylvania Ave., N.W., Suite 800
 Washington, DC 20006
 7 Telephone: (202) 973-4200
 Email: ronnielondon@dwt.com

8 DAN LAIDMAN (State Bar No. 274482)
 DAVIS WRIGHT TREMAINE LLP
 9 865 South Figueroa Street, Suite 2400
 Los Angeles, CA 90017-2566
 10 Telephone: (213) 633-6800
 11 Facsimile: (213) 633-6899
 Email: danlaidman@dwt.com

12 DAVID HALPERIN (Pro Hac Vice)
 13 1530 P Street NW
 Washington, DC 20005
 14 Telephone: (202) 905-3434
 Email: davidhalperindc@gmail.com

15 Attorneys for Plaintiff Public.Resource.Org

16
 17 IN THE UNITED STATES DISTRICT COURT
 18 THE NORTHERN DISTRICT OF CALIFORNIA
 19 SAN FRANCISCO DIVISION

20
 21 PUBLIC.RESOURCE.ORG., a California non-) Case No. 3:13-CV-02789-WHO
 profit organization,)
 22)
 Plaintiff,) **DECLARATION OF PETER**
 23) **SKOMOROCH**
 v.)
 24)
 UNITED STATES INTERNAL REVENUE)
 25 SERVICE,)
)
 26 Defendant.)

1 I, Peter Skomoroch, I, declare as follows:

2 1. I am an equity partner at Data Collective, a venture capital firm in Silicon Valley. I
3 have personal knowledge of the matters stated in this declaration and could competently testify to
4 them if called as a witness.

5 2. From 2009 to 2013, I was a Principal Data Scientist at LinkedIn, the world's largest
6 professional network on the Internet with more than 313 million members in over 200 countries
7 and territories. At LinkedIn, I helped the company build tools that allows individuals and
8 companies to build profiles of their skills and interests. For example, the IRS maintains a profile
9 page about LinkedIn at <https://www.linkedin.com/company/4922>.

10 3. As part of my work as a Data Scientist, I helped LinkedIn develop algorithms that
11 provided users of our site relevant information when they do searches. In addition to the
12 information provided by the companies and individuals directly, we used a variety of outside
13 sources of information to help determine relevance.

14 4. Directories of business information are very helpful in developing these types of
15 algorithms. The IRS Exempt Organizations database is particularly important for the nonprofit
16 sector, but the fact that the data is released as bitmap images (scans) makes it much harder to take
17 advantage of this information.

18 5. The IRS Modernized e-File (MeF) format consists of Form 990 information coded
19 in the XML markup syntax, a format that is significantly easier to work with to extract
20 information.

21 6. Extracting information about companies is a common task for a large number of
22 established and startup companies in Silicon Valley. For example, if you search for "american
23 national standards institute" on Google, you will receive sidebar about the company which
24 includes information from the Wikipedia entry, as well as the Tax ID, the date it was founded, the
25 name of the CEO, and the tax deductibility code which is 501(c)(3).

26 7. Rudimentary information about nonprofit companies can be found in the IRS
27 Exempt Organizations Business Master File. This file can be found at
28

1 [http://www.irs.gov/Charities-&-Non-Profits/Exempt-Organizations-Business-Master-File-Extract-](http://www.irs.gov/Charities-&-Non-Profits/Exempt-Organizations-Business-Master-File-Extract-EO-BMF)
2 [EO-BMF.](http://www.irs.gov/Charities-&-Non-Profits/Exempt-Organizations-Business-Master-File-Extract-EO-BMF)

3 8. The Master File contains a few useful fields about organizations, including their
4 address, income, and assets. A description of the fields in the Master File can be found at
5 http://www.irs.gov/pub/irs-soi/eo_info.pdf.

6 9. If individual returns in MeF format were made available, much more information
7 about organizations could be extracted and this information would be of great use to a number of
8 companies providing information on the Internet.

9 10. The XML format on which Modernized e-File (MEF) is based is a well-known
10 format for programmers. It would be very simple to parse information in these files and use them
11 in web-based services.

12 11. Working with the information currently released, which is in the form of bitmap
13 images, is much more difficult. Optical Character Recognition is difficult to use with low-
14 resolution TIFF images, which is what the IRS returns are released as. OCR is subject to
15 recognition errors and it is difficult to parse individual fields accurately. Data in MeF format is
16 not subject to these limitations.

17 12. I believe that the Exempt Organizations database is a valuable source of business
18 information for an important sector and that information can be used to provide better directory,
19 search, and information services on the Internet. Providing the information in MeF format would
20 make the creation of those services significantly easier and more powerful.

21
22
23
24
25
26
27
28

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

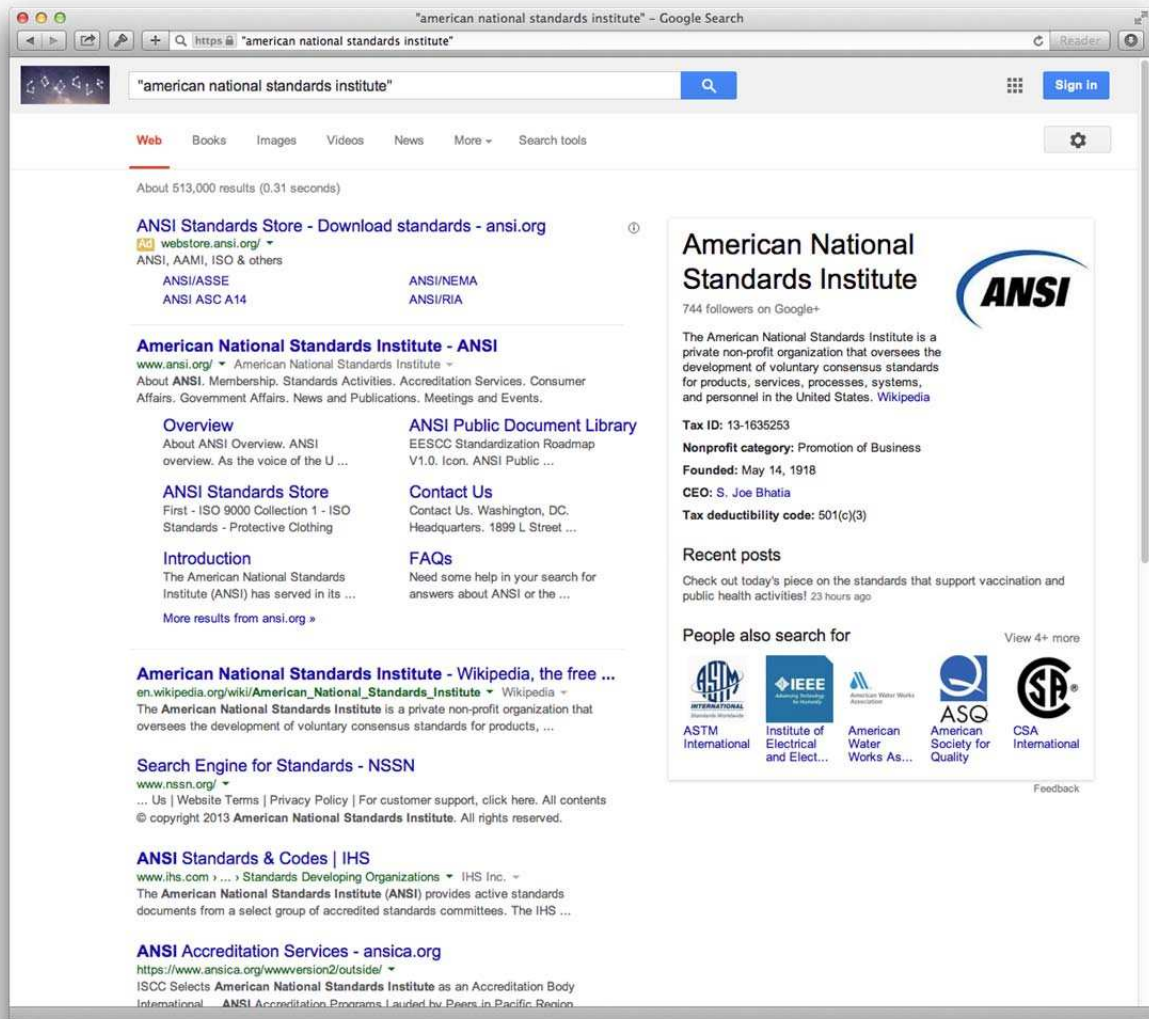


Figure 1: Example of a search on Google for “American National Standards Institute

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct and that this declaration was executed this 23rd day of September, 2014 at San Francisco, California.

/s/ Peter Skomoroch
PETER SKOMOROCH