

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

1 I, Carl Malamud, declare as follows:

2 1. Since 2007, I have been the President and Founder of Public.Resource.Org, a
3 nonprofit corporation and the Plaintiff in this FOIA action. I have personal knowledge of the
4 matters stated in this declaration and could competently testify to them if called as a witness.

5 **Mr. Malamud's Background and Experience**

6 2. My formal education was in Business Economics and Public Policy at the Indiana
7 University School of Business where I completed all coursework for the Doctorate in Business
8 Administration and received an MBA in 1982.

9 3. From 1982 to 1992, I worked professionally in the field of computer networks,
10 including positions at the Board of Governors of the Federal Reserve System, numerous
11 consulting engagements with government groups such as the Department of Defense, wrote as a
12 Contributing Editor and columnist for numerous trade publications such as Communications
13 Week, and authored 8 professional reference books.

14 4. From 1993 to 1996, I served full-time as the founder and executive director of the
15 Internet Multicasting Service, where I started and ran the first radio station on the Internet. As
16 part of my work at the Internet Multicasting Service, I was also responsible for putting the U.S.
17 Securities and Exchange Commission EDGAR system on the Internet and then donating
18 computers and software to the SEC so they could take my system over. I was also responsible for
19 putting numerous other government databases on the Internet for the first time, including the U.S.
20 Patent database.

21 5. In 1998 and 1999, I was the CEO of Invisible Worlds. During that period, I
22 worked with my Chief Technology Officer, Dr. Marshall T. Rose, to help develop the tools used
23 to produce Internet Standards. These tools are based on the XML markup language, which is the
24 same language that the IRS uses for their Modernized e-File (MeF) format. These tools continue
25 to be used as the basis for authoring documents for the Internet standards process. The
26 specifications for this work have been published as Internet Request for Comments 2629, "Writing
27 I-Ds and RFCs using XML." That standard may be found at <http://tools.ietf.org/html/rfc2629>.

28

1 6. In 2004, I was a consultant on documentation strategies to the Internet Systems
2 Consortium, a nonprofit corporation that produces software essential to the operation of the
3 Domain Name System. I was the founding Chairman of the Board of the Internet Systems
4 Consortium in 1994. ISC is the author and publisher of BIND, which is used by many large
5 Domain Name Servers throughout the world and also operates the “F” Root Name server, which is
6 one of the core authoritative name servers that make the Internet function. As a consultant on
7 documentation strategies, I spent a great deal of time working with Docbook, an XML-based
8 authoring language for technical documentation.

9 7. In 2007, I founded Public.Resource.Org, a nonprofit corporation which is based in
10 California. We are responsible for placing the historical opinions of the U.S. Court of Appeals
11 back to the founding of the court on the Internet for the first time. As part of that work we
12 discovered numerous Social Security Numbers (SSNs) in those opinions and notified the Court of
13 the presence of this information. On July 16, 2008, Chief Judge Lee H. Rosenthal thanked us for
14 our efforts on behalf of the Committee on Rules of Practice and Procedure of the Judicial
15 Conference of the United States. That letter may be found at
16 <https://public.resource.org/scribd/7512576.pdf>.

17 8. In 2008 and 2009, I conducted a series of audits on 20 million pages of PACER
18 documents and discovered numerous SSNs. We notified the Chief Judges of 32 U.S. District
19 Courts of these findings and this resulted in changes in the privacy procedures for the PACER
20 documents and acknowledgment of our efforts by several Chief Judges and by the Committee on
21 Rules of Practice and Procedure of the Judicial Conference of the United States.

22 9. In 2007 and then again in 2010, I submitted reports to the Speaker of the House of
23 Representatives concerning my recommendations for broader availability of video from
24 Congressional hearings. On January 5, 2011, the Speaker of the House acknowledged my efforts
25 and authorized me to work with the Committee on Oversight and Government Reform and the
26 House Broadcast Studio, an effort that led to the posting of over 14,000 hours of video from
27 Congressional hearings. The letter from the Speaker may be found at
28 <https://law.resource.org/rfcs/gov.house.20110105.pdf>.

1 10. In 2008, I served as an advisor to the Presidential transition, where I outlined a
2 series of proposed changes in how the Official Journals of Government, including the Federal
3 Register, can be published. Those changes were implemented and have resulted in a substantial
4 improvement in the online system, which is visible at federalregister.gov.

5 11. In 2008, I began a program called FedFlix in cooperation with the National
6 Technical Information Service (NTIS) and the National Archives and Records Administration.
7 The program sent volunteers into the National Archives to copy videos and obtained copies of
8 video from numerous agencies, including the Department of Defense, OSHA, and the Mine Health
9 and Safety Administration. Approximately 6,000 videos were copied and posted to YouTube and
10 the Internet Archive and have since been viewed over 50 million times.

11 **Mr. Malamud's Work with the IRS Exempt Organizations Database**

12 12. In 2008, I began working with the IRS Exempt Organizations database by
13 submitting payment for 6 years of DVDs and developing software to process that data and post it
14 on the Internet with no restrictions on use. Since 2008, I have processed and posted on the
15 Internet over 7,634,050 instances of the Form 990 filed by Exempt Organizations. The data that I
16 processed was made available on our servers, on nonprofit services such as the Internet Archive,
17 and forms the basis for numerous other commercial and non-commercial systems that analyze and
18 host Form 990 data. Our archive of Form 990s is the only one freely available on the Internet with
19 no restrictions on access or use. We make this data available free of charge and with no
20 restrictions, just as we have with court documents and numerous other government databases,
21 because we believe that these Works of Government should be more broadly available.

22 13. As part of my work, I performed audits of the Exempt Organizations database
23 looking for instances of where the IRS has released individuals' SSNs as part of its release of
24 Form 990 data. Our best estimate is that there are close to 600,000 SSNs in the Exempt
25 Organizations data we purchased from the IRS. When I find SSNs in a Form 990, I redact that
26 information and replace the files we made available for public view. I also systematically notify
27 the IRS, GuideStar, the Foundation Center, and others who I know have copies of this database.

1 14. On July 2, 2013, I notified the IRS and the Treasury Inspector General for Tax
2 Administration (TIGTA) of a large number of Social Security Numbers for political organizations
3 filing under Section 527 that were on the IRS web site. That notification can be found at
4 <https://bulk.resource.org/irs.gov/eo/doc/irs.gov.20130702.pdf>. The Inspector General assigned
5 complaint number 63-1307-0025-C to their investigation of this matter.

6 15. On July 15, 2013, Congressman Tom Latham and 41 other members of the House
7 of Representatives wrote to the Acting Commissioner of the Internal Revenue Service to request
8 an explanation of this privacy breach. That letter may be found at
9 <https://bulk.resource.org/irs.gov/eo/doc/irs.gov.20130715.pdf>.

10 16. On September 16, 2013, the Acting Commissioner wrote to Congressman Tom
11 Latham and informed the Congress that the IRS had changed the position on redaction of Social
12 Security Numbers. That letter may be found at
13 <https://bulk.resource.org/irs.gov/eo/doc/irs.gov.20130916.pdf>.

14 17. On December 6, 2013, the Internal Revenue Service updated section 3.20.13.13.2
15 of the Internal Revenue Manual to permit redaction of Social Security Numbers for Section 527
16 Political Organizations. Those changes were effective January 1, 2014. This section of the IRM
17 may be found at http://www.irs.gov/irm/part3/irm_03-020-013r.html.

18 18. On April 22, 2014, I notified the IRS Commissioner and the Inspector General of a
19 large number of Social Security Numbers in returns for Exempt Organizations that are not
20 Political Organizations. That letter may be found at
21 <https://bulk.resource.org/irs.gov/eo/doc/irs.gov.20140422.pdf>.

22 19. On July 7, 2014, I concluded the audit of SSNs and sent the IRS Commissioner and
23 the Inspector General detailed audit results, including copies of 9,392 returns that I had redacted
24 with detailed recommendations on steps the IRS should take to mitigate this problem. The cover
25 letter for this audit may be found at <https://bulk.resource.org/irs.gov/eo/doc/irs.gov.20140707.pdf>.
26 The Inspector General assigned complaint number 63-1407-0060-C to their investigation of this
27 matter.
28

1 Schemas and Business Rules for 1120, 1120S, 1120-F, and 7004 Modernized e-File (MeF).” That
2 information can be found at [http://www.irs.gov/Tax-Professionals/e-File-Providers-&-
3 Partners/2014-Valid-XML-Schemas-and-Business-Rules-for-1120-1120S-1120-F-and-7004-
4 Modernized-e-File.](http://www.irs.gov/Tax-Professionals/e-File-Providers-&-Partners/2014-Valid-XML-Schemas-and-Business-Rules-for-1120-1120S-1120-F-and-7004-Modernized-e-File)

5 27. The name of the file that I examined is
6 Example_TransmissionWithConsolidatedReturn.xml. A copy of the file I examined is available at
7 https://bulk.resource.org/irs.gov/eo/doc/doc/Example_TransmissionWithConsolidatedReturn.xml.

8 28. In order to remove (redact) one element nested inside an XML file, I use a common
9 programmers tool called a “text editor.” Any professional programmer has access to such
10 software. I use a text editor called bbedit on my Apple computer. Other examples of text editors
11 are “vi” on any Unix computer, and “notepad” on any Windows computer. I used the bbedit
12 software on the file named Example_TransmissionWithConsolidatedReturn.xml, removed the
13 element IRS1120LScheduleB, and saved the file with a new name. That entire process took me
14 57 seconds.

15 29. There are a number of techniques used to transform and process XML files. A
16 common technique is the use of Style Sheets, a standard defined by the World Wide Web
17 Consortium, the standards-making body for the World Wide Web. The definition of Extensible
18 Stylesheet Transformations (XSLT) may be found at <http://www.w3.org/TR/xslt>.

19 30. The IRS uses this technique to publish a number of sample files that can be used to
20 transform returns in MeF. These style sheets can be used by businesses, tax preparers, and others
21 to transform a return into another format, such as transforming the XML into HTML for display in
22 a web browser. The IRS publishes these style sheets at [http://www.irs.gov/Tax-Professionals/e-
23 File-Providers-&-Partners/Modernized-e-File-MeF-Stylesheets](http://www.irs.gov/Tax-Professionals/e-File-Providers-&-Partners/Modernized-e-File-MeF-Stylesheets).

24 31. I wrote a very simple style sheet, a true and correct copy of which is attached as
25 Exhibit A, that is based on something called an “identity transformation.” An identity
26 transformation is a style sheet that copies everything that is input to the output with no changes.
27 An example of the identity transformation may be found in Section 7.5 of the XSLT specification.
28 I added a single line to the style sheet which copies every element except the element

1 IRS1120LScheduleB. It took me almost one hour to write this style sheet because it had been
2 several years since I looked at style sheets and had to use Google to understand how to list the
3 namespaces that the IRS uses. Using a free open source program which comes on my computer
4 called xsltproc, I was able to specify the name of an input file, the name of the style sheet, and the
5 name of the output file. I ran that program and produced an XML file with the Schedule B
6 removed. It took 1.429 seconds to execute this command on my desktop computer. I ran this
7 program on a single instance of a Form 990, but this program could also be used, without
8 modification, to process hundreds or thousands of instances of the Form 990. It can also be easily
9 modified to remove multiple schedules.

10 32. Availability of returns in MeF format are significantly easier to work with than the
11 bitmap files produced by the IRS and shipped on DVDs. For my particular application, finding
12 Social Security Numbers in current returns, having the e-file data would have saved me a week of
13 initial processing of the data and would have found much more reliable results.

14 33. In addition to locating SSNs, the availability of the data in MeF format would
15 unlock a large number of other applications. For example, in order to find returns in our collection
16 of over 7.5 million Form 990s, computer programs must use a variety of search indices. With the
17 data the IRS currently provides, we know the name of the nonprofit and rudimentary information
18 such as the city, state, date of filing, and assets. If information were available in MeF format,
19 much more useful search capabilities would be possible using all of the data fields in the return to
20 help the public readily access the information that they desire.

21 34. Public.Resource.Org's request for Exempt Organization returns in MeF format
22 instead of bitmap images would be of substantial use to perform audits for privacy violations of
23 Exempt Organization returns. If the MeF format data were available, I would be able to notify the
24 IRS and other organizations with copies of this data more quickly about any breaches that were
25 discovered. In addition to finding privacy breaches, there would be a large number of other
26 beneficial applications in the public interest. It is my considered technical opinion, based on over
27 30 years as a computer professional, extensive work with the XML standard, and 6 years of
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

experience with the IRS Exempt Organizations database that it is technically very easy to make the MeF version of these public filings available.

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct and that this declaration was executed this 27 day of September, 2014 at Sebastopol, California.



CARL MALAMUD