No. 20-1191

UNITED STATES COURT OF APPEALS
FOR THE FOURTH CIRCUIT

WIKIMEDIA FOUNDATION,

Plaintiff–Appellant,

v.

NATIONAL SECURITY AGENCY, *et al.*,

Defendants–Appellees.

On Appeal from the United States District Court
for the District of Maryland at Baltimore

JOINT APPENDIX—VOLUME 7 OF 7 (JA3879–JA4124)

H. Thomas Byron III
Joseph Busa
Michael Shih
U.S. DEPARTMENT OF JUSTICE
950 Pennsylvania Ave. NW
Washington, DC 20530
Phone: (202) 616-5367
Fax: (202) 307-2551
h.thomas.byron@usdoj.gov

Counsel for Defendants–Appellees

Patrick Toomey
Ashley Gorski
Charles Hogle
AMERICAN CIVIL LIBERTIES
UNION FOUNDATION
125 Broad Street, 18th Floor
New York, NY 10004
Phone: (212) 549-2500
Fax: (212) 549-2654
ptoomey@aclu.org

Counsel for Plaintiff–Appellant
(Additional counsel on next page)

Alex Abdo
Jameel Jaffer
KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY
475 Riverside Drive, Suite 302
New York, NY 10115
Phone: (646) 745-8500
alex.abdo@knightcolumbia.org

Deborah A. Jeon
David R. Rocah
AMERICAN CIVIL LIBERTIES UNION FOUNDATION OF MARYLAND
3600 Clipper Mill Rd., #350
Baltimore, MD 21211
Phone: (410) 889-8555
Fax: (410) 366-7838
rocah@aclu-md.org

Benjamin H. Kleine
COOLEY LLP
101 California Street, 5th Floor
San Francisco, CA 94111
Phone: (415) 693-2000
Fax: (415) 693-2222
bkleine@cooley.com

***Wikimedia Foundation v. National Security Agency, et al.*,**
No. 20-1191 (4th Cir.)

JOINT APPENDIX
Table of Contents

VOLUME 1

U.S. District Court for the District of Maryland, Docket Sheet,
Case No. 1:15-cv-00662 ..... JA0001

Plaintiff Wikimedia Foundation's Amended Complaint
(June 22, 2015), ECF No. 72 ..... JA0036

Exhibits to Wikimedia Foundation's Motion to Compel

Declaration of Patrick Toomey, Counsel for Wikimedia Foundation
(Mar. 26, 2018), ECF No. 125-3 ..... JA0096

Exhibit 1: Chart Identifying Discovery Requests at Issue on
Wikimedia Foundation's Motion to Compel,
ECF No. 125-4..... JA0101

Exhibit 2: Wikimedia Foundation's Requests for Admission
and attachments (Nov. 7, 2017), ECF No. 125-5 ..... JA0118

Exhibits to Defendants' Opposition to Wikimedia Foundation's Motion to Compel

Declaration of Daniel R. Coats, Director of National Intelligence
(Apr. 28, 2018), ECF No. 138-2 ..... JA0170

Declaration of Lauren L. Bernick, Senior Associate Civil Liberties
Protection Officer in the Office of Civil Liberties, Privacy, and
Transparency at the Office of the Director of National Intelligence
(Apr. 28, 2018), ECF No. 138-3 ..... JA0190

Notice of Filing Unclassified & Redacted Version of the Declaration of George C. Barnes, Deputy Director of the NSA (May 11, 2018), ECF No. 141 ..... JA0199

Unclassified & Redacted Version of the Declaration of George C. Barnes, Deputy Director of the NSA (May 11, 2018), ECF No. 141-1 ..... JA0201

Exhibits to Wikimedia Foundation's Reply in Support of Its Motion to Compel

Declaration of Ashley Gorski, Counsel for Wikimedia Foundation (May 18, 2018), ECF No. 143-1 ..... JA0270

Exhibit 1: Chart Identifying Deposition Questions at Issue on Wikimedia Foundation's Motion to Compel, ECF No. 143-2 ..... JA0272

Exhibit 2: Transcript of Deposition of NSA's Designated Witness, Rebecca J. Richards, Pursuant to Fed. R. Civ. P. 30(b)(6) (Apr. 16, 2018), ECF No. 143-3 ..... JA0286

Opinion & Order Denying Wikimedia Foundation's Motion to Compel

Memorandum Opinion (Aug. 20, 2018), ECF No. 150 ..... JA0689

Order Denying Plaintiff's Motion to Compel Discovery Responses & Deposition Testimony (Aug. 20, 2018), ECF No. 151 ..... JA0716

Exhibits to Defendants' Motion for Summary Judgment

Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University (Nov. 13, 2018), ECF No. 164-4 ..... JA0719

Declaration of James Gilligan, Counsel for Defendants (Nov. 13, 2018), ECF No. 164-5 ..... JA0818

Exhibit 3: Wikimedia Foundation's Amended and Supplemental Responses and Objections to NSA's First Set of Interrogatories (Mar. 23, 2018), ECF No. 164-6 ..... JA0821

Exhibit 4: Wikimedia Foundation's Amended Responses and Objections to ODNI's Interrogatory No. 19 (Apr. 6, 2018), including Technical Statistics Chart, ECF No. 164-7 ..... JA0861

Exhibit 5: Wikimedia Foundation's Responses and Objections to NSA's First Set of Interrogatories (Jan. 11, 2018), ECF No. 164-8..... JA0876

VOLUME 2

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment

Declaration of Scott Bradner, Former Senior Technology Consultant for the Harvard University Chief Technology Officer (Dec. 18, 2018), ECF No. 168-2 ..... JA0920

Appendices A through Z to Declaration of Scott Bradner (Dec. 18, 2018), ECF Nos. 168-3 to 168-4 ..... JA1067

VOLUME 3

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Appendices AA through FF to Declaration of Scott Bradner (Dec. 18, 2020), ECF No. 168-5 ..... JA1791

Declaration of Jonathon Penney, Associate Professor at the Schulich School of Law and Director of the Law & Technology Institute at Dalhousie University (Dec. 18, 2018), ECF No. 168-6 ..... JA2151

Declaration of Michelle Paulson, Former Legal Director and Interim General Counsel for Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-7 ..... JA2218

Declaration of James Alexander, Former Manager for Trust and Safety and Former Legal and Community Advocacy Manager at Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-8 ..... JA2244

Declaration of Tilman Bayer, Senior Analyst for Wikimedia Foundation Product Analytics Team (Dec. 18, 2018), ECF No. 168-9 ..... JA2253

Declaration of Emily Temple-Wood (Dec. 18, 2018), ECF No. 168-10 ..... JA2268

Declaration of Patrick Toomey, Counsel for Wikimedia Foundation (Dec. 18, 2018), ECF No. 168-11 ..... JA2278

Exhibit 8: Wikimedia-hosted email list discussing NSA slide with Wikimedia logo, from July to August 2013, ECF No. 168-12 ..... JA2283

Exhibit 9: Wikimedia "Talk page" discussing its non-public information policy, from September to December 2013, ECF No. 168-13 ..... JA2305

Exhibit 10: "OTRS" ticket showing Wikimedia user requesting Tor permissions in September 2013, ECF No. 168-14 ..... JA2349

VOLUME 4

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Exhibit 11: Wikimedia webpage showing Wikimedia user requesting Tor permissions in September 2017, ECF No. 168-15 ..... JA2353

Exhibit 12: Wikimedia document compiling German-user-

community appeal concerning privacy in 2013,
ECF No. 168-16 ..... JA2357

Exhibit 13: Wikimedia "Talk page" discussing NSA
surveillance from June to December 2013,
ECF No. 168-17 ..... JA2363

Exhibit 14: Wikimedia Technical Statistics Chart & Supporting
Exhibits A-G, ECF No. 168-18 ..... JA2396

Exhibit 15: Privacy & Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of FISA* (July 2014), ECF No. 168-19 ..... JA2434

Exhibit 16: FISC Memorandum Opinion, [*Redacted*], 2011 WL
10945618 (Oct. 3, 2011), ECF No. 168-20 ..... JA2631

Exhibit 17: Office of the Director of National Intelligence, *DNI Declassifies Intelligence Community Documents Regarding Collection Under Section 702 of FISA* (Aug. 21, 2013),
ECF No. 168-21 ..... JA2717

Exhibit 18: Defendant NSA's Objections and Responses to
Plaintiff's First Set of Interrogatories (Dec. 22, 2017),
ECF No. 168-22 ..... JA2721

Exhibit 19: FISC Submission, *Clarification of National Security Agency's Upstream Collection Pursuant to Section 702 of FISA*
(May 2, 2011), ECF No. 168-23 ..... JA2743

Exhibit 20: Office of the Director of National Intelligence,
Statistical Transparency Report Regarding Use of National Security Authorities, Calendar Year 2017 (Apr. 2018),
ECF No. 168-24 ..... JA2748

Exhibit 21: FISC Memorandum Opinion & Order
(Apr. 26, 2017), ECF No. 168-25 ..... JA2790

VOLUME 5

Exhibits to Wikimedia Foundation's Opposition to Defendants' Motion for Summary Judgment (Cont'd)

Exhibit 22: FISC Submission, *Government's Response to the Court's Briefing Order of May 9, 2011* (June 1, 2011), ECF No. 168-26 ..... JA2890

Exhibit 23: *Big Brother Watch & Others v. United Kingdom*, App. Nos. 58170/13, 62322/14, 24960/15, Eur. Ct. H.R. (2018), ECF No. 168-27 ..... JA2932

Exhibit 24: NSA Director of Civil Liberties & Privacy Office, *NSA's Implementation of FISA Section 702* (Apr. 16, 2014), ECF No. 168-28 ..... JA3145

Exhibit 25: *Legal Issues Relating to the Testing, Use, and Deployment of an Intrusion-Detection System (EINSTEIN 2.0)*, 33 Op. O.L.C. 1 (Jan. 9, 2009), ECF No. 168-29 ..... JA3157

Exhibit 26: Minimization Procedures Used by the NSA in Connection with Acquisitions of Foreign Intelligence Information Pursuant to Section 702 of FISA (July 2014), ECF No. 168-30 ..... JA3193

Exhibit 27: Glenn Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet,"* Guardian, July 31, 2013, ECF No. 168-31 ..... JA3209

Exhibit 28: NSA slide, excerpted from Exhibit 27 (Greenwald, *XKeyscore: NSA Tool Collects "Nearly Everything a User Does on the Internet"*), ECF No. 168-32 ..... JA3220

Exhibit 29: Morgan Marquis-Boire, et al., *XKEYSCORE: NSA's Google for the World's Private Communications*, Intercept, July 1, 2015, ECF No. 168-33 ..... JA3222

Exhibit 30: NSA slide deck, *XKEYSCORE for Counter-CNE*, published in The Intercept on July 1, 2015, ECF No. 168-34 ... JA3237

Exhibit 31: Wikimedia, *Founding Principles* (accessed Mar. 14, 2018), ECF No. 168-35 ..... JA3259

Exhibit 32: Yana Welinder, *Opposing Mass Surveillance on the Internet*, Wikimedia Blog (May 9, 2014), ECF No. 168-36 ..... JA3262

Exhibit 33: Wikimedia Public Policy, *Privacy* (accessed Mar. 14, 2018), ECF No. 168-37 ..... JA3266

Exhibit 34: Wikipedia, *Sock Puppetry* (accessed Mar. 14, 2018), ECF No. 168-38 ..... JA3273

Exhibit 35: Wikimedia, *Privacy Policy* (accessed Feb. 14, 2018), ECF No. 168-39 ..... JA3286

Exhibit 36: Ryan Lane, *The Future of HTTPS on Wikimedia Projects*, Wikimedia Blog (Aug. 1, 2013), ECF No. 168-40 ..... JA3311

Exhibit 37: Yana Welinder, et al., *Securing Access to Wikimedia Sites with HTTPS*, Wikimedia Blog (June 12, 2015), ECF No. 168-41 ..... JA3317

Exhibit 38: Wikimedia email describing Tech/Ops goals and the importance of HTTPS (May 23, 2014), ECF No. 168-42 .... JA3325

Exhibit 39: Wikimedia document discussing IPsec implementation, including July 8, 2013 statement from a Wikimedia engineer, ECF No. 168-43 ..... JA3328

Exhibit 40: Wikimedia job posting for Traffic Security Engineer (accessed Feb. 8, 2018), ECF No. 168-44 ..... JA3364

Exhibit 41: Michelle Paulson, *A Proposal for Wikimedia's New Privacy Policy and Data Retention Guidelines*, Wikimedia Blog (Feb. 14, 2014), ECF No. 168-45 ..... JA3367

Exhibit 42: Wikimedia's Supplemental Exhibit C in response

to NSA Interrogatory No. 8 (volume of HTTP border-crossing communications by country), ECF No. 168-46 ..... JA3375

Exhibit 43: Wikimedia's Supplemental Exhibit D in response to NSA Interrogatory No. 8 (volume of HTTPS border-crossing communications by country), ECF No. 168-47 ..... JA3388

Exhibit 44: Wikimedia analytics document showing monthly unique visitors to Wikimedia by region, from December 2007 to May 2015, ECF No. 168-48 ..... JA3400

Exhibit 45: Press Release, NSA, *NSA Stops Certain Section 702 "Upstream" Activities*, Apr. 28, 2017, ECF No. 168-49 ..... JA3404

VOLUME 6

Exhibits to Defendants' Reply in Support of Their Motion for Summary Judgment

Second Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University (Feb. 15, 2019), ECF No. 178-2 ..... JA3407

Declaration of Alan J. Salzberg, Principal of Salt Hill Statistical Consulting (Feb. 15, 2019), ECF No. 178-3 ..... JA3452

Second Declaration of James Gilligan, Counsel for Defendants (Feb. 15, 2019), ECF No. 178-4 ..... JA3725

Exhibit 9: Wikimedia Foundation's Responses and Objections to DOJ's First Set of Interrogatories (Jan. 11, 2018), ECF No. 178-5 ..... JA3728

Exhibit 10: Relevant Portions of the Deposition of James Alexander, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6), ECF No. 178-6 ..... JA3761

Exhibit 11: Relevant Portions of the Deposition of Michelle

Paulson, Wikimedia Foundation witness taken pursuant to Fed. R. Evid. 30(b)(6), ECF No. 178-7 ..... JA3777

Exhibit 12: Wikimedia Foundation, *Securing access to Wikimedia sites with HTTPS*, June 12, 2015 (WIKI0007108-7114), ECF No. 178-8 ..... JA3791

Exhibit 13: Wikipedia: Village pump (technical)/Archive 138 (WIKI0006872-6938), ECF No. 178-9 ..... JA3800

Exhibit 14: Jimmy Wales and Lila Tretikov, "Stop Spying on Wikimedia Users," N.Y. Times, Mar. 10, 2015, ECF No. 178-10 ..... JA3869

Exhibit 15: Wikimedia Foundation, *Wikimedia v. NSA: Wikimedia Foundation files suit against NSA to challenge upstream mass surveillance*, Mar. 10, 2015, ECF No. 178-11 ..... JA3873

VOLUME 7

Exhibits to Wikimedia Foundation's Sur-reply in Opposition to Defendants' Motion for Summary Judgment

Second Declaration of Scott Bradner, Former Senior Technology Consultant for the Harvard University Chief Technology Officer (Mar. 8, 2019), ECF No. 181-1 ..... JA3879

Second Declaration of Jonathon Penney, Associate Professor at the Schulich School of Law and Director of the Law & Technology Institute at Dalhousie University (Mar. 8, 2019), ECF No. 181-2 ..... JA3940

Second Declaration of Michelle Paulson, Former Legal Director and Interim General Counsel for Wikimedia Foundation (Mar. 8, 2019), ECF No. 181-3 ..... JA4006

Second Declaration of Tilman Bayer, Senior Analyst for Wikimedia Foundation Product Analytics Team (Mar. 8, 2019), ECF No. 181-4 ..... JA4012

Second Declaration of Emily Temple-Wood (Mar. 8, 2019),
ECF No. 181-5 ..... JA4015

Exhibits to Defendants' Sur-reply in Support of Their Motion for Summary Judgment

Third Declaration of Henning Schulzrinne, Julian Clarence Levi Professor of Computer Science at Columbia University
(Mar. 22, 2019), ECF No. 182-2 ..... JA4019

Second Declaration of Alan J. Salzberg, Principal of Salt Hill Statistical Consulting (Mar. 22, 2019), ECF No. 182-3 ..... JA4048

Opinion & Order Granting Defendants' Motion for Summary Judgment

Memorandum Opinion (Dec. 16, 2019), ECF No. 188 ..... JA4073

Order Granting Defendants' Motion for Summary Judgment
(Dec. 16, 2019), ECF No. 189 ..... JA4123

Wikimedia Foundation's Notice of Appeal

Notice of Appeal (Feb. 14, 2020), ECF No. 191 ..... JA4124

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 1

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 1:15-cv-00662-TSE

REPLY DECLARATION OF SCOTT BRADNER

TABLE OF CONTENTS

I. MY CONCLUSION HAS NOT CHANGED. 1

II. INTRODUCTION 1

III. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY CONFLICTS WITH WHAT IS PUBLICLY KNOWN ABOUT UPSTREAM COLLECTION. 5

A. The four key foundations of my conclusion in my original declaration 5

1. Foundation 1: To conduct upstream collection of international public Internet communications traversing a circuit, the NSA must be copying, reassembling, and reviewing transactions on that circuit. 5
2. Foundation 2: Wikimedia communications are transported on all international circuits originating or terminating in the United States. 6
3. Foundation 3: The NSA conducts upstream collection on at least one international Internet link. 6
4. Foundation 4: The NSA is copying, reassembling, and reviewing Wikimedia communications on the international Internet links it is monitoring. 9

B. Why the NSA is copying, reassembling, and reviewing all communications on the international Internet links it is monitoring 10

1. The NSA has acknowledged that, on international Internet links it is monitoring, it does not apply IP filters. 10
2. The PCLOB has explained that upstream collection has been implemented in a technological manner designed to "comprehensively" acquire the communications of the NSA's targets. 13
3. Other technical and practical necessities make clear that the NSA is copying, reassembling, and reviewing Wikimedia's communications. 17

IV. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY ASSUMES THE IMPLAUSIBLE USE OF WHITELIST AND BLACKLIST FILTERS. 19

A. Whitelist and blacklist filters 19

1. Filters on international Internet circuits. 20
2. Whitelist filters are incompatible with the government's public descriptions of the upstream collection program. 20
   a. Whitelist filters are almost useless in the upstream collection program because using them would require that the NSA know unknowable information. 21
   b. Dr. Schulzrinne's proposed whitelisting is based on other assumptions or simplifications that are also inconsistent with what is known about the upstream collection program. 23
      i. *Number of NSA targets.* 24

ii. *Impossible targeting*.....25
iii. *Whitelist complexity and dynamism*.....27
c. Whitelist filters assume that the NSA wants to avoid almost all of the world's communications.....28
3. Blacklist filters are incompatible with the government's public descriptions of the upstream collection program.....29
a. Improbable non-targeting.....29
b. Blacklist filters would not guarantee that the NSA would avoid copying, reassembling, and reviewing Wikimedia communications.....31
4. Still copying packets.....33
5. Probing for blind spots.....33
B. "About" communications.....35

V. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY IS IMPLAUSIBLE FOR NUMEROUS OTHER TECHNICAL AND PRACTICAL REASONS.....37
A. Copy-then-filter vs. in-line filter.....37
1. Fiber-optic splitter.....38
2. Configuring the ISP router to mirror communications.....40
3. All packets are copied.....41
4. Configuring the ISP router to filter before mirroring.....41
B. Collecting "web activity".....44
C. Collecting web communications.....45
D. Collecting encrypted communications.....46
E. GCHQ surveillance.....48
F. ISP-operated copy device.....52
G. EINSTEIN.....52

VI. SUMMARY.....53

1. My name is Scott Bradner. I have been asked by the plaintiff's counsel in *Wikimedia Foundation v. National Security Agency*, No. 1:15-cv-006622-TSE (D. Md.), to provide an expert report addressing the government's reply to the plaintiff's brief and to my declaration, both of which were dated December 18, 2018. My qualifications to express an opinion in the case as well as my compensation and CV are as stated in my previous declaration.

2. A list of the documents provided to me by plaintiff's counsel was attached as Appendix B to the previous declaration.

I. MY CONCLUSION HAS NOT CHANGED.

3. I have carefully reviewed Dr. Schulzrinne's reply declaration as well as the government's reply brief. These documents do not change my basic conclusion in this case that "*it is virtually certain that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications*."

II. INTRODUCTION

4. My conclusion rests on four basic foundations:

   i. The NSA is copying packets, reassembling them into communications and then reviewing the communications for the presence of selectors as part of the upstream program.

   ii. Wikimedia's traffic traverses every circuit carrying public Internet traffic into and out of the country (i.e., "international internet links").

   iii. The NSA is monitoring at least one such circuit under the upstream collection program.

   iv. On any circuit it is monitoring, the NSA must be copying, reassembling, and reviewing transactions, including Wikimedia communications, to find those communications that are to or from its targets.

5. The government and Dr. Schulzrinne do not dispute the first two foundations, and they do not seriously dispute the third foundation. But they do dispute the fourth foundation.

6. Dr. Schulzrinne disputes the fourth foundation of my conclusion primarily by describing what I will call a "Wikimedia-avoidance theory"—a hypothetical architecture for an upstream collection program that intentionally avoids Wikimedia's communications (and potentially many other types of communications), rather than having as its goal comprehensively collecting the communications to and from the NSA's targets. This hypothetical architecture is deliberately designed *not* to be comprehensive—because it is designed to avoid entire categories of Internet communications on the off chance that there might be Wikimedia communications present. In offering his Wikimedia-avoidance theory, Dr. Schulzrinne is effectively ignoring the inescapable technical implications of the government's own descriptions of the upstream collection program. Dr. Schulzrinne does not cite any evidence in either of his declarations that the NSA is actually using the extensive filters he describes, nor does he cite any evidence that the NSA is actually avoiding every one of the billions of Wikimedia communications.

7. I disagree with Dr. Schulzrinne and believe that his hypothetical architecture is inconsistent with what the government has disclosed about the upstream collection program. His architecture conflicts with the government's definitive statement that the NSA "*will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being*

monitored by NSA."[1] Dr. Schulzrinne's hypothetical architecture also conflicts with the NSA's goal "*to comprehensively acquire communications that are sent to or from its targets*."[2] The architecture also conflicts with other technical and practical necessities of conducting a program that has collected millions of communications to or from tens of thousands of targets dispersed around the world. Each of these conflicts independently disproves Dr. Schulzrinne's speculation that the NSA is using his Wikimedia-avoidance theory in its upstream collection program, and each independently supports my conclusion concerning the NSA's monitoring of Wikimedia communications.

8. Dr. Schulzrinne argues that his hypothetical architecture, based on extensive whitelist and blacklist filters, does not conflict with government disclosures about the upstream collection program. The bulk of Dr. Schulzrinne's reply declaration is a set of nuanced discussions that are not relevant to the first two conflicts between his hypothetical architecture and the government's disclosures about the upstream collection program. He devotes little space to showing that his hypothetical architecture is consistent with the government's definitive admission that the NSA will acquire wholly domestic communications under some conditions or with the NSA's described goal to comprehensively acquire the communications to or from the NSA's targets. His few arguments relating to these two conflicts are either technically incorrect or consist basically of pleas to ignore the plain meanings of the government's disclosures.

[1] Appendix P (FISC Opinion at 45 (Oct. 3, 2011), *available at* ECF No. 168-4 at 562-643 ("FISC Opinion")).

[2] Appendix F (Privacy and Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of FISA* at 10 (July 2, 2014), *available at* ECF No. 168-3 at 199-395 ("PCLOB Report")).

9. As I explain below, Dr. Schulzrinne is also incorrect in his response to the other technical and practical necessities that support my conclusion.

10. Before getting to these explanations, I want to respond to a thread that runs through Dr. Schulzrinne's declaration. He insists that it is impossible for him or me to know the NSA's practices, priorities and capabilities, or the relative likelihood of different technical implementations of the upstream collection program.[3] While absolute assurance may be difficult, the NSA must operate in the real world and deal with the technical and operational limitations inherent in the Internet and in the telecommunications providers it compels to assist it. This need to operate in the real world constrains the ways the NSA could have implemented and be operating the upstream collection program and enables informed deduction of the NSA's actual implementation. Where different implementations present certain technical or practical trade-offs, I have tried to clearly state the degree of certainty or confidence I have in my conclusions.

11. The government has made numerous disclosures relating to the upstream collection program over the years. Many of these disclosures have specific technical implications. The descriptions of the upstream collection program in my first declaration as well as in this declaration are not speculative; instead they are based on the application of my technical expertise to analyze the government's disclosures to understand those technical implications.

[3] Schulzrinne Reply Decl. ¶ 3.

III. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY CONFLICTS WITH WHAT IS PUBLICLY KNOWN ABOUT THE UPSTREAM COLLECTION PROGRAM.

12. Dr. Schulzrinne's theory of how the NSA might be conducting its upstream collection program conflicts with what the government has publicly said about the upstream collection program. To see why, it is helpful to begin with my original declaration. The final conclusion that I reached in my original declaration was based on four key foundations. I provided support for each of the foundations, summarized here:

A. The four key foundations of my conclusion in my original declaration

1. *Foundation 1: To conduct upstream collection of international public Internet communications traversing a circuit, the NSA must be copying, reassembling, and reviewing transactions on that circuit.*

13. The first foundation for the final conclusion in my previous declaration was "*the NSA must be copying packets, reassembling them into communications and then reviewing the communications for the presence of selectors as part of the upstream program*."

14. I discussed this foundation in my original declaration.[4]

15. In summary, in order to determine if one or more selectors are present in a communication, the NSA must first copy the packets that make up a communication, then reassemble the packets into a copy of the communication. Only after the communication has been reassembled can the NSA review the contents of the communication to determine if the communication contains one or more selectors. The NSA must do this whether or not selectors are present.

[4] Bradner Decl. ¶¶ 250-320, ECF No. 168-2.

16. Neither the government nor Dr. Schulzrinne disputes the requirement that packets be copied and reassembled before they could be reviewed for the presence of selectors.

2. *Foundation 2: Wikimedia communications are transported on all international circuits originating or terminating in the United States.*

17. The second foundation for the final conclusion in my previous declaration was "*Wikimedia's international communications traverse every circuit carrying public Internet traffic on every international cable connecting the U.S. to other countries*" (i.e., "international internet links").

18. I discussed this foundation in my original declaration.[5]

19. In summary, considering the volume of Wikimedia's international communications and the fact that there are users of Wikimedia's U.S.-based services in all of the world's inhabited continents and islands, there must be Wikimedia communications traversing all of the international Internet links connecting the U.S. to the rest of the world.

20. Neither the government nor Dr. Schulzrinne disputes this foundation.

3. *Foundation 3: The NSA conducts upstream collection on at least one international Internet link.*

21. The third foundation for the final conclusion in my previous declaration was "*the NSA is monitoring at least one such circuit under the upstream collection program*."

22. I discussed this foundation in my original declaration.[6]

[5] Bradner Decl. ¶¶ 336-50.

[6] Bradner Decl. ¶¶ 150-153, 222-28, 260-64, 291, 331-35.

23. In summary, based on disclosures the government has made, the NSA is monitoring at least one international Internet circuit that is transporting Wikimedia communications as part of the upstream collection program.

24. Dr. Schulzrinne does not dispute this foundation. The government disputes this foundation and makes, in my opinion, an unpersuasive argument on pages 5 and 6 of their reply brief as to whether government disclosures confirm that the NSA has monitored at least one international Internet circuit.

25. I draw support for my conclusion that the NSA is monitoring at least one such circuit under the upstream collection program from:

a. the NSA's response to plaintiff's Interrogatory No. 12, in which the NSA acknowledges that the "Internet backbone" includes international Internet circuits or links that convey Internet traffic "*internationally via terrestrial or undersea circuits*,"[7]

b. the FISC Opinion of October 3, 2011, in which the FISC stated that "*the government readily concedes that NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA*,"[8]

[7] Appendix D (NSA Response to Plaintiff's Interrogatory No. 12, at 18 (Dec. 22, 2017), *available at* ECF No. 168-3 at 77-98).

[8] Appendix P (FISC Opinion at 45).

c. NSA representative Rebecca J. Richards's April 16, 2018 deposition, in which she testified that the "*will acquire*" sentence in the FISC Opinion "*is accurate*,"[9]

d. the PCLOB Report of July 2, 2014, which includes the same "*will acquire*" concession by the government,[10]

e. the NSA's 2014 targeting procedures, which make it clear that, at least in some circumstances, the NSA does not make use of an IP filter to discard wholly domestic communications,[11]

f. the explanation in my first declaration that it is logical and unsurprising that, in designing a program to intercept international Internet communications, the NSA would monitor international Internet links.[12]

26. Taken individually and together, these references make it clear that the NSA has monitored at least one international Internet circuit. Because Wikimedia communications are present on all international Internet circuits, the NSA has monitored at least one international Internet circuit that carries Wikimedia communications.

[9] Appendix K (Transcript of Deposition of Rebecca J. Richards 160:4-17 (Apr. 16, 2018), *available at* ECF No. 168-4 at 105-507 ("Richards Dep.")).

[10] Appendix F (PCLOB Report at 41 n.157).

[11] Appendix T (NSA Section 702 Targeting Procedures at 2 (2014), *available at* ECF No. 168-4 at 1062-1071).

[12] Bradner Decl. ¶¶ 222-24, 293, 332.

4. *Foundation 4: The NSA is copying, reassembling, and reviewing Wikimedia communications on the international Internet links it is monitoring.*

27. The fourth foundation for the final conclusion in my previous declaration was that on any circuit it is monitoring, the NSA must, for a variety of technical reasons, be copying, reassembling, and reviewing all transactions, including Wikimedia communications, to find those communications that are to or from (or about) its targets

28. I discussed this foundation throughout my original declaration.[13]

29. In summary, based on the foundations discussed above, Wikimedia international communications will be transported over at least one of the international Internet circuits the NSA is monitoring. As I explain in detail below, there are several independent reasons why, in the process of monitoring such a link, the NSA must be copying, reassembling, and reviewing at least all international communications transported on the link.

30. Both the government and Dr. Schulzrinne dispute this foundation. Dr. Schulzrinne maintains that the NSA could be using whitelist filters (that is, filters that enumerate the specific IP addresses and/or protocols the NSA wants to review) and/or blacklist filters (that is, filters that enumerate specific IP addresses and/or protocols that the NSA does not want to review) to avoid copying, reassembling and reviewing Wikimedia communications. I disagree that the use of such filters would be technologically consistent with the government's public descriptions of the upstream collection program, and I also disagree that the use of such filters would avoid Wikimedia communications. I will discuss my objections below.

[13] Bradner Decl. ¶¶ 36-48, 272-89, 293-94, 301-18, 333, 335.

31. I will focus on Dr. Schulzrinne's objections, leaving it to counsel to address any of the government's objections that are not based on Dr. Schulzrinne's objections.

B. Why the NSA is copying, reassembling, and reviewing all communications on the international Internet links it is monitoring

32. My conclusion that the NSA is copying, reassembling, and reviewing all communications on at least some of the circuits it is monitoring is supported by at least three independent bases. Each of these bases shows that it is a virtual certainty that the NSA is copying, reassembling, and reviewing Wikimedia's communications. I will now provide a short explanation of each of these bases and of the way in which Dr. Schulzrinne's Wikimedia-avoidance theory conflicts with what is publicly known about the upstream collection program.

1. *The NSA has acknowledged that, on international Internet links it is monitoring, it does not apply IP filters.*

33. The government has disclosed that it does not always apply filters to the traffic on circuits it is monitoring. On at least some circuits, the government has acknowledged that it does not rely on filters to eliminate wholly domestic communications.[14]

34. The government has been quite clear about one of the circumstances in which it does not apply filters. It has acknowledged that it does not rely on filters on the international links it is monitoring.[15]

[14] Bradner Decl. ¶ 291.

[15] Bradner Decl. ¶¶ 292-300.

35. In particular, the NSA has acknowledged to the FISC that it "*will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA*."[16] Note that this disclosure does not say "*may acquire*" or "*might acquire*." Instead, the statement is definitive—the NSA *will* acquire wholly domestic about communications, i.e. wholly domestic communications that include one or more selectors, if they are routed through an international Internet link the NSA is monitoring.[17]

36. Dr. Schulzrinne quotes this statement in ¶ 56 of his declaration. But two paragraphs later he seems to dismiss the categorical nature of the FISC statement and says "*wholly domestic communications of the kinds described above* ***could*** *still be copied and scanned by the NSA*" (emphasis added).[18]

37. There are multiple reasons to believe that the FISC statement should be taken at face value.

38. **Reason for the FISC Opinion**. Considering the circumstances that led to the decision that this quote is part of, it is hard to imagine that the FISC is not being as precise as it possibly could be. The Opinion was the result of a sequence of multiple hearings that were called after the FISC learned of previously undisclosed surveillance activity by the NSA. The hearings involved multiple submissions to the FISC explaining the details of the surveillance activity. It is clear from this that the FISC considered the

[16] Appendix P (FISC Opinion at 45).

[17] Bradner Decl. ¶¶ 292-294.

[18] Schulzrinne Reply Decl. ¶ 58.

situation to be very important and deserving of a very careful Opinion. Thus, there is every reason to believe that the FISC was very careful in what it wrote.

39. **Precise use of language**. It is clear that the FISC was purposeful in its choice of the phrase "*will acquire*." The FISC used a different, less emphatic phrase in a different circumstance only a few paragraphs away.[19]

40. **Richards deposition**. NSA representative Rebecca J. Richards confirmed the sentence in the FISC Opinion was accurate during her April 16, 2018 deposition.[20]

41. The excerpt from the FISC Opinion specifically discusses "about" collection, but that discussion shows that the NSA does not employ IP filters at least in some circumstances.

42. As a technological matter, the only way the NSA *will* acquire wholly domestic "about" communications on the international Internet links it is monitoring, as the government has disclosed, is if it is not applying any filters to that traffic before reviewing the communications to see if they contain one or more selectors. If the NSA were applying, for example, a filter that discarded web (ports 80 and 443) communications, the NSA would miss web-based "about" communication, which would not be consistent with the FISC's statement that all "about" communications "*will*" be acquired. Therefore the NSA must not be applying any filters at these locations. Thus, even in the improbable case where the NSA were deploying the whitelist and blacklist filters of Dr. Schulzrinne's Wikimedia-avoidance theory, based on the FISC Opinion, the filters would not be deployed at the international Internet links being monitored by NSA.

[19] Appendix P (FISC Opinion at 35).

[20] Appendix K (Richards Dep. 160:4-17).

43. The FISC disclosure concerning wholly domestic "about" communications is consistent with other government disclosures that IP filters are not always used, as discussed above. The lack of all filters on these international Internet links means that the NSA would copy, reassemble and review all communications on the link, including all Wikimedia communications that happen to be on the link, for the presence of selectors. This is how the NSA would find the "about" communications. Note that, since it is undisputed that there are Wikimedia communications on every international Internet circuit, there will be Wikimedia communications on any such international Internet circuit that the NSA is monitoring.

44. The definitive statement in the FISC Opinion does not provide any room for any filters, such as the whitelist or blacklist filters Dr. Schulzrinne hypothesizes could be used, because the use of such filters would discard some wholly domestic about communications.

45. Thus, at least on the international Internet links—where it does not employ filters—the NSA must be copying, reassembling and reviewing all communications, including any Wikimedia communications that traverse the link, in order to determine which communications contain targeted selectors.

2. *The PCLOB has explained that upstream collection has been implemented in a technological manner designed to "comprehensively" acquire the communications of the NSA's targets.*

46. The government has said that the aim of the NSA is "*to comprehensively acquire communications that are sent to or from its targets*."[21]

[21] Appendix F (PCLOB Report at 10, 123, 143); Bradner Decl. ¶ 333.

47. Dr. Schulzrinne summarily dismisses the government's own statement as an unrealistic or unrealized goal. He also implies that we should not accept that the PCLOB meant what it said when it used the term "*comprehensively*."[22]

48. I find no reason to doubt the PCLOB's use of the word "*comprehensively*," which it used in describing the technical reasons why the NSA collects "about" communications. Specifically, the PCLOB states that this collection is "*an inevitable byproduct of the government's efforts to comprehensively acquire communications that are sent to or from its targets*."[23] A corollary statement is that the NSA would not be collecting "about" communications if it was not striving to be comprehensive in its collection of the communications of its targets. The PCLOB meant the term "*comprehensively*" to explain the need for the NSA's specific technological implementation of the upstream collection program, and so I find it an appropriate basis on which to explain the technological implementation of the upstream collection program.

49. Even if Dr. Schulzrinne were correct that the NSA is not being comprehensive in which international Internet circuits it is monitoring, it would not follow that the monitoring on the circuits it *does* monitor would not be comprehensive. In fact, the FISC's description of the improper collection of "about" communications only makes sense if the NSA were comprehensively monitoring individual circuits.

[22] Schulzrinne Reply Decl. ¶¶ 72-74.

[23] Appendix F (PCLOB Report at 10).

50. It may be useful to note that the NSA worked "*extensively*" with PCLOB while the PCLOB was preparing the Report to ensure the Report's accuracy.[24] Because of the obvious care that was taken in preparing the Report, it is appropriate to understand the term "*comprehensively*" literally, as the PCLOB obviously intended it to be taken. It is also useful to note that Ms. Richards confirmed in her deposition that the NSA reviewed every sentence of the PCLOB Report, that it identified any inaccuracies in the Report to the PCLOB, and that it does not believe there to be any inaccurate statements of fact in the Report.[25]

51. The use of whitelist and/or blacklist filters as Dr. Schulzrinne hypothesizes is incompatible with a goal of comprehensively acquiring communications that are to or from NSA targets. Such filters work against a goal of being comprehensive. Any such filter inevitably discards communications that might include communications that are to or from the NSA's targets. For example, Dr. Schulzrinne hypothesizes a blacklist filter that would discard traffic to or from Wikimedia servers.[26] He imagines that such a blacklist would discard all communications accessing information on Wikimedia websites by an NSA target. As Dr. Schulzrinne states, I do not know the NSA's surveillance priorities, practices, and capabilities insofar as they are unstated or not inferable based on the NSA's extensive public disclosures.[27] But any such blacklist would, by definition, be incompatible with the government's *stated* goal of completeness.

[24] Appendix K (Richards Dep. 105:20-106:13); Appendix F (PCLOB Report at 3-4).

[25] Appendix K (Richards Dep. 101:22-102:5, 105:7-12, 105:20-106:13, 107:1-5, 108:11-15, 145:9-12).

[26] Schulzrinne Reply Decl. ¶¶ 12-13, 39-42.

[27] Schulzrinne Reply Decl. ¶ 3.

52. The use of whitelist filters, as imagined by Dr. Schulzrinne, is even less compatible with the concept of completeness. At least with a blacklist, one is specifying the relatively few addresses or protocols that are not of interest among the billions of possible addresses and thousands of possible protocols. In that case, communications to or from unknown addresses or using unknown protocols will still be reviewed. With a whitelist, however, one is specifying the relatively few addresses or protocols that are of interest. This means that communications to or from most of the billions of possible addresses or using most of the thousands of possible protocols will be discarded and not reviewed to see if they are from or to NSA targets. It is certainly technically possible to design an Internet surveillance program using a whitelist, but doing so would purposefully ignore most of the Internet, and it would be inconsistent with the publicly known details about the upstream collection program.

53. If the NSA were using a whitelist, or even a blacklist, of the sort that Dr. Schulzrinne speculates, the PCLOB would not have been able to say that the NSA's goal was "*to comprehensively acquire communications that are sent to or from its targets*." I will discuss a number of other flaws in Dr. Schulzrinne's concept of using whitelist and/or blacklist filters below.

54. Based on the technical detail in the PCLOB Report, the NSA must be comprehensively copying, reassembling, and reviewing all communications on a circuit that it is monitoring if the NSA is to meet its goal "*to comprehensively acquire communications that are sent to or from its targets*." This goal is technologically incompatible with any significant use of filters other than those IP filters that ensure that at least one end of a communication is outside the U.S., and, as discussed above, the

government has disclosed that IP filters are not used on international Internet circuits. Since there are Wikimedia communications on every international Internet circuit, there will be Wikimedia communications on any international Internet circuit that the NSA is monitoring and, if the NSA is comprehensively copying, reassembling, and reviewing all packets on circuits that it is monitoring so that it can comprehensively acquire communications that are sent to or from its targets, the NSA will be copying, reassembling, and reviewing Wikimedia communications.

3. *Other technical and practical necessities make clear that the NSA is copying, reassembling, and reviewing Wikimedia's communications.*

55. There is no reliable way for the NSA to know if an individual packet on a circuit is part of a communication that contains one or more selectors without reviewing a reassembled communication containing that packet. Thus, the NSA cannot know in advance which packets need to be copied because they are part of a communication that contains selectors. Thus, the NSA must be copying, reassembling, and reviewing at an absolute minimum those communications it wishes to scan for the presence of selectors.[28]

56. The NSA could not be making use of extensive whitelist and/or blacklist filters, such as the Wikimedia-avoidance architecture Dr. Schulzrinne imagines, because of the technical inability to know in advance which packets on a circuit are part of communications to or from the NSA's targets. Dr. Schulzrinne speculates that the NSA could know its targets' IP addresses and communications protocols in advance, but as I explain at length in the next section, that is not possible given all that we know about the scale and purposes of the upstream collection program.

[28] Bradner Decl. ¶¶ 236-48, 30-18, 333, 335.

57. The only way that the upstream collection program could possibly avoid all of Wikimedia's ubiquitous communications is if the NSA had actively strived to eliminate them, and Dr. Schulzrinne presents no evidence that the NSA has ever attempted to do so or any plausible explanation for why it would do so. Even setting this fact aside, the whitelist and blacklist filters Dr. Schulzrinne imagines would not, in fact, guarantee that the NSA would be able to avoid all Wikimedia communications. Even if they could do so, there is no plausible technical or practical reason why the NSA actually would want to avoid all Wikimedia communications. Some Wikimedia communications, for example, those communications that could reveal what the NSA's foreign intelligence targets are reading and writing on Wikimedia websites such as Wikipedia, would provide information that the NSA could consider to be potentially of interest. In any case, it is almost inconceivable that the NSA went out of its way to try to specifically ensure that upstream collection would not encounter even a single Wikimedia communication.

58. Dr. Schulzrinne's Wikimedia-avoidance architecture is pure speculation, unsupported by any of the government's disclosures concerning the upstream collection program or by any plausible technical or practical consideration. My explanations of the upstream collection program are based on the public record about it and on my expert analysis of that record and of the technology of Internet surveillance. Dr. Schulzrinne's theory, in contrast, is merely a thought experiment, conducted without any consideration of whether the architecture he has imagined would remotely satisfy the purposes of the upstream collection program.

59. I will address Dr. Schulzrinne's responses to this basis for my opinion in much greater detail in the next section.

60. The above three bases show that the NSA is not employing the types of ubiquitous whitelist or blacklist filters that Dr. Schulzrinne imagines.

IV. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY ASSUMES THE IMPLAUSIBLE USE OF WHITELIST AND BLACKLIST FILTERS.

61. As I explained above (¶¶ 32-60), there are three independent technological bases that support my final conclusion in this case. Dr. Schulzrinne and the government focus most of their attention on the third basis—that technical and practical necessities make clear that the NSA is copying, reassembling, and reviewing Wikimedia's communications—and do not seriously challenge the first two bases. In the previous section, I focused on the first two bases, and in this section I will mostly focus on Dr. Schulzrinne's responses to my third basis. I will also address at least some of the points he attempts to make concerning fine points in the other bases.

62. Most of Dr. Schulzrinne's response to my third basis involves his Wikimedia-avoidance theory for the upstream collection program—an architecture based on whitelist and blacklist filters.

A. Whitelist and blacklist filters

63. Dr. Schulzrinne's declaration largely focuses on the proposition that the NSA could use whitelist and/or blacklist filters to limit the scope of the upstream collection program at least enough to avoid copying, reassembling, and reviewing Wikimedia communications but still be compatible with government disclosures on the operation of the upstream collection program.[29] He also says that the use of such filters would ensure that Wikimedia communications were not among the communications the

[29] See, e.g., Schulzrinne Reply Decl. ¶¶ 50-51.

NSA copies, reassembles and reviews in the process of searching for communications to or from its targets.

64. I disagree on both suggestions. The use of some filters may be compatible with the government disclosures that discuss the use of IP filters to discard wholly domestic communications, but the use of whitelist or blacklist filters is not compatible with the disclosures such as the ones in the FISC Opinion and PCLOB Report that describe monitoring without the use of IP filters.[30] As I explain below, there are other reasons why whitelisting and blacklisting filters are technologically incompatible with what is publicly known about the upstream collection program, and anyway, such filters would not reliably avoid Wikimedia communications.

1. *Filters on international Internet circuits.*

65. Even if it were the case that the NSA was making use of whitelist and/or backlist filters in some circumstances, for example where filters are also used to discard wholly domestic communications, it does not follow that the NSA would add special whitelist or blacklist filters where it is not using IP filters, such as international Internet circuits as noted in the FISC Opinion. (See above at ¶¶ 33-45.)

2. *Whitelist filters are incompatible with the government's public descriptions of the upstream collection program.*

66. As noted above, a whitelist filter is a filter that enumerates the specific IP addresses and/or protocols that the NSA wants to review. All incoming packets that do not have an IP source or destination address that matches an IP address in the filter list, or are using a protocol that is not listed in the filter list of protocols, will be discarded.

[30] Appendix P (FISC Opinion at 45); Appendix F (PCLOB Report at 36-37).

67. There are multiple reasons that the use of whitelist filters would be incompatible with the public descriptions of the NSA's upstream collection program.

a. Whitelist filters are almost useless in the upstream collection program because using them would require that the NSA know unknowable information.

68. As I said in my previous declaration: Whitelisting requires knowing in advance all of the IP addresses that might be used by each of the NSA's targets as well as assuming that those targets are not moving around and thereby changing their IP addresses.

69. Basically, the underlying assumption inherent in the use of whitelist filters is that the NSA has up to date, comprehensive and accurate information on where its targets will be, what sites they will be communicating with and what protocols they will be using in advance of the start of any such communications.[31] This assumption would be impossible to meet for communications to or from the NSA's targets, but even harder to meet for "about" communications because the NSA would have to (1) know which non-targets will be talking about targets and (2) have comprehensive and accurate information on which IP addresses these non-targets will be using, what sites they will be communicating with, and what protocols they will be using, in advance of the start of any such communications. In other words, the use of whitelist filters involves an assumption of precognition.

70. Dr. Schulzrinne seems to think that developing and maintaining whitelists is easy, but it would be virtually impossible to do so for a surveillance program meant to capture the communications of thousands of targets, and in fact impossible to do so for a

[31] Bradner Decl. ¶ 366(d).

program meant to capture the communications of *unknown non-targets* about targets.[32] The examples he provides for an IP address-based whitelist include unspecified individual IP addresses or blocks of IP addresses,[33] the IP addresses of VPN and e-mail servers,[34] and the IP addresses of selected web servers, webmail or chatroom sites.[35] Dr. Schulzrinne's list focuses on servers rather than clients. This makes sense in the same way as looking under a streetlight to find your lost keys makes sense. As long as the NSA's targets use these servers, the NSA will intercept the targets' communications. But there are billions of IP addresses and countless e-mail, web, and other servers in the world, and it is trivial to set up even more of these kinds of servers. Anyone can set up new servers of the kinds Dr. Schulzrinne lists. For example, I have both an e-mail server and a web server in my house. They were easy to set up. If the NSA were restricting itself to the IP addresses of known servers, it would be deliberately foreclosing its ability to capture large amounts of target traffic. Communications making use of new or temporary servers, including those that have been temporarily set up to facilitate communication by NSA targets, would escape the NSA's upstream collection program.

[32] Dr. Schulzrinne speculates that the NSA might obtain "about" communications exchanged by individuals using whitelisted IP addresses (Schulzrinne Reply Decl. ¶ 58), but the PCLOB has made clear that "about" collection permits the NSA to acquire communications between entirely unknown non-targets (Appendix F (PCLOB Report at 121, 126)). It is not possible to whitelist the IP addresses of unknown non-targets in way that would reliably acquire about communications traversing circuits being monitored by the NSA.

[33] Schulzrinne Reply Decl. ¶¶ 8, 43.

[34] Schulzrinne Reply Decl. ¶ 58.

[35] Schulzrinne Reply Decl. ¶¶ 35, 37.

71. Note that Dr. Schulzrinne admits that an IP whitelist that does not include Wikimedia IP addresses would not exclude Wikimedia communications if someone using a whitelisted address communicated with Wikimedia. [36]

72. In theory, protocol-based whitelist filters would not be quite as useless as IP address-based ones for conducting upstream-*style* collection. But, as I pointed out in my previous declaration, there is nothing that restricts Internet users to using assigned port numbers for their applications.[37] Thus, protocol-based whitelist filters could easily miss a lot of communications that the NSA would otherwise want to review, including those using new, non-public or ad hoc protocols—for example, ad hoc protocols used to facilitate the communications of the NSA's targets.

73. That said, the only example Dr. Schulzrinne provides for a protocol whitelist is one for web protocols.[38] From government disclosures, it is already known that the NSA copies, reassembles and reviews web communications, so it is clear that the NSA is not using protocol-based whitelist (or blacklist) filtering to exclude the web protocols.[39] (See ¶ 130 below.)

b. Dr. Schulzrinne's proposed whitelisting is based on other assumptions or simplifications that are also inconsistent with what is known about the upstream collection program.

74. Dr. Schulzrinne makes a number of significant assumptions or simplifications as he argues that the NSA could be using whitelists and blacklists to

[36] Schulzrinne Reply Decl. ¶ 12.

[37] Bradner Decl. ¶ 109.

[38] Schulzrinne Reply Decl. ¶ 10.

[39] Bradner Decl. ¶¶ 314-315.

implement its upstream collection program without copying, reassembling or reviewing Wikimedia communications.

i. *Number of NSA targets.*

75. As I explained in my last declaration, one reason that whitelisting of the sort Dr. Schulzrinne describes is not remotely possible for the upstream collection program has to do with the number and mobility of the NSA's targets.[40] Dr. Schulzrinne responds by implying that the NSA might only have a few targets for the upstream collection program. As he points out, the government has indicated that the NSA has over 120,000 Section 702 targets, but has not stated explicitly that all of these targets are part of the upstream collection program.[41]

76. While that is true, there must be a significant number of upstream collection program targets or they must be prolific communicators—the government has disclosed that 26 million communications were collected under the upstream collection program in 2011.[42] For example, if there were only a thousand upstream collection targets, they would have to average 26,000 communications captured under the upstream collection program each per year in order for the NSA to have collected 26 million communications per year. While the actual number has not been publicly released, based on how many communications were collected in 2011 there are almost certainly tens of thousands of targets for the upstream collection program.

[40] Bradner Decl. ¶ 366(d).

[41] Bradner Decl. ¶ 334; Schulzrinne Reply Decl. ¶ 46.

[42] See, e.g., Appendix F (PCLOB Report at 37); Appendix P (FISC Opinion at 26, 30-34, 73).

ii. *Impossible targeting.*

77. Another reason that whitelisting of the sort Dr. Schulzrinne describes is not remotely possible for the upstream collection program has to do with the requirement that the NSA know, in advance, the IP addresses of its targets or the services the targets are using, even when the targets move around and their IP addresses change.[43] In particular, it would also, as a general rule, be difficult to identify IP addresses that are exclusively used by particular individuals or groups, including NSA targets, which would be required if the aim is to limit the possible copying, reassembly and review to NSA targets.

78. In addition, not all communications with targets will contain a target's IP address. For example, the IP addresses of targets do not appear in the communications when a target is using an intermediary or a communications service that involves multiple hops.[44] Nor do IP addresses associated with targets appear in "about" communications (¶¶ 108-109). In the former case, a whitelist would miss communications to or from the target, and in the latter case, communications about a target. If the whitelist included the IP addresses of intermediary or communications services that involve multiple hops, it would sweep in the communications of all other users of the services, which could include Wikimedia communications.

79. Dr. Schulzrinne argues that the movements of the NSA's targets could be limited to a "*given geographical area*." He notes that the NSA could use a whitelist that

[43] See, e.g., Bradner Decl. ¶¶ 171, 173-74, 229-30, 366(d).

[44] Bradner Decl. ¶¶ 244-7; Appendix P (FISC Opinion at 33-35).

includes a "*set of IP addresses . . . associated with geographical areas where the target is believed to be located*."[45]

80. The use of ranges of IP addresses in a whitelist can simplify the whitelist creation and maintenance, but their use would broaden the range of addresses that would be accepted by the whitelist. Considering the broad geographic distribution of Wikimedia users, the broader the range the more likely that the whitelist would result in the copying and review of Wikimedia communications.

81. Dr. Schulzrinne's suggestion that target movements could be limited to a given geographic area is theoretical at best. There is no particular reason to think that the NSA targets are so limited.

82. But, even if the target's mobility were limited, that does not mean that their use of the Internet would be restricted to any particular range of IP addresses. Because a target could use different ISPs at different times and at different nearby locations and, because the IP addresses that ISPs use are not geographically assigned, a target could move from IP address range to IP address range as they moved around.[46] Moreover, it is reasonable to infer that the NSA's targets are widely distributed across a number of geographic regions, given that the foreign intelligence, counterterrorism, weapons proliferation, and cyber-security uses of Section 702 surveillance implicate foreign governments, organizations, and actors around the world, so that the IP address ranges would be numerous and varied.[47]

---

[45] Schulzrinne Reply Decl. ¶ 47.

[46] Bradner Decl. ¶¶ 159-60.

[47] *'Section 702' Saves Lives, Protects the Nation and Allies*, NSA/Central Security Service (Dec. 12, 2017), https://perma.cc/3JAL-WVV2 ("*'Section 702' Saves Lives*").

83. In general, I do not think it is possible to reliably predict how a user's IP address may change over time or as they move, much less the IP addresses of thousands of users. This is made even more complicated by the fact that the actual users of a given IP address can change over time.

84. Dr. Schulzrinne also suggests that the NSA could be "*whitelisting the IP addresses of websites, webmail services, and/or chatrooms of interest.*" I will note that a webmail service is a website. In addition, many websites and services are now making use of content distribution networks, which, by design, can have different, and changing, IP addresses in different parts of the world. Many websites and services are also making use of cloud-based services, such as Amazon AWS, which also can have multiple and changing IP addresses. Keeping track of the set of IP addresses in use by a particular service at any point in time is, at best, difficult.

iii. *Whitelist complexity and dynamism.*

85. As discussed above in ¶ 84, Dr. Schulzrinne imagines that the NSA could get by with a very simple set of whitelist rules. But he does not mention the fact that the rules would have to be frequently updated as NSA targets were added or removed, or as they changed their locations or methods of operation. He paints a picture of only having to list the IP addresses of some servers along with some IP addresses of a few individuals and ranges of IP addresses.

86. Dr. Schulzrinne implies that the targets of the NSA's upstream collection program do not include individuals.[48] The government's own public documents indicate that the targets include individuals.[49]

[48] Schulzrinne Reply Decl. ¶ 47.

87. Dr. Schulzrinne suggests the NSA could be using whitelists that do not include Wikimedia addresses.[50] A whitelist that included all of the possible IP addresses or IP address ranges that NSA targets could be using yet excluded Wikimedia addresses would be very large indeed since Wikimedia is using only a handful of the approximately 4 billion possible Internet addresses. (There are a bit more than 4 billion possible IP version 4 addresses, as well as billions and billions of times more IP version 6 addresses. For this declaration I will focus on IP version 4 addresses because those are the addresses in most common use.)

88. The only way that a whitelist used in the upstream collection program could be simple while still ensuring that the NSA is comprehensively collecting communications to and from its targets is if the whitelist included most non-U.S. IP addresses and most protocols, in which case there is little reason to have a whitelist in the first place. Of course, any such whitelist would unquestionably include IP addresses and protocols used by Wikimedia users and thus would unquestionably include Wikimedia communications.

c. Whitelist filters assume that the NSA wants to avoid almost all of the world's communications.

89. If the NSA is using a whitelist filter, it means that the NSA is only interested in the people and processes it already knows about and that it has decided to actively ignore everything else. The use of a whitelist in the NSA upstream collection program would be the equivalent of deciding to only look at the material coming through

[49] See, e.g., Office of the Director of National Intelligence, *Section 702 Overview* at 5-7, https://perma.cc/J9X6-YME6 ("*Section 702 Overview*"); *'Section 702' Saves Lives*, *supra* note 47: Appendix F (PCLOB Report at 36); Appendix N (FISC Submission at 4 (Aug. 16, 2011)).
[50] Schulzrinne Reply Decl. ¶ 12.

a few select holes of a sieve. Since whitelisting is specifically designed to ignore most of the Internet, it would be extraordinarily easy for the NSA's targets to avoid being monitored.

3. *Blacklist filters are incompatible with the government's public descriptions of the upstream collection program.*

90. As noted above, a blacklist filter is a filter that enumerates the specific IP addresses and/or protocols that the NSA does not want to review. All incoming packets that have an IP source or destination address that matches an IP address in the filter list or is using a protocol that is listed in the filter list of protocols will be discarded.

91. The government has disclosed that the NSA does impose one type of blacklist filter in at least some circumstances. Wholly domestic communications are filtered out, at least at some—but not all—locations (¶¶ 33-45). This type of filtering can be done with a blacklist that discards packets whose source and destination IP addresses are both within the U.S.

92. There are multiple reasons that any additional use of blacklist filters would be incompatible with the public descriptions of the NSA's upstream collection program and why the use of blacklist filters would not mean that the NSA was avoiding Wikimedia communications. I will now review them.

a. Improbable non-targeting.

93. Dr. Schulzrinne suggests that the NSA could be using blacklist filters to avoid all communications that are to or from the IP addresses of Wikimedia servers.[51] In hypothesizing that the NSA is using such a filter, Dr. Schulzrinne presupposes that the

[51] See, e.g., Schulzrinne Reply Decl. ¶ 12.

NSA would have had a reason to deliberately avoid Wikimedia communications in the upstream collection program.[52] This is different than in a whitelist filter where the NSA is deciding which communications it wants to look at.

94. Dr. Schulzrinne does not provide any evidence, let alone any creditable reason that the NSA would have specifically decided that it did not want to include Wikimedia communications in the upstream collection program. It would be difficult for him to do so because, as he points out, he does not have any specific knowledge of the NSA's priorities. But, based on the available public information about the operation of the upstream collection program and the program's intelligence-gathering purpose, I find it impossible to infer that the NSA would have singled out Wikimedia communications, among the vast array of communications on the Internet, as communications that should be ignored.

95. Dr. Schulzrinne does talk about reducing the load on the devices the NSA is using in the upstream collection program, but he does not indicate why a desire to reduce load would have led the NSA to exclude Wikimedia communications.[53]

96. Dr. Schulzrinne notes that the inter-regional Internet capacity is very large these days—as much as 295 terabits per second.[54] To put the volume of Wikimedia communications into context, in the six-month period between August 1, 2017 and January 31, 2018, Wikimedia engaged in approximately 760 billion international communications.[55] This works out to about 48 thousand communications per second.

---

[52] Bradner Decl. ¶ 367(a).

[53] Schulzrinne Reply Decl. ¶¶ 20-22.

[54] Schulzrinne Reply Decl. ¶ 20.

[55] Bradner Decl. ¶ 346.

Assuming that an average Wikimedia communication is 81 packets,[56] and that the average packet length is 1,500 8-bit bytes, that means Wikimedia communication averages about 47 Gbps per second. Thus, Wikimedia represents about 0.016% of the inter-regional Internet capacity. Even if I were way off in my estimate of an average communication length of 81 packets, the percentage of inter-regional Internet capacity, measured in bits per second, represented by Wikimedia communications is still extraordinarily low. Blacklisting the Wikimedia IP addresses would not make any measurable difference to the load experienced by the NSA's upstream collection system. Thus, it is very unlikely that the NSA would have decided to specifically blacklist Wikimedia communications to reduce the load on the upstream collection program systems, even if, as Dr. Schulzrinne suggests, it might be easy to do.[57] Note also that although Dr. Schulzrinne says that, according to a statistics website, one of the Wikimedia websites is the 5th most popular website globally, it does not follow that Wikipedia generates a significant amount of traffic measured in bits per second compared to the inter-regional Internet capacity.

b. Blacklist filters would not guarantee that the NSA would avoid copying, reassembling, and reviewing Wikimedia communications.

97. I noted a number of situations under which NSA's use of a blacklist filter designed to block the communications to or from the IP addresses of Wikimedia servers would not guarantee that the NSA would avoid copying, reassembling and reviewing

[56] Bradner Decl. ¶¶ 144-45.

[57] Schulzrinne Reply Decl. ¶ 41.

Wikimedia communications in my previous report.[58] Dr. Schulzrinne responded to the discussion in my previous declaration in his reply declaration.[59] In his response, Dr. Schulzrinne addressed three of the cases in which I said that Wikimedia communications could still be copied, reassembled and reviewed even if there were a blacklist filter in place that discarded communications to or from the IP addresses of Wikimedia servers. Dr. Schulzrinne responds to my MCT, e-mail and VPN examples.

98. For each of these three cases, Dr. Schulzrinne creates a list of conditions that he says must be true for Wikimedia communications to be copied, reassembled and reviewed.

99. For example, in each of his sets of conditions, he says that the relevant communication would have to *not* be blacklisted in order for that communication to be acquired, as though that would be a difficult condition to meet. But, in reality, the possibility that these communications *would* be blacklisted in each example is far-fetched. For the MCT example, in which a Wikimedia communication is enclosed in an MCT traversing an international Internet link, Dr. Schulzrinne says that the MCT would have to not be blacklisted. Such an MCT could be one to or from a mail server. In the e-mail example, Dr. Schulzrinne says that the e-mail itself would have to not be blacklisted. In this case the communication would also be to or from an e-mail server. And, in the VPN example, he says that the VPN server communications would have to not be blacklisted.

[58] Bradner Decl. ¶ 367(b).

[59] Schulzrinne Reply Decl. ¶¶ 78-87.

100. It is technically possible for the NSA to block communications to and from e-mail servers and to or from VPN servers but it is hard to understand why the NSA would do this. These types of servers are exactly the types of services that NSA targets could be using, and so a world-wide blacklist blocking communications to and from e-mail and VPN servers would block just the type of sites bad actors would be using and the NSA would have an incentive to target.

101. All three of Dr. Schulzrinne's sets of conditions include the obvious requirements that either the user or the server be located outside the U.S., but not both, that communications between the user and server transit an international Internet link the NSA was monitoring, and that the users must be communicating with Wikimedia. These are all requirements I assumed when I described the cases. In my opinion, these requirements are obvious and would likely be frequently met.

4. *Still copying packets.*

102. As I discuss below and in my original declaration, even if the NSA were employing some sort of filter that discarded packets that were part of Wikimedia communications, the NSA would most likely be copying those packets before discarding them.[60]

5. *Probing for blind spots.*

103. In my original declaration I mentioned that if the NSA were using blacklist or whitelist filters to ignore protocols or IP address ranges, targets could probe to see if they could discover the lapses in coverage.[61] Dr. Schulzrinne questions the

[60] See ¶¶ 114-122; Bradner Decl. ¶¶ 269-279.

[61] Bradner Decl. ¶ 366(b).

possibility that the NSA's targets are sophisticated enough to probe for gaps in the NSA's coverage.[62] I will note that foreign intelligence officers and services are among the people and groups that are lawful targets for the upstream collection program,[63] and they are among our most sophisticated adversaries. I will also note that such probing is more likely to require process rather than technical sophistication. For example, it could involve purposefully communicating information about an information resource over a protocol you suspect the NSA is not monitoring and then monitoring the information resource to see if it is accessed, or purposefully conveying actionable information, such as the identity of a foreign agent, over such a protocol and seeing if action is taken against the agent.

104. There is another significant risk to the NSA's use of whitelists and/or blacklists to limit what its surveillance devices copy, reassemble, and review, and it would not require probing.

105. To implement the kind of whitelisting and blacklisting that Dr. Schulzrinne hypothesizes in the way that he hypothesizes, ISPs would need to configure their routers with the whitelists and blacklists. It is standard practice for ISPs to backup their router configurations in their network management systems, so that they can quickly deploy, modify or re-deploy the configurations as needed. There have been too many cases where ISPs' network management systems have been compromised, and so whitelisting and blacklisting of the sort Dr. Schulzrinne describes would create the unnecessary risk of compromise of the NSA's whitelists and blacklists. In addition,

[62] Schulzrinne Reply Decl. ¶ 32.

[63] 50 U.S.C. § 1881a.

multiple ISP technicians generally have access to the management stations, increasing the number of people that would have to be trusted.

106. Dr. Schulzrinne strangely suggests that the NSA might not care if foreign actors knew of ways to get around NSA monitoring.[64] I agree that I am not privy to the NSA's priorities that are not public, but I would find it quite extraordinary if the NSA did not care whether targets of its surveillance had a roadmap for evading its surveillance.

107. In fact, George C. Barnes, Deputy Director of the NSA, specifically stated that revealing information that could be used to help an adversary evade the NSA would be a problem:

> *Revealing which channels* [of communication] *are free from surveillance and which are not could also reveal sensitive intelligence methods, and thereby help an adversary evade detection and capitalize on limitations in the NSA's surveillance capabilities*.[65]

B. "About" communications

108. Dr. Schulzrinne's proposed use of whitelist and blacklist filters is also entirely inconsistent with "about" collection. About collection is a process within the upstream collection program that involves the collection of communications between two non-targets that contain one or more selectors associated with an NSA target. Dr. Schulzrinne discusses "about" communications in ¶¶ 49-52 of his reply declaration.

109. As a general rule, the IP addresses on the packets that make up the "about" communications will likely have no relation to any targets. For this reason, the use of

[64] Schulzrinne Reply Decl. ¶ 33.

[65] Barnes Decl. ¶ 57, ECF No. 141-1.

whitelist and blacklist filters of the sort that Dr. Schulzrinne describes is not compatible with "about" collection, because that kind of filtering would guarantee that the NSA's upstream collection devices miss "about" communications.

110. Dr. Schulzrinne responds to this fact by describing a two-step process for the collection of "about" communications that he claims is compatible with the use of whitelist and blacklist filters. The first step uses a whitelist IP address-based filter, but to set up this whitelist filter the NSA would have to know in advance which non-targets' IP addresses to whitelist, or what servers' IP addresses to whitelist in order to find the "about" communications.

111. Dr. Schulzrinne oversimplifies the problem by saying that "about" communications would be collected if the communication containing the "about" selector were whitelisted. Dr. Schulzrinne ignores the fact that in order to include such communications in a whitelist, the NSA would first have to know in advance which non-targets were going to be talking about targets and also know in advance what IP addresses the non-targets would be using. If the NSA were following Dr. Schulzrinne's description, they might capture an occasional "about" communication if one of the non-targets was using an IP address or service that the NSA had whitelisted for some other reason, but they could not normally capture them.

112. Under Dr. Schulzrinne's model, the only way the NSA could reliably capture about communications would be to whitelist all non-wholly domestic communications, in which case the whitelist would be guaranteed to result in the copying and review of Wikimedia communications.

V. DR. SCHULZRINNE'S WIKIMEDIA-AVOIDANCE THEORY IS IMPLAUSIBLE FOR NUMEROUS OTHER TECHNICAL AND PRACTICAL REASONS.

113. Dr. Schulzrinne made a number of other points in his reply brief. I will respond to some of them now.

A. Copy-then-filter vs. in-line filter

114. In my original declaration, I expressed the opinion that the NSA was most likely using a copy-then-filter architecture for its upstream collection program.[66] The government treats this as a significant concession, but the government completely misrepresents how this point relates to my ultimate conclusion. My discussion of the copy-then-filter implementation is a discussion of an independent reason to believe that the NSA is copying Wikimedia's communications as they travel across international Internet links—even if the NSA were performing filtering of the kind Dr. Schulzrinne hypothesizes. Specifically, a copy-then-filter architecture renders all of Dr. Schulzrinne's speculation about filtering irrelevant. That is because a copy-then-filter architecture involves the NSA copying all communications on a circuit independent of any filtering that might subsequently be performed on the copies of the packets. That of course includes the copying of Wikimedia's communications.

115. It bears emphasis, however, that even if the NSA is not using a copy-then-filter architecture, but instead is using the in-line filtering architecture Dr. Schulzrinne describes (which he refers to as "*filter-then-copy-and-scan*"), some Wikimedia communications will be copied. Dr. Schulzrinne's Wikimedia-avoidance architecture is entirely implausible for the many independent reasons I have explained above (including

[66] Bradner Decl. ¶¶ 265-289.

the multiple ways in which his theory conflicts with the government's public disclosures) and explain below (including the multiple ways in which his theories are at odds with the technical and practical necessities of conducting the upstream collection program as it has been described). In short, even if implemented as Dr. Schulzrinne imagines, Dr. Schulzrinne's filters will miss some Wikimedia communications and those communications will be copied, reassembled and reviewed.

116. I will now respond to Dr. Schulzrinne's points about the implications of using the one implementation (copy-then-filter) versus the other (in-line filtering). But, it continues to be my opinion that the copy-then-filter architecture is the simplest, most reliable and easiest to operate architecture for the NSA to use for the upstream collection program. If the NSA uses this architecture, all packets on a communications circuit being monitored by the NSA, including the packets that make up Wikimedia communications, are copied. It is possible that the NSA uses different architectures in different monitoring locations, but, in my opinion, the advantages of the copy-then-filter architecture mean that it is most likely the default architecture. Even if the copy-then-filter architecture is not being used everywhere, all packets on the circuits where the copy-then-filter architecture is used are copied, as are many packets on the circuits where an in-line filter architecture is used.

1. *Fiber-optic splitter.*

117. The simplest option involves the use of a fiber-optic splitter, which Dr. Schulzrinne suggested in his first declaration and I discussed in my original declaration.[67] A fiber-optic splitter produces a copy of all communications on a fiber by splitting the

[67] Schulzrinne Decl. ¶¶ 55-56, ECF No. 164-4; Bradner Decl. ¶¶ 275-77.

light on the fiber into two streams.[68] Dr. Schulzrinne notes that adding a fiber-optic splitter into a network adds a potential failure point and causes loss of optical power.[69] He is correct that a fiber-optic splitter could be a failure point, but a fiber-optic splitter is a passive device that does not include a processor or software that could have bugs or need upgrading, and it does not require any configuration or power so the probability of failure is very low and the possibility that misconfiguration or that a power failure could impact the network is nonexistent. A fiber-optic splitter does reduce the optical power that would be received by the ISP's receiving device but, as long as the ISP knows the splitter is in-line, the receiving device can be configured to compensate for the loss.

118. On the other hand, using a router's mirror function (as Dr. Schulzrinne describes), would have a much higher failure and disruption probability because the router requires power and because the router includes a computer and software that can have bugs, would need updating, and would be vulnerable to hacking. If the ISP used an existing router to filter and then copy communications, the *added* risks would be a little bit less significant, but the need to constantly reconfigure the device with updated blacklists and whitelists would create the risk of misconfiguration or overloading. Either way, the risks of failure are greater for the in-line device Dr. Schulzrinne proposes than for a fiber-optic splitter.

119. Dr. Schulzrinne notes, as I did in my original declaration, that the use of a fiber-optic splitter would mean that it would need to be coupled with an opto-electronic

[68] Bradner Decl. ¶¶ 5, 275-77.

[69] Schulzrinne Reply Decl. ¶ 27.

device to split out the channels the NSA wanted to monitor.[70] Dr. Schulzrinne paints a pretty dire picture of the requirements for this device, but, as I pointed out in my original declaration, not all channels on the cables are used to transport international Internet communications.[71] The splitting device only needs to pay attention to the circuits that are so used and, of those, only the circuits that the NSA wishes to monitor. Note that the splitting device Dr. Schulzrinne mentions is not an esoteric piece of equipment; it is the same device that ISPs routinely use to split the light on optical fiber into different channels and is normally included in the router that the optical fiber is plugged into.

2. *Configuring the ISP router to mirror communications.*

120. The other option to support the copy-then-filter architecture is, as Dr. Schulzrinne suggests, for the NSA to command the ISP to configure its router to mirror the communications on the circuits the NSA wants to monitor and send the mirrored packets to one or more NSA-operated devices.[72]

121. The operationally simplest way to do this is to command the ISP to configure its router to mirror all of the packets on the channels the NSA wants to monitor. Such a configuration is very simple and would not have to change over time. This way also means that the ISP personnel are not exposed to any of the NSA's collection methods other than the fact that data collection is being done on a particular circuit.

[70] Bradner Decl. ¶ 277; Schulzrinne Reply Decl. ¶ 21.

[71] Bradner Decl. ¶¶ 214-25.

[72] Bradner Decl. ¶ 278.

3. *All packets are copied.*

122. All packets on a circuit, including the packets comprising any Wikimedia communications, are copied with either of the above two designs.

4. *Configuring the ISP router to filter before mirroring.*

123. Dr. Schulzrinne hypothesizes that the NSA could be applying whitelist or blacklist filters to the packets *before* the router mirrors the packets to the NSA upstream collection devices.[73] He says, and I agree, that if the NSA is doing so, then the logical place to apply such filters is in an ISP router that will be processing the stream of packets entering or exiting a communications channel the NSA wants to monitor. Dr. Schulzrinne and I disagree as to whether the NSA could in fact be using such whitelist or blacklist filters on all of the circuits that it monitors, even if it uses them on some, and we disagree as to what specific configurations would be likely for whitelist or blacklist filters. I discuss this disagreement at length above.[74]

124. There are different categories of whitelist or blacklist filters that have different implications when it comes to configuring the filters. In this case, I believe it most likely that the actual router configuration is being performed by ISP personnel since, in my experience, an ISP would be unlikely to allow non-ISP personnel to configure its routers. In theory, the NSA could create some configuration data files that the ISP personnel could then load into the router—this would put the NSA directly in charge of the filter details while avoiding having the fingers of non-ISP personnel in the routers. In this case the ISP personnel would still have access to the details of the

[73] Schulzrinne Reply Decl. ¶ 28.

[74] See ¶¶ 63-107.

configuration. In any case, the more complex and detailed the filter configurations, the more often they will change. If an ISP had to implement all of the whitelist and blacklist filters Dr. Schulzrinne has suggested, the ISP would have to be updating its router configuration all the time. Operationally, this is not a good idea from an ISP's perspective because it increases the chance of human error impacting the ISP operations.

125. In his reply declaration, Dr. Schulzrinne says that he does not suggest installing an NSA-operated device in the middle of an ISP's network.[75] There are only two options for the type of in-line filter that Dr. Schulzrinne proposes: (1) having an NSA-operated device perform the filtering, or (2) exposing ISP personnel to the details of the NSA's collection program. Each of these options has significant downsides. If in-line filters are used at all, it may well be that there is no single answer. For example, the NSA might want to operate its own device in smaller ISPs or ISPs demonstrating less technical expertise but delegate the operation to ISP personnel in larger ISPs or those with more technical expertise.

126. As I discussed in my previous declaration, it is my opinion that the NSA would want to limit the exposure of the details of at least some of the types of whitelist and blacklist filters Dr. Schulzrinne suggests could be used. Dr. Schulzrinne dismisses that concern by pointing to the fact that the NSA shares target information with telecommunications providers under the PRISM program.[76] But Dr. Schulzrinne ignores the fact that the categories of information are quite different. With PRISM, the identities of one or more individual targets are exposed to the personnel of the telecommunications

[75] Schulzrinne Reply Decl. ¶ 23.

[76] Schulzrinne Reply Decl. ¶ 18.

companies it compels to provide assistance. But the detailed configuration of blacklist and whitelist filters can provide a roadmap that can be used to entirely avoid NSA surveillance.

127. Note that the NSA has to share confidential target information with telecommunications providers under PRISM, since only the telecommunications providers have the ability to retrieve the communications of interest. The NSA does not have that constraint in the upstream collection program. The NSA could:

a. use a copy-then-filter approach with a fiber-optic splitter, which would reveal no confidential information to the provider other than the fact that the NSA was monitoring one or more channels on a fiber;

b. use a copy-then-filter approach by commanding the ISP to provide the NSA with a full copy of the packets on a particular channel, which would increase the shared confidential information to include the specific channels being monitored;

c. operate its own in-line filter device which would also share the information about which channels were being monitored; or

d. command the ISP to operate an in-line filter, which would require the NSA to share the details of exactly what it is monitoring and not monitoring (i.e., the configurations for the whitelist and blacklist filters Dr. Schulzrinne hypothesizes).

128. Given the fact that the NSA has a choice, it seems reasonable to infer that the NSA would want to minimize the confidential information that it had to share to

operate the upstream collection program. (See also ¶ 107 above, citing Barnes Decl. ¶ 57.)

129. It is one thing to expose the fact that the NSA has asked for Mr. Smith's e-mail and a very different thing to expose the fact that the NSA is using a whitelist filter that discards packets to or from large chunks of the Internet or discards all packets that are not e-mail or web traffic.

B. Collecting "web activity"

130. Dr. Schulzrinne questions the idea that the NSA might be monitoring communications to or from web servers.[77] I discussed the government's disclosure that it is monitoring "web activity" in my original declaration.[78] Dr. Schulzrinne suggests that the government might have been sloppy and was referring to overall Internet activity when they wrote "web activity."[79] That seems to me to be a very tenuous argument. I have seen no indication in any of the government's released documents that they are that sloppy. In particular, the reference to "web activity" is in a formal and highly technical government submission to the FISC in response to a highly technical request from the court, hardly a place that anyone would be sloppy. And, the context in which the phrase was used makes total sense if the government was using the phrase precisely, as a reference to the world wide web protocols (HTTP/HTTPS).

131. It is also well known that terrorists make use of communications tools that use HTTP/HTTPS. Most human-to-human communications on the Internet are transported using HTTP/HTTPS, whether over older mediums (websites) or more recent

[77] Schulzrinne Reply Decl. ¶ 36(b).

[78] Bradner Decl. ¶¶ 314-15.

[79] Schulzrinne Reply Decl. ¶ 36(b).

ones (messaging services). For example, terrorists make use of the Telegram application, which can operate over port 80 (HTTP).[80] There are obviously examples of widely used protocols that do not use HTTP/HTTPS—for example, FTP/SMTP/IMAP/POP/etc. — but an increasing amount of Internet communications do use HTTP/HTTPS to increase security (HTTPS) or to bypass firewalls (both).

C. Collecting web communications

132. Dr. Schulzrinne spends some time discussing ways the NSA could be filtering out web or encrypted traffic to make it less likely that it is copying, reassembling and reviewing Wikimedia communications. I will now review his suggestions.

133. Dr. Schulzrinne suggests that the NSA might be configuring a blacklist to block both HTTP and HTTPS communications.[81] First of all, the government has acknowledged that it is capturing web activity, i.e. HTTP and/or HTTPS communications (¶ 130). The NSA cannot be both blocking all web communications and be collecting web activity.

134. Note that web traffic makes up a very large percentage of Internet traffic. For example, one study of traffic between the Japanese WIDE Project ISP and its backbone ISP showed that web traffic was 75% of the overall traffic.[82] As Dr. Schulzrinne notes, web traffic is not restricted to web sites such as cnn.com and

[80] Joby Warrick, *The 'App of Choice' for Jihadists: ISIS Seizes on Internet Tool to Promote Terror*, Wash. Post (Dec. 23, 2016), http://wapo.st/2hzoY6P; *MTProto Mobile Protocol: Transport*, Telegram, https://perma.cc/T6FL-WWP8.

[81] Schulzrinne Reply Decl. at 15 n.7.

[82] Chia-ling Chan, et al., *Monitoring TLS Adoption Using Backbone and Edge Traffic* (2018), https://perma.cc/6C8D-GWCT.

wikipedia.org; web protocols are also used for webmail and chatrooms.[83] Ignoring web traffic as Dr. Schulzrinne has suggested would exclude the vast quantities of human-to-human communications that are transported by the web.

135. As I mentioned in my original declaration, any such discarding of all web communications would leave a very large hole in the NSA's coverage[84]—contrary to any notion of completeness such as that noted in the PCLOB Report

136. Dr. Schulzrinne suggests that the NSA could be restricting its collection of web communications to a few sites such as "*specific webmail and chatroom sites*." As I discussed above, this would also leave very large holes in the NSA's coverage. (¶ 134) Any such filtering would also be contrary to the aim of the "about" collection program, which is to collect communications between unknown non-targets.

D. Collecting encrypted communications

137. Dr. Schulzrinne says that the NSA might not be collecting HTTPS communications using the authority in Section 702 to collect encrypted communications.[85] Dr. Schulzrinne makes no actual argument that the NSA is not collecting at least some encrypted communications, he just says that my arguments are not technical.

138. Multiple government disclosures make it clear that the NSA collects encrypted communications. For example:

a. The PCLOB noted: "*With respect to all of the agencies, extensions from these age-off requirements may be sought from*

[83] Schulzrinne Reply Decl. ¶ 35.

[84] Bradner Decl. ¶ 366(f).

[85] Schulzrinne Reply Decl. ¶ 36(a).

a high-level agency official. Other limited exceptions apply, such as to communications that are still being decrypted." (Note the phrase "*still being decrypted.*")[86]

b. The PCLOB also noted: "*The NSA may also retain communications beyond the normal age-off period if it is still decrypting the communication or using the communication to decrypt other communications*." (Note the phrases "*still decrypting*" and "*decrypt other communications.*")[87]

c. The NSA's minimization procedures note: "*In the context of a cryptanalytic effort, maintenance of technical data bases requires retention of all communications that are enciphered or reasonably believed to contain secret meaning, and sufficient duration may consist of any period of time during which encrypted material is subject to, or of use in, cryptanalysis*." (Note the phrase "*encrypted material is subject to.*")[88]

139. Dr. Schulzrinne hypotheses that while the NSA may collect encrypted communications it may only be doing so under PRISM.[89] This seems unlikely. The telecommunications providers assisting the NSA in the case of PRISM will frequently have direct access to the user's unencrypted communications, for example in an e-mail

[86] Appendix F (PCLOB Report at 60).

[87] Appendix F (PCLOB Report at 63).

[88] Appendix S (NSA Section 702 Minimization Procedures at 10 (2014), *available at* ECF No. 168-4 at 1046-1061).

[89] Schulzrinne Reply Decl. ¶ 36(a).

server, whereas communications across the Internet are increasingly being encrypted.[90] If the upstream collection program were to ignore encrypted communications it would be increasingly unable to collect any communications.

E. GCHQ surveillance

140. I discussed some public disclosures from the U.K.'s signals intelligence agency, Government Communications Headquarters (GCHQ), in my first declaration in order to "*reinforce my conclusions that the NSA relies on the copy-then-filter configuration to conduct the upstream collection program and that it does not selectively filter traffic prior to copying it as Dr. Schulzrinne hypothesizes it could*."[91]

141. Those include the disclosure by the GCHQ that, under a surveillance program analogous to upstream collection, "*it is necessary to intercept the entire contents of a bearer [circuit], in order to extract even a single specific communication for examination*."[92]

142. Dr. Schulzrinne dismisses the disclosures as only being "*the roughest outline*" of the process the GCHQ uses, and as being "*non-technical*."[93] I disagree and find the disclosures have enough detail for me to draw my conclusions.

143. He also says that the disclosures, even if not detailed, are "*quite comparable*" to his suggested filter first approach. I disagree with this as well.

144. Dr. Schulzrinne provides a citation describing the GCHQ's bulk interception as purported proof of his contention.[94] The citation Dr. Schulzrinne provides

[90] John Maddison, *Encrypted Traffic Reaches a New Threshold*, NETWORKComputing (Nov. 28, 2018), https://perma.cc/6VFY-YEGL.

[91] Bradner Decl. ¶ 369.

[92] Bradner Decl. ¶ 368.

[93] Schulzrinne Reply Decl. ¶¶ 59-60.

is from a filing by the U.K. government in the European Court of Human Rights. The U.K. filing, in turn, refers to a "*Bulk Powers Review*" of "*the operational case for various intelligence gathering powers.*"[95] The U.K. filing quotes the *Bulk Powers Review* to provide a summary on how the GCHQ interception program works.[96] Dr. Schulzrinne's citation is of the U.K. filing's quoting of the *Bulk Powers Review*.

145. Dr. Schulzrinne says that the U.K. filing "*actually describe*[*s*] *a collection approach quite comparable (at least at a general level) to the type of IP address and port and protocol number filtering described in my earlier declaration*."[97] I disagree with Dr. Schulzrinne's analysis of the U.K. filing he provides in support of his conclusion. I do not believe that the citation shows any evidence that the GCHQ is filtering traffic on a channel before copying the traffic; in fact, the citation shows the opposite. The paragraph in the *Bulk Powers Review* that immediately follows the outline Dr. Schulzrinne cites makes this clear:

> ***The two major processes***
>
> *2.19. A description is given in the 2015 ISC report (paras 61-73), of two major and distinct processes that apply to interception under bulk warrants. Those processes are identified in more detail in the closed version of the report, and I have been briefed on each of them. In summary:*
>
> *(a)* ***The "strong selector" process*** *(2015 ISC report, paras 61-64) operates on the bearers that GCHQ has chosen to access. As the*

---

[94] Schulzrinne Reply Decl. ¶ 61.

[95] David Anderson Q.C., U.K. Independent Reviewer of Terrorism Legislation, *Report of the Bulk Powers Review*, Cm 9326 (August 2016), https://perma.cc/V3ME-QZED ("*Bulk Powers Review*").

[96] *Bulk Powers Review*, *supra* note 101, at 23-24.

[97] Schulzrinne Reply Decl. ¶ 61.

internet traffic flows along those chosen bearers, the system compares the communications against a list of strong selectors in near real-time. Any communications which match the selectors are automatically collected and all other communications are automatically discarded. The nature of the global internet means that the route a particular communication will take cannot be predicted and a single communication is broken down into packets which can take different routes. In order to identify and reconstruct the wanted communications of subjects of intelligence interest, GCHQ's processing relies on accessing the "related communications data" (secondary data) in the bearer.

A copy of all the communications on a bearer has to be held for a short period in order to allow the strong selectors to be applied to those communications. This process accordingly requires a bulk warrant under the Bill. However, in the opinion of the ISC, "while this process has been described as bulk interception because of the numbers of communications it covers, it is nevertheless targeted since the selectors used relate to individual targets".

(b) ***The "complex query" process*** *(2015 ISC report paras 65-73) is used where GCHQ is looking to match much more complicated criteria, for example with three or four elements. This process operates across a far smaller number of bearers. These bearers are not chosen at random, as GCHQ focuses its resources on those most likely to carry communications of intelligence value. As a first step in the processing under this method the system applies an initial set of processing rules. Those rules seek to select communications of potential intelligence value while discarding those least likely to be of intelligence value. The selected communications are not available to GCHQ staff to search through at will. Further complex searches draw out the*

> *communications of intelligence value. By performing searches combining a number of criteria, the odds of a 'false positive' are considerably reduced.*
>
> *This second process is closer to true bulk interception, since it involves the collection of unselected content and/or secondary data. It permits types of analysis and selection that are not currently achievable in the near real-time environment of the strong selector process (2.19(a) above). But as with the first process, it remains the case that communications unlikely to be of intelligence value are discarded as soon as that becomes apparent.*[98]

146. The description of the "strong selector" process specifically says "*[a] copy of all the communications on a bearer has to be held for a short period in order to allow the strong selectors to be applied to those communications."* Combining this statement with the description of the third stage of collection in the extract Dr. Schulzrinne provided makes it clear that all of the communications on a bearer (GCHQ's term for a circuit) are copied and stored at least temporarily so that those communications that contain selectors, if any, can be located.

147. The description of the "complex query" process says that the process is "*closer to true bulk interception, since it involves the collection of unselected content and/or secondary data.*"[99] This statement by itself notes that the GCHQ is collecting "*unselected content*." To do so, it is copying communications that it has not checked for the presence of selectors.

---

[98] *Bulk Powers Review*, *supra* note 101, at 24-25.

[99] *Bulk Powers Review*, *supra* note 101, at 25.

148. In both cases, GCHQ is copying all the contents of a bearer, which is the point I made in my original declaration.

F. ISP-operated copy device

149. Dr. Schulzrinne suggests that a fiber-optic splitter, or even an electronic device such as a router, could be operated by the ISP and the copy of the communications created by such a device could be sent to a filtering device operated by the ISP with the output of the filtering device sent to the NSA.[100] Such an arrangement would not actually change the fact that the NSA is creating a copy, since the copy device and filter would be operated at the direction and auspices of the NSA. As I noted in my first declaration, work performed at the direction of the NSA is still work done by the NSA.[101] In addition, as with other copy-then-filter configurations, all the packets on the circuit, including packets that are part of Wikimedia communications, are copied.

G. EINSTEIN

150. I mentioned in passing the U.S. government-operated EINSTEIN 2 & 3 systems in my original declaration.[102] Dr. Schulzrinne made rather much more of the mention than I had in mind. I just mentioned EINSTEIN as an example of a deep packet inspection (DPI) device. But I will comment on Dr. Schulzrinne's discussion of EINSTEIN.[103]

151. Dr. Schulzrinne tries to differentiate EINSTEIN from the NSA's upstream collection program in two ways: (1) he says that EINSTEIN has to look at all traffic

[100] Schulzrinne Reply Decl. ¶ 64.

[101] Bradner Decl. ¶ 5.

[102] Bradner Decl. ¶¶ 259, 286.

[103] Schulzrinne Reply Decl. ¶¶ 67-69.

whereas the upstream collection program, at least as he imagines it, does not; and (2) EINSTEIN is a cybersecurity system and upstream collection program is not.

152. Relating to (1), Dr. Schulzrinne says "*cyber attacks can use any protocol, originate from any external Internet host, and can target any destination system, to be effective an intrusion-detection system must inspect all incoming traffic*."[104] Of course, the same can be said, with the exception of the discarding of wholly domestic communications, of the upstream collection program. The main theme of Dr. Schulzrinne's declaration is how to limit the upstream collection program so that it avoids Wikimedia communications at the expense of being able to capture communications from, to or about its targets which "*can use any protocol, originate from any external Internet host, and can target any destination system*."[105]

153. Relating to (2), the use of DPI is not limited to cybersecurity systems. The upstream collection program uses DPI to find the selectors in the communications it reviews. In addition, the NSA's public materials include mention of the use of the upstream collection program for cybersecurity purposes.[106]

VI. SUMMARY

154. In order for Dr. Schulzrinne to be correct in his speculation that the NSA could operate the upstream collection program without ever copying, reassembling, or reviewing even a single Wikimedia communication, every one of the following assumptions, which are prerequisites for his claim, must be true:

[104] Schulzrinne Reply Decl. ¶ 69.

[105] Schulzrinne Reply Decl. ¶ 69.

[106] See, e.g., *'Section 702' Saves Lives*, *supra* note 47; *Section 702 Overview*, *supra* note 49, at 4.

1. That the government's unambiguous concession to the FISC—"*that NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA*"[107]—was false, despite the NSA's representative confirming its accuracy at her deposition.[108]
2. That the PCLOB's statement—that the technical design of the upstream collection program supports the NSA's goal "*to comprehensively acquire communications that are sent to or from its targets*"—was false.[109]
3. That the NSA decided to significantly limit the scope of its collection under the upstream collection program by doing at least one of the following:
    a. Deliberately blacklisting all Wikimedia IP addresses.
    b. Deliberately whitelisting a subset of IP addresses other than Wikimedia's IP addresses.
    c. Deliberately excluding all of the communications protocols that Wikimedia uses, through blacklists or whitelists configured to exclude:
        i. All web activity (i.e., HTTP/S), notwithstanding the government's concession that upstream collection involves the collection of "web activity," *and*

[107] Appendix P (FISC Opinion at 45).

[108] Appendix K (Richards Dep. 160:4-17).

[109] Appendix F (PCLOB Report at 10).

ii. All e-mail activity (i.e., SMTP), notwithstanding the government's concession that, under the upstream collection program, it uses e-mail addresses as selectors.[110]

4. That the NSA has limited the scope of its upstream collection in this way on *every* one of the international Internet links it monitors.

5. That the limitations on collection above are in fact entirely effective at avoiding Wikimedia's communications, even though there are multiple circumstances in which they would not be. For example, Dr. Schulzrinne's claim requires that none of the following could ever occur:

   a. If the NSA whitelisted IP addresses other than Wikimedia's IP addresses:

      i. A user of a whitelisted IP address communicates with Wikimedia, and the communication traverses an international Internet link monitored by the NSA.

   b. If the NSA blacklisted all Wikimedia IP addresses:

      i. One of Wikimedia's communications is enclosed in a multi-communication transaction (MCT) that is not blacklisted, and it traverses an international Internet link monitored by the NSA.

      ii. One of Wikimedia's communications passes through an intermediary that replaces Wikimedia's IP address (such as an e-mail server, VPN, or other communication service)

[110] Appendix F (PCLOB Report at 7).

with an IP address that is not blacklisted, and it traverses an international Internet link monitored by the NSA.

iii. A user of a whitelisted IP address communicates with Wikimedia, and the communication traverses an international Internet link monitored by the NSA.

c. If the NSA excluded the protocols that Wikimedia's communications use:

i. One of Wikimedia's communications is enclosed in a multi-communication transaction (MCT) using a protocol that is not excluded, and it traverses an international Internet link monitored by the NSA.

ii. One of Wikimedia's communications passes through an intermediary using a protocol that is not excluded, and it traverses an international Internet link monitored by the NSA.

6. That the NSA uses an in-line filter architecture to accomplish the upstream collection program in all cases, rather than ever using a copy-then-filter architecture.

155. Given these considerations and my analysis of the NSA's disclosures regarding the upstream collection program, it remains my opinion that it is virtually certain that the NSA has, in the course of the upstream collection program, copied, reassembled and reviewed at least some of Wikimedia's communications.

I declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Date: 3/8/2019

Scott Bradner

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 2

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION,

Plaintiff,

v.

NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*,

Defendants.

No. 1:15-cv-00662-TSE

REPLY DECLARATION OF JONATHON PENNEY

TABLE OF CONTENTS

I. INTRODUCTION 1
II. DR. SALZBERG'S ANALYSIS IS FLAWED 1
III. RESPONSES TO DR. SALZBERG'S SIX METHODOLOGICAL CRITIQUES 18

I. INTRODUCTION

1. My name is Jonathon Penney. I have been asked by the plaintiff's counsel in *Wikimedia Foundation v. National Security Agency*, No. 1:15-cv-00662-TSE (D. Md.), to provide this reply declaration to address the Defendants' reply to the Plaintiff's brief and to my declaration, both of which were dated December 18, 2019. Results of my additional analysis in support of this reply declaration are included herein in the Appendix.

2. My qualifications and expertise are discussed in detail in my opening Declaration ("Declaration"). *See* ECF No. 168-02. Unless otherwise stated, I have personal knowledge of the facts herein.

II. DR. SALZBERG'S ANALYSIS IS FLAWED

3. In support of their Reply motion, Defendants submit the Declaration of Dr. Alan Salzberg ("Salzberg Declaration"). *See* ECF No. 178-3. Dr. Salzberg's analysis misunderstands my interrupted time series (ITS) design and study in fundamental ways and ignores relevant literature on methodological best practices for ITS studies. Furthermore, in critiquing my study, the Salzberg Declaration relies primarily on a visual inspection of data, which can often be misleading,[1] rather than formal testing mechanisms that can be verified. Salzberg's reliance on a visual inspection of the data causes him to formulate misguided critiques and conclusions as to my study's reliability and validity. In actuality, my methodology, method of analysis, and regression model is superior to any alternatives he suggests.

4. **First**, Salzberg's use of disaggregated line plots (see ¶¶ 12-14 and Figure 1 of the Salzberg Declaration) to analyze the Wikipedia page view data is an inferior method of analysis

[1] GENE S. FISCH, *EVALUATING DATA FROM BEHAVIORAL ANALYSIS: VISUAL INSPECTION OR STATISTICAL MODELS?* 54 Behavioural Processes 137, 137 (2001) (quoting Howard Wainer: "A graph is nothing but a visual metaphor. To be truly evocative, it must correspond closely to the phenomena it depicts… If a graphic depiction of data does not faithfully follow this notion it is almost sure to be misleading.").

for ITS studies, as compared to the segmented linear regression trend analysis of aggregated data that I used. Dr. Salzberg's approach ignores relevant literature on methodological best practices. Indeed, the recommended method of analysis for ITS design studies is segmented linear regression analysis, which I employed in my study, as it allows researchers to: (1) control for prior trends in the data; (2) measure the dynamics of change in response to an intervention; (3) tolerate fewer time points than alternative methods; (4) adjust for serial correlation in the data; and (4) apply these methods to aggregate level data.[2] By contrast, Salzberg's disaggregated line-plots method offers none of these strengths or advantages and is neither recommended nor discussed in relevant literature.

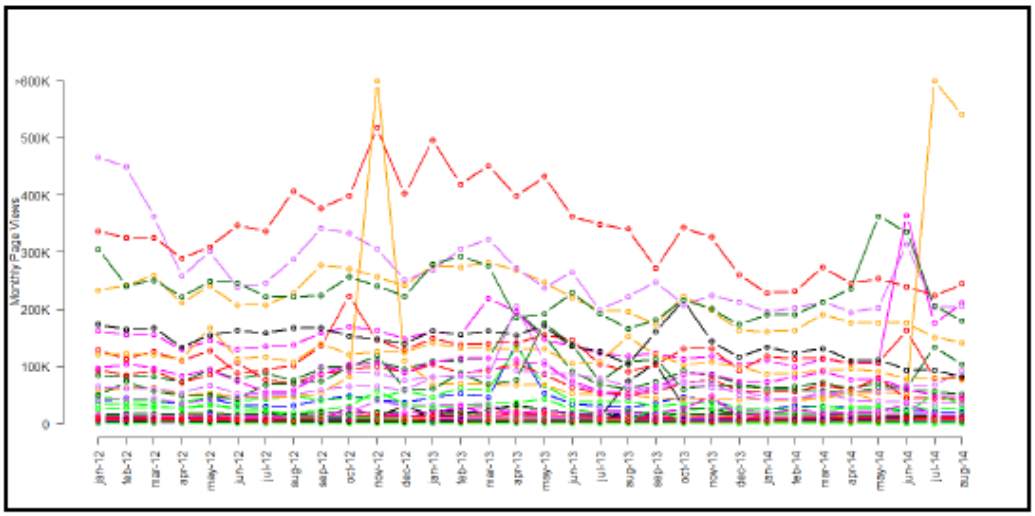
5. **Second**, Salzberg's use of disaggregated line plots adds "noise," both visual and statistical, which masks actual overall trends in the data best understood through analysis of *aggregated* monthly page view data. Consistent with my approach, a majority of ITS design studies use aggregated data.[3] This is to reduce or remove noise in the data and to allow for more

[2] The leading peer reviewed works on ITS design, methodology, and analysis recommends segmented linear regression analysis over several alternatives methods and models, including generalized estimating equations (GEE) method and autoregressive integrated moving average (ARIMA) modeling. FANG ZHANG, A.K. WAGNER, ET AL., *METHODS FOR ESTIMATING CONFIDENCE INTERVALS IN INTERRUPTED TIME SERIES ANALYSES OF HEALTH INTERVENTIONS*, 62:2 J. Clinical Epidemiology 143, 143-144 (2009) (discussing the advantages of segmented linear regression analysis for ITS designs compared to alternatives); A.K. WAGNER ET AL., *SEGMENTED REGRESSION ANALYSIS OF AN INTERRUPTED TIME SERIES IN MEDICATION USE RESEARCH*, 27 J. Clinical Pharmacy & Therapeutics 299, 299, 208 (2002) (describing segmented linear regression analysis as a "powerful statistical method or estimating intervention effects" in ITS studies, and describing "strengths"); MYLENE LAGARDE, *HOW TO DO (OR NOT TO DO) … ASSESSING THE IMPACT OF A POLICY CHANGE WITH ROUTINE LONGITUDINAL DATA,* 27:1 Health Policy and Planning 76, 79 (2012) (noting this method "controls for secular trends and can also adjust for potential serial correlation of the data"); ROBERT B. PENFOLD & FANG ZHANG, *USE OF INTERRUPTED TIME SERIES ANALYSIS IN EVALUATING HEALTH CARE QUALITY IMPROVEMENTS*, 13:6 Acad. Pediatrics S38 (2013) (discussing the advantages and limitations of employing time series analysis to understand and explore the impact of health policy changes).

[3] Wagner (2002), *id.*, at 308 ("Segmented regression typically aggregates individual-level data by time point" and noting a leading ITS study where the "unit of analysis" was a monthly aggregated data, as used in this study); JANDOC, ET AL., *INTERRUPTED TIME SERIES ANALYSIS IN DRUG UTILIZATION RESEARCH IS INCREASING: SYSTEMATIC REVIEW AND RECOMMENDATIONS*, 68 J. Clinical Epidemiology 950, 950 (2015) ("Interrupted time series methods use aggregate data collected over equally spaced intervals before and

sophisticated statistical tests and analysis.[4] Salzberg departs from this approach, which is standard in a majority of ITS studies, leading him to incorrect inferences and conclusions in his analysis. For example, Figure 1 of the Salzberg Declaration distorts and hides important trends by plotting individual line plots for the 48 Terrorism articles. Here, line plots for a majority of the 48 Articles cannot be seen as they have page views too small to be visualized with the large page view scale (0 to 600,000 page views) used on the vertical axis of the graph:

Figure 1: Individual Page Views for Each of the Articles Within the Terror 48, Which The Penney Declaration Hypothesized Show an Immediate Decline Beginning in June 2013



6. Salzberg's Figure 1 creates a false impression there are no patterns or trends for overall page views over 32 months. But those trends are easily visible when individual article page views are analyzed as aggregated monthly page views as visualized in **Figure 1a** below. This

after an intervention, with the key assumption that data trends before the intervention can be extrapolated to predict trends had the intervention not occurred"); EMMA BEARD, *USING TIME-SERIES ANALYSIS TO EXAMINE THE EFFECTS OF ADDING OR REMOVING COMPONENTS OF DIGITAL BEHAVIOURAL INTERVENTIONS AND ASSOCIATIONS BETWEEN OUTCOMES AND PATTERNS OF USAGE*, Centre for Behaviour Change (CBC) Conference, University College of London 15 (2017), https://www.ucl.ac.uk/behaviour-change/events/presentations-17/beard.pdf (noting that a "[m]ajority of studies use aggregated data". She also specifically notes that linear regression may be used for "interrupted time series design" if autocorrelation is controlled).

[4] BEARD, *id.*, at 15 (noting that a "[m]ajority of studies use aggregated data" as this "removes noise and allows for more sophisticated tests which require continuous or rate type data").

figure plots the median aggregated page views for the 48 Terrorism Articles with trend lines included to understand the shift in trend.

Figure 1a: Aggregated Median Page Views for 48 Terrorism Wikipedia Articles

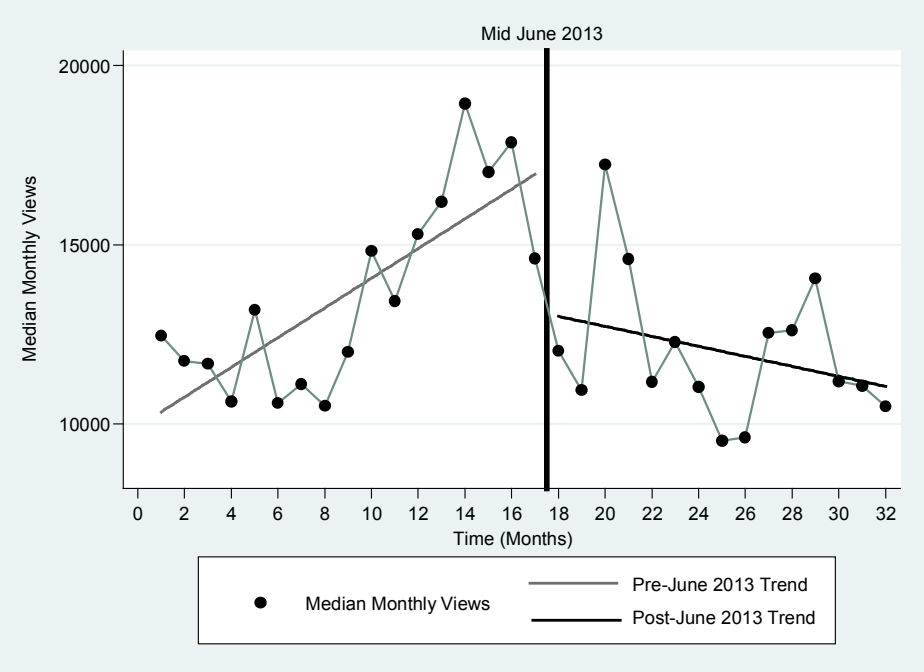
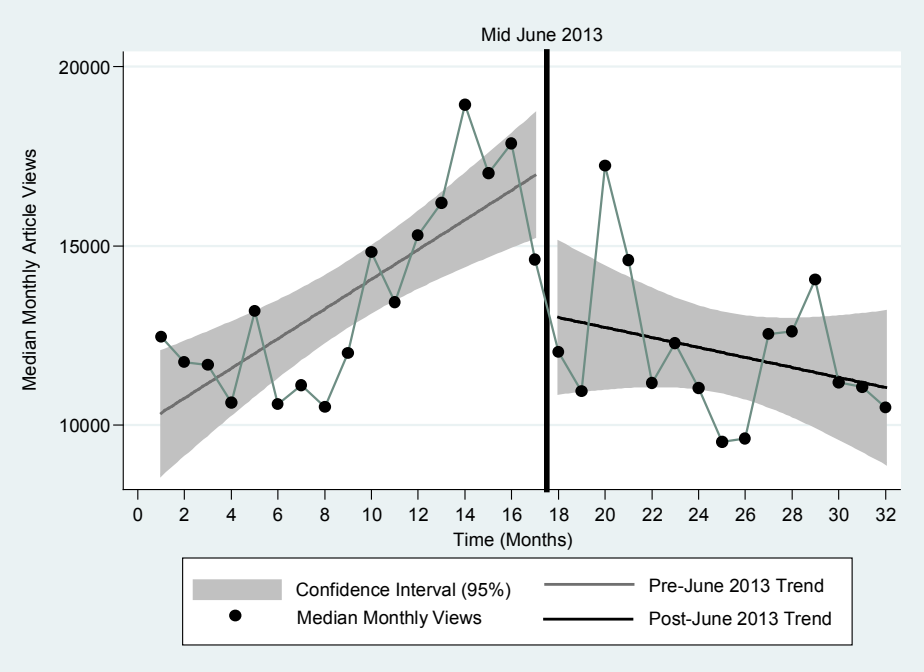


Figure 1b: Aggregated Median Page Views for 48 Terror Articles (With C.I.)

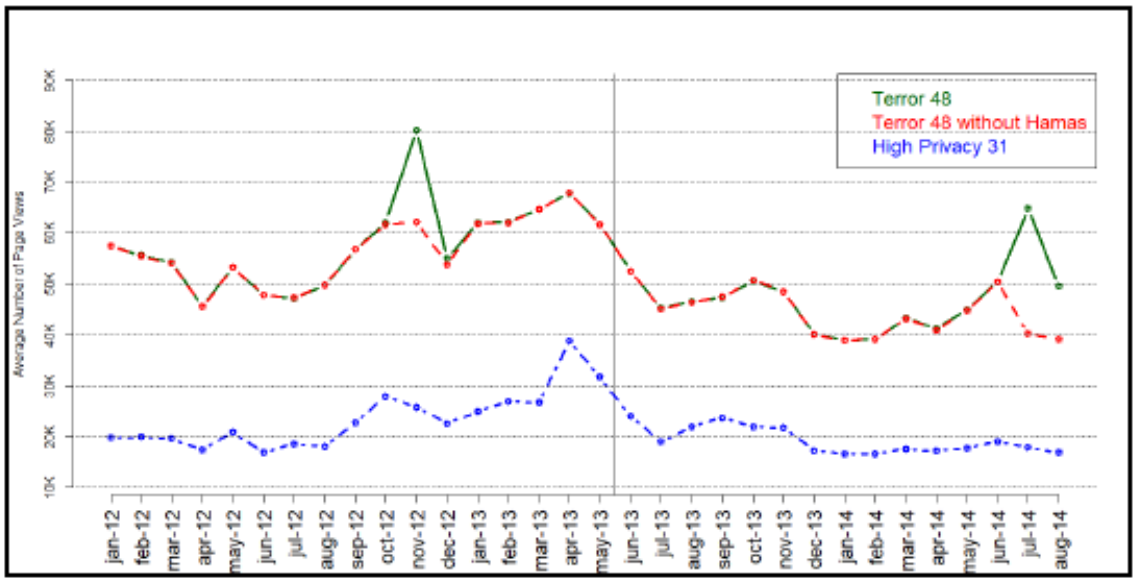


7. **Figure 1b** visualizes the same data with pre/post June 2013 trend lines and

confidence intervals (the gray shaded area), and demonstrates that the trend change before and after June 2013 was statistically significant, as there is no overlap of confidence intervals during these two periods. Specifically, this visualization of the data demonstrates a statistically significant drop in June 2013 and reduction in overall monthly page views. This is a far clearer visualization, with clear trends, compared to Salzberg's Figure 1, which masks these trends in the "noise" of 48 disaggregated line plots.

8. Figures 5 and 6 are of the Salzberg Declaration are similarly distorted. These figures present line-plots for the 48 Terror Articles, 47 Terror Articles, and 31 High Privacy Articles together. For example, Figure 5 visualizes a line plot for the average monthly page views for those three article sets:

Figure 5: Average Page Views Show a Peak in April 2013 or Before



9. Again, by presenting the data associated with these three different sets of articles on the same graph with the same scale on the vertical axis (10,000 to 90,000 average page views), the Figure distort the presentation of the data, creating a false impression that both the 47 Terror Articles and the 31 articles with more privacy sensitive ratings ("31 Higher Privacy Articles") have

a flat trend over the course of the 32 months. These distortions are easy to visualize when the average monthly page views for the 31 Higher Privacy Articles are plotted and presented with an appropriate scale as in **Figures 2a** and **2b** below.

Figure 2a: Average Monthly Page Views for the 31 Higher Privacy Articles Plotted Alone Show Increase Until June 2013 And Then A Drop-Off After That Month

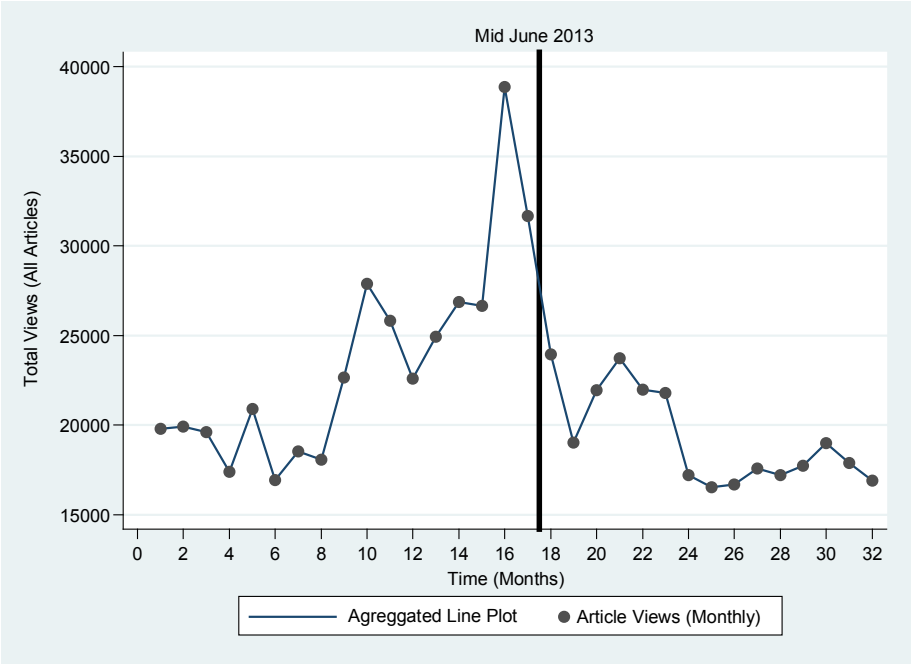
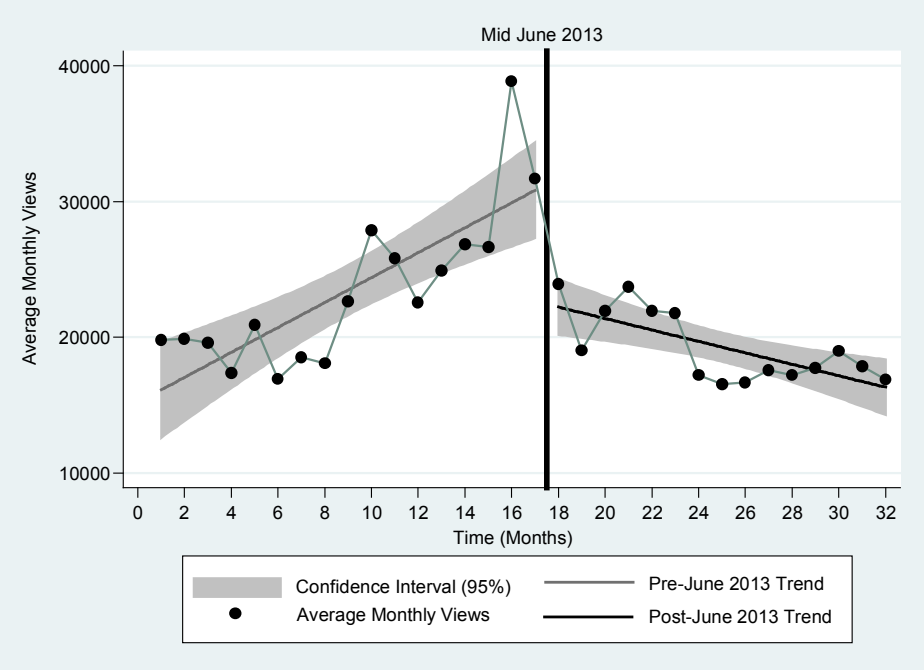


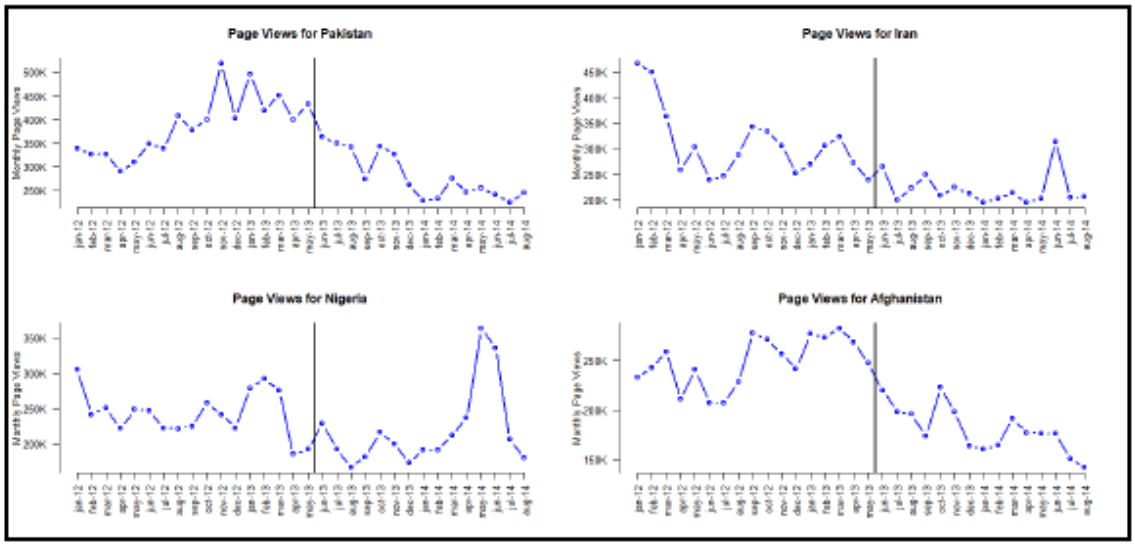
Figure 2b: Average Monthly Views for the 31 Higher Privacy Articles Plotted Alone Show Increase Until June 2013 And Then A Sharp Drop-Off (With Trend Lines)



10. In **Figure 2a** above, the average monthly page views for the 31 Higher Privacy Articles show a clear trend and not the flattened pattern reflected in Figure 5 of the Salzberg Declaration. With an appropriate scale on the vertical axis, page views increase until mid-June 2013 and declined thereafter. This point is even clearer in **Figure 2b**, which plots the very same data but adds trend lines and confidence intervals for clarity.

11. **Third**, Salzberg focuses his analysis on cherry-picked individual articles that obscure and mislead about actual trends in the data. For example, Figure 2 of the Salzberg Declaration visualizes monthly page view line-plots for four articles: Pakistan, Iran, Nigeria, and Afghanistan:

Figure 2: Individual Articles show no Association of June 2013 with a Decline in Page Views



12. While these four Wikipedia articles did form part of the 48 Terror Article set, they have among the lowest privacy-sensitivity scores among all articles in the set (*see* Table 12 of my Declaration). That is, these articles raised few privacy concerns for survey participants. As discussed in my opening declaration, 415 independent Internet users participated in a survey in which they provided feedback on how keywords associated with each of the 48 Terrorism Articles may raise privacy-related concerns ("Privacy Evaluation Survey").[5] In the Privacy Evaluation Survey, the combined average privacy-sensitivity rating for all 48 Terror Articles was 2.15 and the median was 2.07. The articles that Salzberg cherry-picked fell far below that mean and median: Pakistan (1.82), Iran (1.85); Nigeria (1.71); Afghanistan (1.83). Since the hypothesis that I tested in my study concerns a ***privacy***-based chilling effects theory [6]—i.e., that Wikipedia users avoided privacy-sensitive Wikipedia articles due to awareness of NSA Upstream Surveillance—it is inappropriate to rely on trends for these four articles, which fell far below the mean and median privacy-sensitive rating. As such, to the extent Salzberg isolates and relies on these four articles to

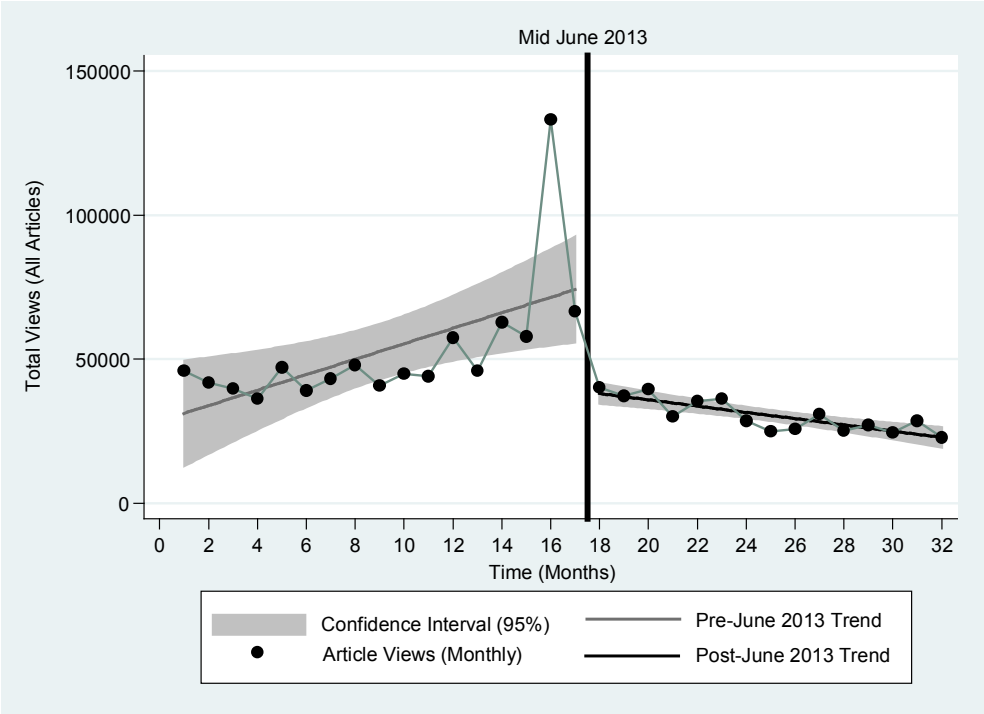
[5] See ¶¶ 32-33 of my Declaration dated December 18, 2018.
[6] See ¶¶12-21 of my Declaration dated December 18, 2018.

reject a chilling effects hypothesis (*see* paragraph 16 of the Salzberg Declaration), his analysis is misleading, unreliable, and masks actual trends found in my study.

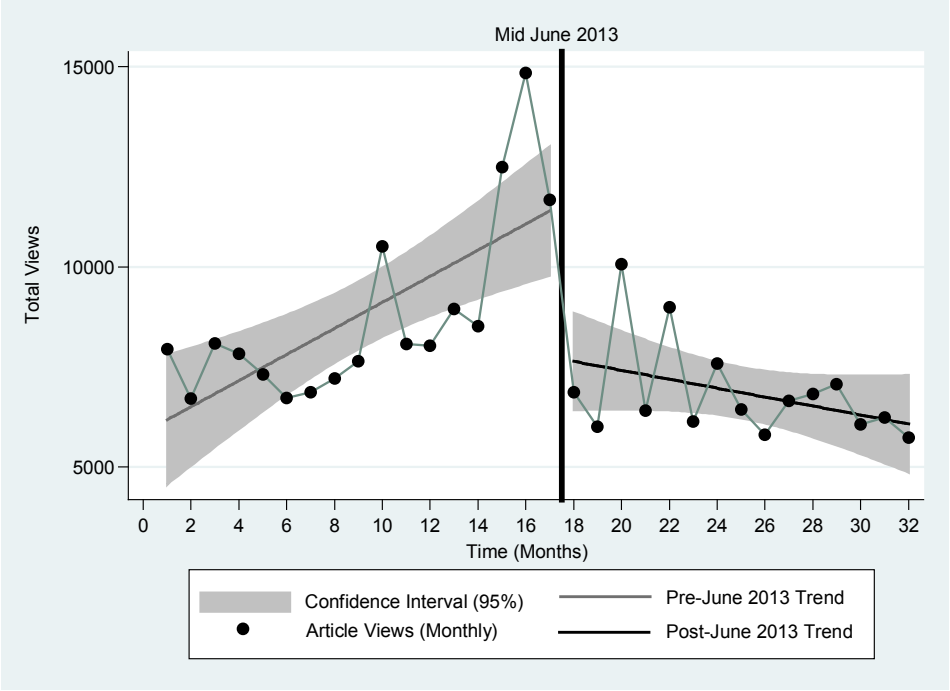
13. **Fourth**, the Salzberg Declaration focuses on disaggregated line-plots that mask aggregated data trends or less privacy-sensitive or privacy-concerning articles like the four articles noted above (Pakistan, Iran, Nigeria, and Afghanistan). This approach is inappropriate given my study tests a chilling effects hypothesis based on a privacy theory. Moreover, Salzberg's approach leads to a flawed analyses and conclusions. In fact, the page views for the four articles (among the 48 Terrorism Articles) with the very highest privacy-sensitivity scores according to the Privacy Evaluation Survey—improvised explosive device (2.86), dirty bomb (2.81), car bomb (2.81), and ammonium nitrate (2.61)—are entirely consistent with a chilling effects hypothesis in June 2013. **Figure 3a** depicts monthly page views for each of these most privacy-sensitive articles. Each figure demonstrates page view trends consistent with a chilling effects hypothesis: a monthly increase in page views leading up to June 2013, an abrupt statistically significant decline and subsequent change in trend to a monthly decrease in page views.

Figure 3a: Page Views For The Four Most Privacy-Sensitive Articles Are Consistent With A Chilling Effects Hypothesis in June 2013

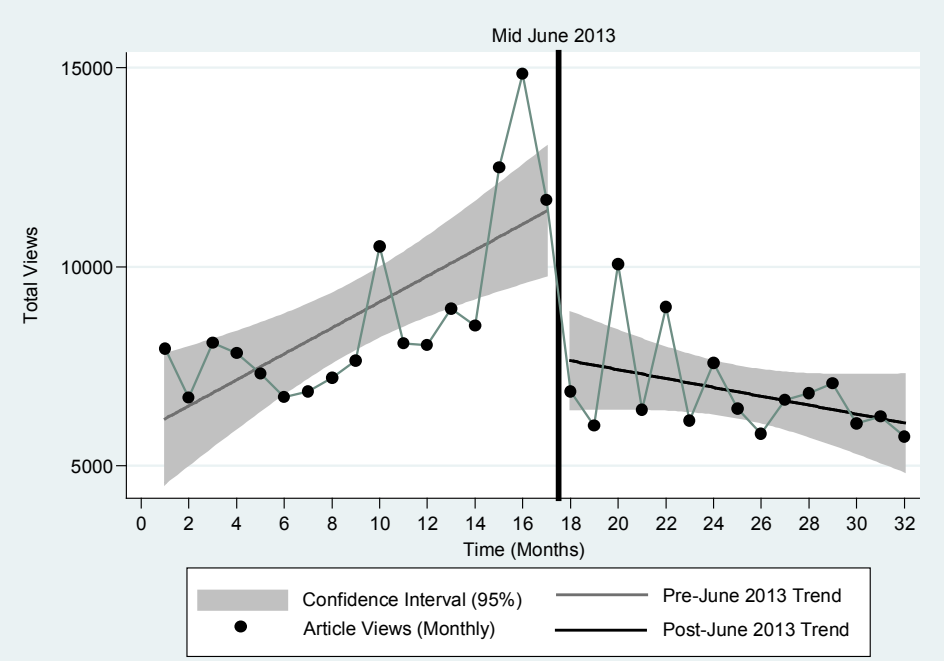
Improvised Explosive Device



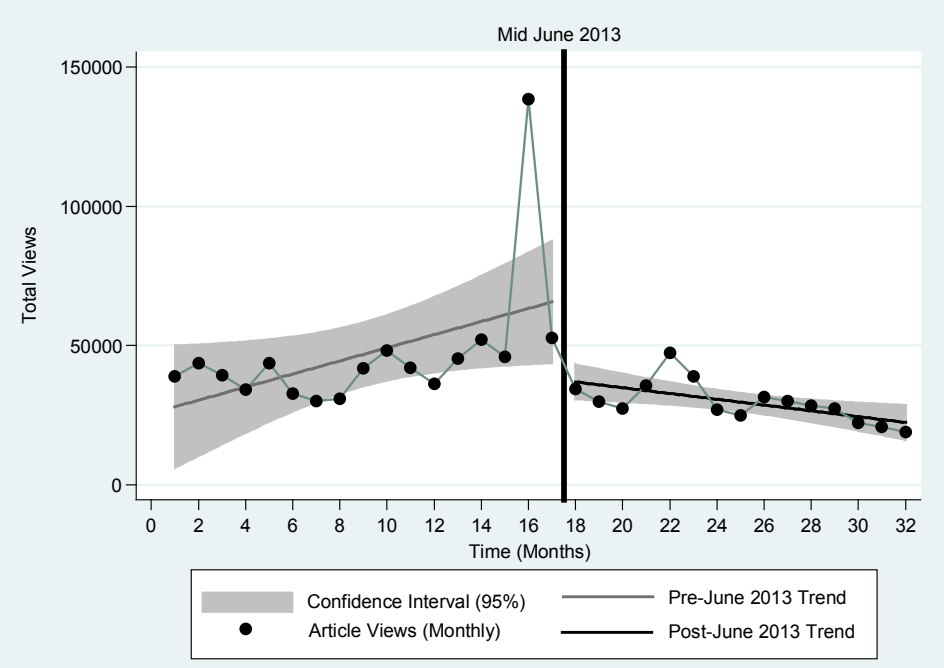
Dirty Bomb



Car Bomb

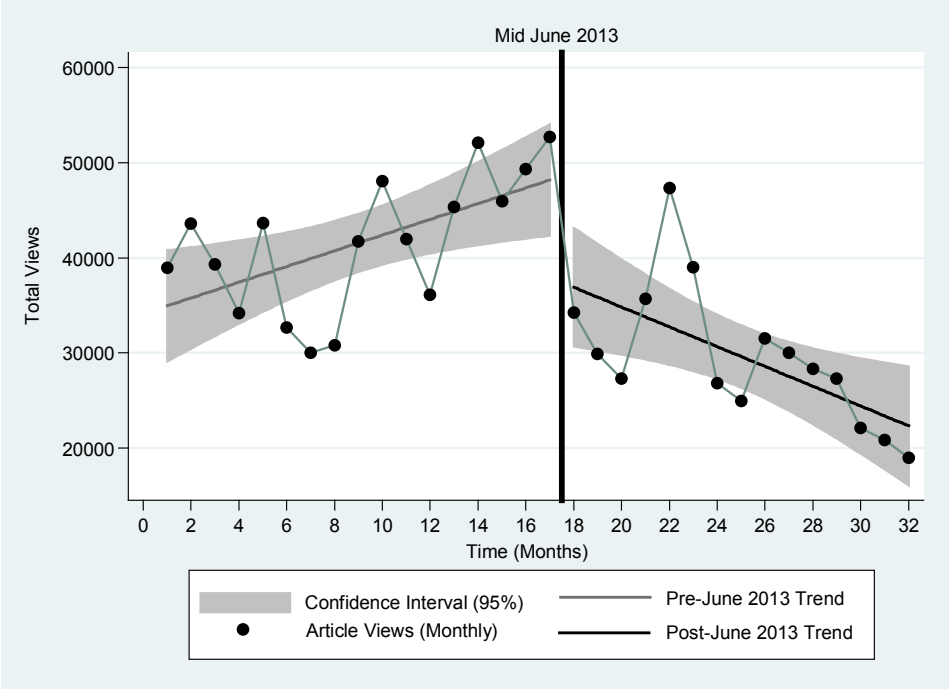


Ammonium Nitrate



14. Salzberg claims that the "ammonium nitrate" article is an outlier in its number of page views in April 2013.[7] Even assuming this is correct, if the article's page views for that month are normalized,[8] the overall trend for the ammonium nitrate article remains consistent with a chilling effect hypothesis. **Figure 3b** provides the ammonium nitrate article's normalized page views over 32 months, which are consistent with a June 2013 chilling effects hypothesis: increasing monthly articles views in the months leading up to June 2013, and then an abrupt statistically significant decline in June and a subsequent monthly reduction in views:

Figure 3b: Page Views for Normalized Ammonium Nitrate Article Are Consistent With June 2013 Chilling Effects Hypothesis



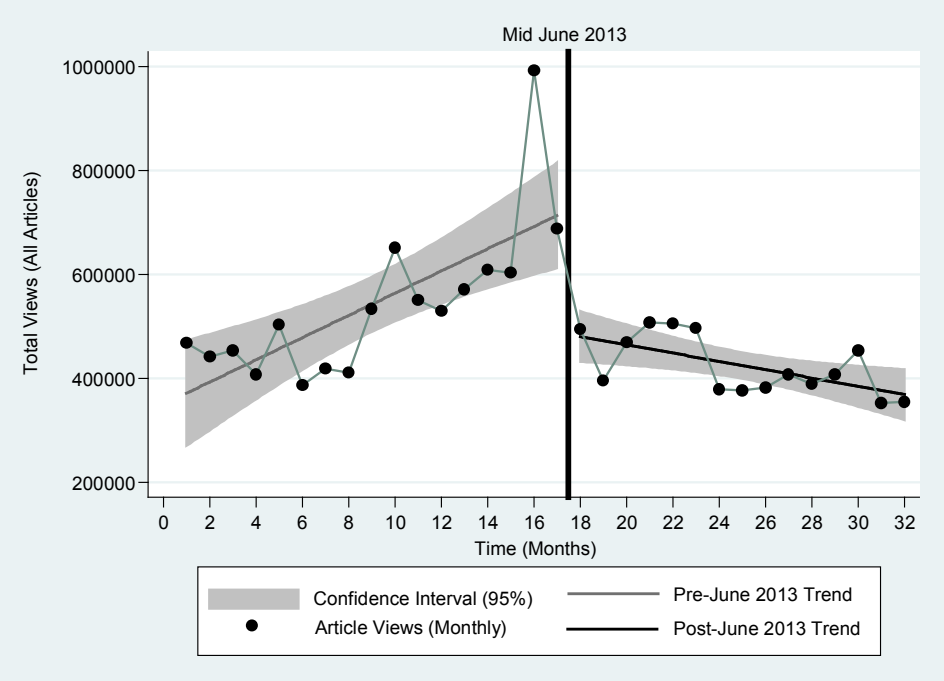
[7] See ¶ 60 of the Salzberg Declaration.

[8] I replaced the outlier value for the article in April 2013 (138363) with an average of the total page views for the article (49316) in the two adjacent months (May and March 2013). Correcting, modifying, or deleting an outlier value or observation in a data set is consistent with best-practices in dealing with outliers: HERMAN AGUINIS, RYAN K. GOTTFREDSON & HARRY JOO, *BEST-PRACTICE RECOMMENDATIONS FOR DEFINING, IDENTIFYING, AND HANDLING OUTLIERS*, Organizational Res. Methods 8, 20–23 (2014), http://orm.sagepub.com/content/early/2013/01/11/1094428112470848.abstract ("Once error outliers have been identified, the correct procedure is to either adjust the data points to their correct values or remove such observations from the dataset").

15. Furthermore, the aggregate total monthly page views, average monthly page views, and median monthly page views for the 23 most privacy-sensitive articles among the set of 48 Terrorism Articles in the study[9] are also consistent with a chilling effect hypothesis in June 2013, when measured over a 32-month period. **Figure 4** visualizes monthly median page views for these 23 most privacy sensitive Wikipedia articles:

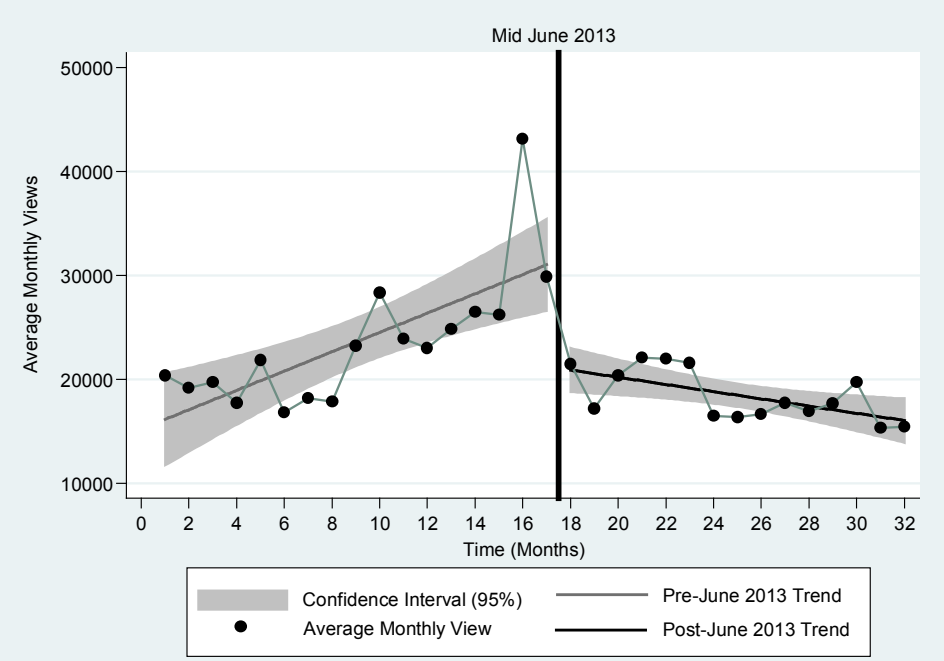
Figure 4: Total Monthly Page Views, Average Monthly Page Views, and Median Monthly Page Views for 23 Most Privacy-Sensitive Wikipedia Articles Over 32 Months Are All Consistent With A Chilling Effects Hypothesis

Raw Total Monthly Page Views: 23 Most Privacy Sensitive Articles

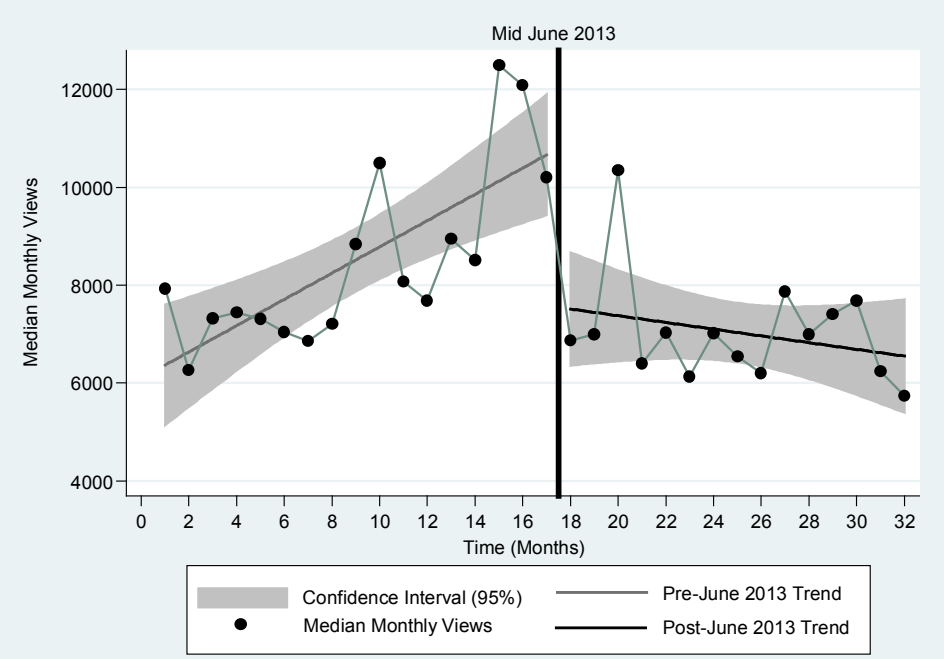


[9] This set of the 23 Most Privacy-Sensitive Article includes all articles in the 48 Terrorism Article group with a combined average privacy-sensitivity score greater than the median of those combined scores (2.07). This group includes: improvised explosive device (2.86), dirty bomb (2.81), car bomb (2.81), ammonium nitrate (2.61), biological weapon (2.60), chemical weapon (2.51), suicide attack (2.50), suicide bomber (2.44), Nuclear Enrichment (2.39), environmental terrorism (2.39), eco terrorism (2.39), weapons grade (2.39), jihad (2.35), Al Qaeda (2.34), terrorism (2.30), conventional weapon (2.27), Taliban (2.22), AL Qaeda in the Arabian Peninsula (2.17), Al Qaeda in the Islamic Maghreb (2.17), terror (2.15), Abu Sayyaf (2.14), Tehrik-i-Taliban Pakistan (2.12), and attack (2.08).

Average Monthly Page Views: 23 Most Privacy Sensitive Articles



Median Monthly Page Views: 23 Most Privacy Sensitive Articles



16. Here, in each graph set out in **Figure 4**, the total raw, average, and median monthly page views for these 23 articles increase in the time period leading up to June 2013. They also

demonstrate a statistically significant decline in June 2013, as well as a statistically significant trend reversal, with monthly page views declining after June 2013. Again, these findings are entirely consistent with a chilling effect hypothesis.

17. In the end, these results—focused on the most privacy-sensitive articles, analyzed both on an individual disaggregated analysis and aggregate monthly analysis—are entirely consistent with a chilling effect hypothesis. By ignoring the privacy theory upon which the chilling effect hypothesis is based, the Salzberg Declaration is deeply flawed.

18. **Fifth,** one of Salzberg's primary critiques of my analysis rests on a false premise: that my study "assumes a single peak in May 2013." This premise is false because my study makes no such assumption. My study hypothesizes a surveillance chilling effect in June 2013. Consistent with other ITS design studies, my study analyzes an outcome variable measured at consistent intervals (monthly privacy-sensitive Wikipedia article view data) to test that hypothesis over 32 months, by examining for statistically significant changes in level and trend in that data both before and after June 2013.

19. **Sixth,** Salzberg claims that my study's model can be altered to "prove" an April 2013 peak or earlier peak (based on a theory that the Boston Marathon bombings caused the page view trend reversal). However, he cites no cross-validation analysis to compare models and results to support his claim.[10] Cross-validation analysis is an established technique for understanding whether the results of a statistical test are robust – i.e., if we leave out datapoints at random from our dataset, do the results still hold? To answer that question and disprove Salzberg's theory, I

[10] LUKE JOHN KEELE, SEMIPARAMETRIC REGRESSION FOR THE SOCIAL SCIENCES 86 (Wiley & Sons, 2008) (describing "cross-validation" as a "general technique for assessing model fit based on resampling that can be applied to most statistical models").

conducted a "leave one out" cross-validation analysis[11] on both the 23 Most Privacy-Sensitive Wikipedia Articles set, as well as the larger 47 Terrorism Article set (the 48 Terrorism Articles without the Hamas article) to compare different statistical models based on a March, April, May, or June 2013 intervention effect. For comprehensive analysis, I used three data sets for each of these article sets—raw total page views, average monthly page views, and median monthly page views. Furthermore, for the 47 Terrorism Article Set, I excluded the "fundamentalism" article, making it a set of 46 total articles. (Salzberg noted that the "fundamentalism" article had too similar values to the "recruitment" article in the broader 48 Article set, so I have excluded it from this supplemental analysis. (Salzberg Decl. ¶ 7.)) I also performed the analysis both including and excluding the "ammonium nitrate" and "jihad" articles from the sets (Salzberg claims these two articles have outlier values).

20. The results of this cross-validation analysis show that for the 23 Article set of the Most Privacy-Sensitive Wikipedia articles, a statistical model based on a June 2013 intervention effect was superior to models based on March, April, and May 2013 interventions ***in every single data set analyzed*** (raw total monthly page views, average monthly page views, median monthly page views). That is, a June 2013 statistical model resulted in fewer estimation errors (lower room mean square errors (RMSE) and mean absolute errors (MAE)) than the other models. These results held even when the "ammonium nitrate" and "jihad" articles were removed from for analysis.

21. For the 46 Terrorism Article Set, a statistical model based on a June 2013

[11] *Id*. at 8 (describing "leave one out" cross validation as "probably the most commonly used method" as it "works well with most any sample size"; also that with "leave-one-out cross-validation, one observation is randomly selected and then omitted from the data set. The analyst then fits one of the possible models to this slightly truncated data set and calculates measure of fit. Next, a new data point is dropped, and the measure of fit is calculated again. This process is repeated as each of the data points is removed from the data set. The cross validation score is the averaged measure of model fit and can be used to compare different model specifications.")

intervention effect was also superior to models based on March, April, and May interventions for both the raw monthly page views set and the average monthly page views data set. These results held even when the "ammonium nitrate" and "jihad" articles were removed from for analysis. For the median monthly page view set, a June 2013 model was also superior to models based on both April and March 2013 interventions, but not May, where the RMSE and MAE scores for both models were very close (only a 2.5% difference in the MAE and 5% difference in the RMSE scores). These results all held or without "ammonium nitrate" and "jihad" in the sets.

22. In short, these results demonstrate the strength and robustness of my June 2013 model and its findings: it proved superior to comparable models in 46 of 48 total tests, and even in those two remaining tests, the difference in results were minimal. Moreover, when focused on the most privacy-sensitive Wikipedia articles, whether including or excluding the "jihad" and "ammonium nitrate" articles that Salzberg claims were outliers due to the Boston Marathon Bombing-- my model was a better "fit" to the data than the alternative models that Salzberg proposes based on earlier interventions in every single data set and scenario.

23. **Seventh,** Salzberg's comparative analysis of recent page view data is fundamentally undermined by the fact that Wikimedia's "page view" definition has changed over time. (*See* Salzberg Decl. ¶¶ 27-32.) Wikimedia has publicly published explanations on recent changes in the page view definition. (*See, e.g.*, Wikimedia Downloads: Analytics Datasets, https://dumps.wikimedia.org/other/analytics/; Research:Page View, https://meta.wikimedia.org/wiki/Research:Page_view.) The page view data that Salzberg relies on from July 2015 through November 2018 includes data on *mobile* page views, and therefore is incomparable to the data from the time period that I studied. As I explained in my Declaration, my study "used data for English language article view counts from stats.grok.se, an online portal

that provided access to *non-mobile* Wikipedia article view count data on a daily and monthly basis." (Decl. ¶ 34 (emphasis added); *see id.* at Table 3, 8, 9 (expressly indicating non-mobile data used.) Salzberg provides an example link to the Pageviews tool that he used to gather the more recent data, which shows that he selected all "Platform" types, including mobile. (*See* Salzberg Decl. ¶ 27, n. 17.) The difference between page views with non-mobile vs. mobile data included is often very significant, and therefore Salzberg's "extended data" comparison analysis is deeply flawed at the source and should be ignored. For example, using the "Hamas" example Salzberg offers, the difference between the "All" platforms and the "Desktop" (non-mobile) data for the month of May 2018 is over 100,000 views.

III. RESPONSES TO DR. SALZBERG'S SIX METHODOLOGICAL CRITIQUES

24. In his Declaration, Salzberg presents a series of purported critiques regarding my analysis. (Salzberg Decl. ¶¶ 47-66.) I respond to these issues below.

25. **Salzberg's first critique**: Aggregation "masks the differences in the changes over time by article" and was "performed without any analysis of the individual datasets" to determine whether it was the appropriate method. Standard methods for analyzing this kind of "panel data" were ignored. (Salzberg Decl. ¶¶ 48–50.)

26. **My Response:**

(a) As noted earlier, my method of analysis to test the June 2013 surveillance chilling effect hypothesis was an ITS design using aggregated data with segmented regression trend analysis. I chose an ITS design because it is an "ideal design" for assessing the impact of a "population-wide" intervention—like the effects of mass online government surveillance—that "affects the whole population and where randomization or a control group is impossible."[12] ITS

[12] N. BRUCE BASKERVILLE, ET AL., *IMPACT OF CANADIAN TOBACCO PACKAGING POLICY ON USE OF A TOLL-FREE QUIT-SMOKING LINE: AN INTERRUPTED TIME-SERIES ANALYSIS*, 6(1) CMAJ Open E59, E64 (2016)

design studies have also been commonly used in contexts like this one, to study information systems context (e.g., computing context)[13] and the impact of media coverage.[14] Within ITS design studies, use of segmented regression to analyze aggregated data to understand pre/post intervention trends in the data is not only "standard," but the ***recommended*** method and approach.[15]

(b) Second, there is no single determinative method or factor to decide whether an aggregated or disaggregated analysis of data is appropriate. Most ITS design studies use aggregated data,[16] because such time series designs "examine aggregate effects"[17] and are "strong designs for estimating the effects of instituting uniform, full-coverage programs or the effects of

---

(discussing "interrupted time-series design" as an "ideal design for assessing the effects of a population-wide intervention"; a "robust method for the evaluation of a policy that affects the whole population and where randomization or a control group is impossible"); RICHARD MCCLEAR ET AL., DESIGN AND ANALYSIS OF TIMES SERIES EXPERIMENTS 7, 297 (2017) (describing interrupted time series designs as "the major application of time series data for causal inference" and as a "strong quasi-experimental design… when random assignment was unfeasible"); CHESTER L. BRITT, DAVID J. BORDUA, & GARY KLECK, *A REASSESSMENT OF THE D.C. GUN LAW: SOME CAUTIONARY NOTES ON THE USE OF INTERRUPTED TIME SERIES DESIGNS FOR POLICY IMPACT ASSESSMENT*, 30 Law & Soc'y Rev. 361, 361 (1996) ("Interrupted time series designs provide one of the most common means for assessing the impact of a change in law or in social policy"); D.T. CAMPBELL, *REFORMS AS EXPERIMENTS*, 24(4) American Psychologist 409 (1969) (this seminal article by Campbell was among the first to advocate for interrupted time series designs in cases where natural experiments are not possible); WAGNER ET AL., *supra* note 1, at 308 (describing ITS designs as the "strongest, quasi-experimental designs" to estimate intervention effects in "non randomized settings").

[13] *See e.g.*, S. ASGARI & NUNES BAPTISTA, *EXPERIMENTAL AND QUASI-EXPERIMENTAL RESEARCH IN INFORMATION SYSTEMS,* IADIS International Workshop Information Systems Research Trends: approaches and methodologies (ISRTAM 2011), 20-26 July (noting ITS designs have been "used often in the field of [Information Systems]").

[14] *See e.g.,* MELANIE A WAKEFIELD, BARBARA LOKEN,& ROBERT C HORNIK, *USE OF MASS MEDIA CAMPAIGNS TO CHANGE HEALTH BEHAVIOR*, 376 The Lancet 1261, 1262-1263, (2010) (discussing interrupted time series analyses studies in the health context); RANDY ELDER ET AL., *EFFECTIVENESS OF MASS MEDIA CAMPAIGNS FOR REDUCING DRINKING AND DRIVING AND ALCOHOL-INVOLVED CRASHES*, 27(1) Am. J. Prev. Med. 57 (2004) (ROBERTO GRILLI ET AL., *MASS MEDIA INTERVENTIONS: EFFECTS ON HEALTH SERVICES UTILIZATION*, 1 Cochrane Database of Systematic Reviews 1, 1 (2002) (providing a comprehensive review of research studying the impact of media coverage on health service use— and noting that among the "twenty studies" reviewed in the work, all used interrupted time series designs).

[15] See works cited at *supra* note 2 and accompanying text.

[16] See also works cited at *supra* note 3 and accompanying text.

[17] JEFFREY M. WOOLDRIDGE, INTRODUCTORY ECONOMETRICS A MODERN APPROACH 15 (5th ed., 2012) ("…time series data are often used to look at aggregate effects. An example of a time series data set on unemployment rates and minimum wages…").

making changes in such programs."[18] NSA surveillance, including Upstream surveillance, was "uniform" and "full coverage," within the meaning of this guidance. It also has aggregate effects, as the entire U.S. Internet-using population is subject to their reach. It was therefore appropriate that my study employed aggregate page view analysis, since I sought to make aggregate level inferences about large scale NSA surveillance effects.[19]

(c) Third, aggregated data and analysis is further appropriate in ITS studies where the aim is to explore national or major regional rates and trends;[20] to reduce or remove "noise" in the data;[21] and to allow for more sophisticated statistical tests and analysis.[22] All of these circumstances apply to this study. First, my study was focused on evaluating the large-scale national aggregated effects or impact of mass awareness of NSA surveillance in June 2013 and after. Thus, examining the Wikipedia article page view data in aggregate is consistent with that aim. Second, analyzing the Wikipedia articles in aggregate helped reduce "noise" in the data given that, inevitably, individual Wikipedia article page views would like fluctuate and vary widely over 32 months; if analyzed in aggregate, broader overall trends or patterns could be discerned. Analyzing aggregated data makes particular sense in this ITS study, as Wikipedia page view data has a particularly high signal-to-noise ratio—that is, where the signal or "true" patterns in data (like longer term trends due by chilling effects) may be obscured by "noise," that is, more

---

[18] PETER ROSSI ET AL., EVALUATION: A SYSTEMATIC APPROACH 352-354 (6th ed., 1999) (noting "[m]ost existing times series involve aggregated data" as they often aim to study national, state, or large regional subjects).
[19] L. LEYDESDORFF, *THE SCIENCE CITATION INDEX AND THE MEASUREMENT OF NATIONAL PERFORMANCE IN TERMS OF NUMBERS OF SCIENTIFIC PUBLICATIONS*, 1-2 Scientometrics 111, 113 (1989) ("In general, one should prefer aggregated data for inferences at the aggregated level, since otherwise methodological problems of inference may emerge.").
[20] ROSSI ET AL., *supra* note 18, at 354.
[21] BEARD, *supra* note 4, at 2 (noting that a "[m]ajority of studies use aggregated data" as this "removes noise and allows for more sophisticated tests which require continuous or rate type data").
[22] *Id*.

temporary variations for individual article page views for other reasons.[23] Several prior studies have observed that Wikipedia page view data has such "noise" and those studies likewise used aggregated page views for analysis.[24] Third, analyzing the Wikipedia article page view data in aggregate also allowed for the more sophisticated statistical tests and analysis. In this case, that was segmented regression trend analysis pre/post June 2013, which is the recommended method of analysis for ITS design studies. In short, an aggregated analysis of the Wikipedia article page view data was both a "standard" method and entirely justifiable.

(d) Fourth, my opinions *are* supported by an analysis of individual article page views. Specifically, I examined the page view trends for individual and smaller groups of articles with higher privacy-sensitivity scores to verify the results of my aggregate data analysis showing: (1) a statistically significant drop in June 2013, and (2) trend change in monthly page views from increasing views before that month to a monthly decline after that month. As the chilling effect hypothesis I was testing is based on privacy sensitivity, then page view trends for the most privacy-sensitive Wikipedia articles would reveal page view trends consistent with the aggregate data results. As seen from **Figures 3a, 3b, and 4**, those page views were consistent with a June 2013 chilling effects hypothesis. Furthermore, I also examined and analyzed individual article page views to identify and investigate outliers. For example, this was done with the "Hamas" article, among others. Examining individual and aggregate level data allowed me to identify overly

---

[23] N. GENEROUS, ET AL., *GLOBAL DISEASE MONITORING AND FORECASTING WITH WIKIPEDIA*, 10(11) PLoS Comput Biol 1, 8, 12 (2014)(study on using Wikipedia page view data to track and predict global diseases using aggregated data as "signal-to-noise" ratio in Wikipedia page view data may mean "[t]rue patterns in data may be obscured by noise", in this case, the "noise" being variations in page views of health information on Wikipedia unrelated to the personal diagnosis); J.D. SHARPE, ET AL., *EVALUATING GOOGLE, TWITTER, AND WIKIPEDIA AS TOOLS FOR INFLUENZA SURVEILLANCE USING BAYESIAN CHANGE POINT ANALYSIS: A COMPARATIVE ANALYSIS*, 2(2) JMIR public health and surveillance 1, 2, 4 (2016) (noting the "signal-to-noise" ratio in Wikipedia page view data can be "problematic" and also aggregating).

[24] *Id*.

influential articles. This is a standard aspect of conducting regression model diagnostics. Additionally, I used individual articles as part of my cross-validation analysis.

(e) Fifth, Salzberg describes my data as "panel data." This is accurate in the broad definition of the term—my data is "longitudinal," involving repeated measurements from a group to study the impact of an intervention (monthly measurements of individual Wikipedia article page views aggregated into the larger sets).[25] However, using more precise definitions informed by the aim and design of my study, the data is more accurately described as time series data. This is because the subject of my study is the aggregate Wikipedia page view trends both before and after June 2013 in order to test the aggregate effects of NSA surveillance on Wikipedia and its users. In short, *time* is a central unit of analysis in my study.[26] By contrast, panel data and panel studies typically follow *individuals* over time, and include multiple-dimensional observations from each individual (e.g., schooling, employment, marital status, training, child rearing, health, etc.).[27] This conclusion is supported by Woolbridge in *Introductory Econometrics: A Modern Approach*, the same text Salzberg cites. It states that a "time series data set consists of observations on a variable or several variables over time," with time an "important dimension in a time series data set."[28] As I was interested in examining the aggregate effects of media coverage of NSA surveillance introduced at a specific point in time (June 2013), and trends before and after that time, the "variable" in my study was aggregate Wikipedia page views, and a central dimension was time. Woolbridge also describes panel data as longitudinal data that follows or "attempts to follow" individuals over time, collecting multiple observations on a range of different data points,

[25] CHENG HSIAO, PANEL DATA ANALYSIS 1 (CUP, 2017); ROSSI ET AL., *supra* note 18, at 352.
[26] GREGORY B. MARKUS, ANALYZING PANEL DATA 7-8 (sage, 1979) (describing the difference between time series data and panel data, with the former having "time" as the central unit of analysis, while panel studies take individuals as the central unity of analysis).
[27] HSIAO, *supra* note 25, at 1; MARKUS, *id* 7-8.
[28] WOOLBRIDGE, *supra* note 17, at 8.

with repeated measures not necessarily uniform or taken at regular intervals.[29] This creates additional layers of complexity for analysis suited for panel studies.[30] The "standard" methods for that Salzberg recommends at paragraph 50 of his Declaration, in fact, concern panel data analysis, specifically data collected at only two or three points in time.[31] These are not applicable to my time series data set collected at regular intervals over 32 months.

27. **Salzberg's second critique**: Penney's model "assumes a single peak in May 2013" rather than "letting the data reveal where, if anywhere a peak in the data exists." Penney's model can be altered to "prove" an April 2013 peak and support the theory that the Boston Marathon bombings caused the page view trend reversal. A "polynomial model" further shows that Penney's hypothesized peaks in page views were incorrect. (Salzberg Decl. ¶¶ 51–54.)

28. **My Response:**

(a) First, my ITS design does not "assume a single peak"—it tests for the effects of real-world events happening at a particular time—June 2013. In order to test a surveillance chilling effect hypothesis in that month, I examined page view trends before an intervention point. This is a standard approach in naturalistic studies like this, where the aim is to test the impact of an intervention at a given point in time (here, NSA surveillance revelations and media coverage in June 2013). Any peak before or after June 2013 arises from the data itself, and is not any assumption or requirement in my model.

(b) Second, Salzberg's approach of "letting the data reveal where, if anywhere, a peak in the data exists," is not a sound or reliable social scientific approach and can lead to substantial bias in results. My ITS study, designed to examine a June 2013 impact due to media

[29] WOOLBRIDGE, *supra* note 17, at 448.
[30] WOOLBRIDGE, *supra* note 17, at 448.
[31] WOOLBRIDGE, *supra* note 17, at 459-474.

coverage of NSA surveillance programs, was based on a hypothesis grounded in existing empirical and theoretical research on privacy, surveillance studies, and chilling effects. By contrast, Salzberg's approach of visually inspecting data and running various statistical models, including a "polynomial model," until a "fit" showing "earlier peaks in 2013" is found, is a biased approach.[32] Any earlier "peak" or polynomial model is not grounded on any *a priori* hypothesis, theory or research. Nevertheless, even assuming a "polynomial model" estimating earlier peaks in 2013 is not biased. It also does not discount or disprove a chilling effects hypothesis in June 2013. My statistical model and results did not require or assume any such "peak" in May. A statistically significant drop in June 2013 based on prior trends, or a reduction in monthly views thereafter, would each be consistent with a chilling effect hypothesis, notwithstanding any earlier peaks in the page view data in the 32 months. My analysis found both, but neither required any "peak" in May or April 2013.

(c) Third, my statistical model based on a June 2013 intervention is a superior fit for the page view data compared to models based on March, April, and May 2013 interventions in every single data set analyzed. This is demonstrated by my cross-validation analysis on both the 23 Article Set (raw monthly page views, average monthly page views, median monthly page views) and the 28 Article Set (raw monthly page views, average monthly page views, median monthly page views).[33] My June 2013 model was a better "fit" when focused on the page view data for the most privacy-sensitive Wikipedia articles compared to Salzberg's alternative models,

---

[32] ANDREW GELMAN & ADAM ZELIZER, *EVIDENCE ON THE DELETERIOUS IMPACT OF SUSTAINED USE OF POLYNOMIAL REGRESSION ON CAUSAL INFERENCE*, Research & Politics (2015) (describing reported effects, based on a "curve fitting" polynomial model that is "statistically significant but substantively dubious, and are sensitive to model choice"); MEGAN L. HEAD ET AL., *THE EXTENT AND CONSEQUENCES OF P-HACKING IN SCIENCE*, 13(3) PLoS Biol 1 (2015) ("Inflation bias… occurs when researchers try out several statistical analyses and/or data eligibility specifications and then selectively report those that produce significant results.").

[33] *See supra* ¶¶ 16-17.

notwithstanding any impact of the Boston Marathon media coverage in April or May 2013. In short, my ITS approach, statistical regression model, and method of analysis is the best way to understand the page views trends for the most privacy-sensitive Wikipedia articles.

29. **Salzberg's third critique:** Penney's model is "oversimplified, leaving out virtually all factors that could affect page views of terror-related articles from the model." For example, the model fails to account for seasonality or major news events. The model tacitly acknowledges this failure in how it handles the Hamas outlier data, which is ultimately manipulated in a way that is favorable to the hypothesis. Penney does not consider other real world variables that may not be favorable to the hypothesis, like the Boston Marathon bombings, which happened shortly before the NSA disclosures. (Salzberg Decl. ¶¶ 55–60.)

30. **My Response:**

(a) My analysis ***does*** account for various external factors that may affect page views in the model. First, while seasonality is a confounding concern in ITS designs, there is no basis to expect large seasonal effects with these page views—that, for example, Internet users tend to view terrorism-related content in the spring but not in the summer. In any event, to account for possible seasonality and seasonal effects in the data in the ITS design, I went beyond the "general recommendation" for 12 data points before and after the hypothesized intervening chilling effect (June 2013),[34] and instead collecting data for 17 data points before (January 2012 through May 2013) and 14 data points following (July 2013 through August 2014). This longer study period allows for better assessment of overall page view trends by identifying and accounting for seasonal trends. This is particularly challenging with Wikipedia page view data because of its high signal-

[34] WAGNER ET AL., *supra* note 2, at 301 ("A general recommendation is for 12 data points before and 12 data points after the intervention (8), although this number is not based on estimates of power. Rather, with 24 monthly measures, the analyst can adequately evaluate seasonal variation.").

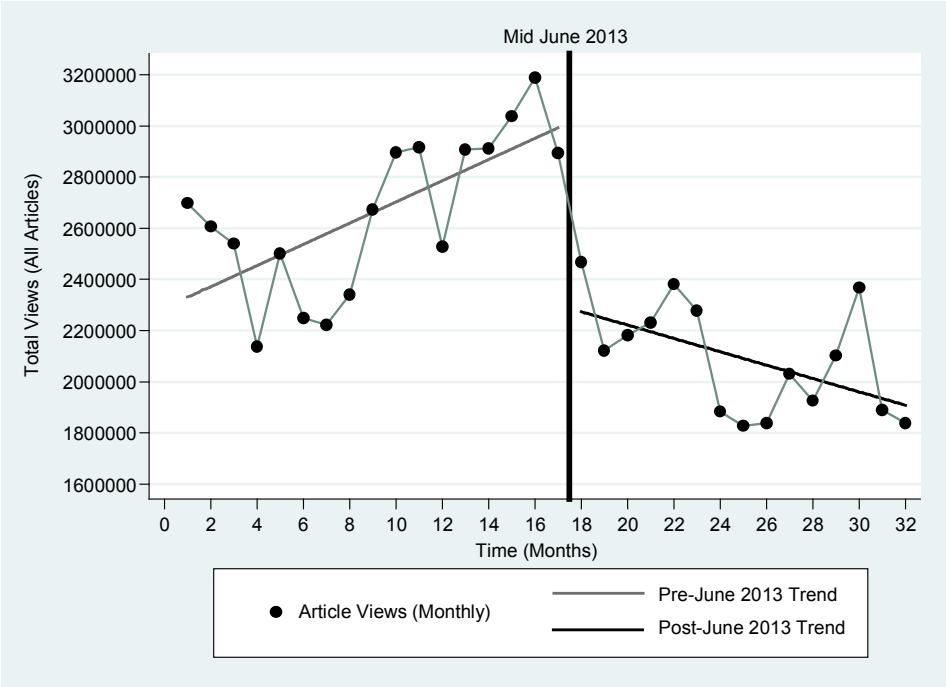
to-noise ratio, that is, variability of page views over the course of a study period due to a range of daily factors.[35]

(b) Second, the data ***was*** also analyzed for seasonal trends as well for real-world events to determine whether any such events would have an outsized effect on page views. This is precisely how I identified the "Hamas" article as an outlier and related it to real world events. This was not done by cherry-picking or manipulation, but through standard regression model diagnostics, as well as best practices for identifying and dealing with outlier observations.[36]

(c) Third, no seasonal or "real world" event-related variations identified by Salzberg explain the actual trends apparent in the aggregated data before and after June 2013. Salzberg points to a seasonal "trough" in the summer of 2012 and a "peak" due to the Boston Marathon bombing in April 2013 for page views in the 47 Terrorism articles. These are visible in **Figure 5** below:

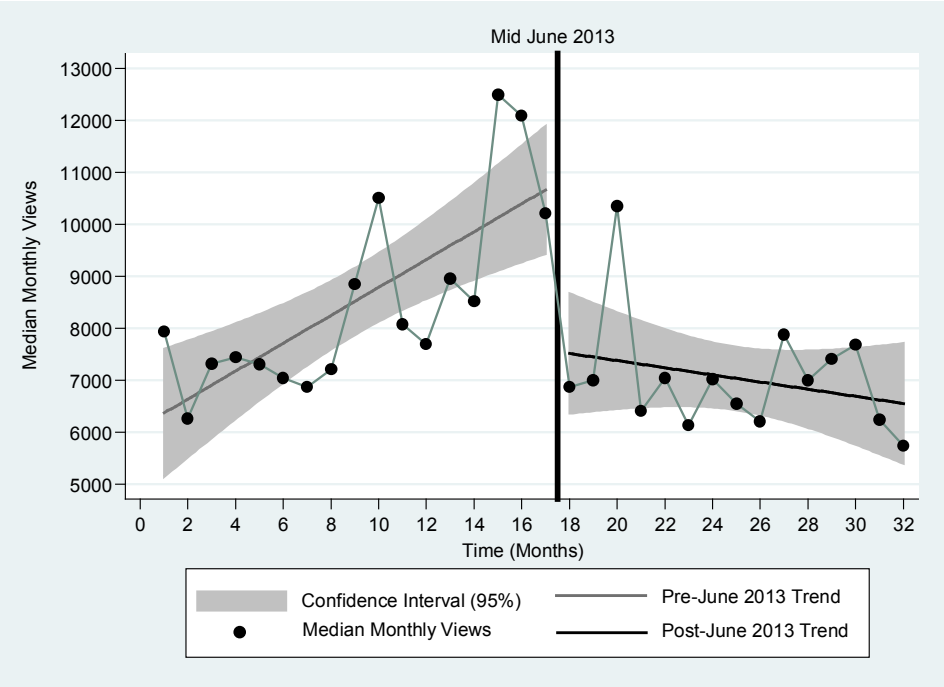
[35] See works cited at *supra* note 23 and accompanying text.
[36] The Hamas article was identified using standard regression model diagnostics, including examining cook's distance, leverage, residual values, z-scoes, among others. Handling of the influential outlier was also done according to best practices, with the outlier removed but results before and after removal conveyed. See AGUINAS ET AL., *supra* note 8, at 20-23.

Figure 5: Aggregated Total Monthly Page Views for 48 Terror Articles (Without Hamas)



However, the added trend analysis of page views in **Figure 5** demonstrates that despite these possibly seasonal and "real world" event-related variations over the course of the 17 months prior to June 2013 and 14 mounts after that period, there is no variation or "trough" comparable to the one that occurs between the total page views as of the beginning of June 2013 (2,893,553 page views) and that at the end of July 2013 (2,121,335). Indeed, during time period there was a decline of 772, 218 page views, or roughly 27%. By contrast, the decline between April and May 2013 that Salzberg highlights (294, 820 page views) is slightly greater than 9%. Nor is there any point in the entire 32-month study period where there are consistently fewer monthly page views than in the months of December 2013 through August 2014. These points are consistent with a June 2013 chilling effects hypothesis. These points are even clearer when examining the median monthly page views for the 23 Most Privacy Sensitive Articles visualized in **Figure 6:**

Figure 6: Aggregated Median Monthly Page Views for 23 Most Privacy-Sensitive Articles



(d) Here, again, there is a significant drop from the median monthly page views as of the beginning of June 2013 (12,090) and the end of July 2018 (6,864) totaling 5,226, which is slightly greater than a 45% drop in median monthly views. Nor, despite some variation, is there any other period in the 32 months where monthly median page views are trending so clearly lower over time as the months after June 2013 onto August 2014. Again, this is consistent with a chilling effects hypothesis. Salzberg offers no alternative explanation beyond identifying "peaks" and "troughs" in 2012 and 2013, which do not account or explain these findings.

(e) Fourth, in a naturalistic study outside the experimental context, it is not possible to control for all confounding factors, like the impact of all real world events on page views over 32 months. However, the ITS design was chosen for the very fact that its pre/post design *can* help control for other explanatory factors as any such known or unknown confounding variables would be present in both the pre and post measurements (monthly page views), thus any

changes after the intervention (June 2013) can be attributed to intervention itself. This dimension of the ITS design can be further strengthened by adding one or more comparators,[37] which was also done in this study. Comparators help control for confounding factors and seasonality.[38] Here, the comparator groups—security, infrastructure, popular, and Wikipedia English home-page traffic—not expected to be impacted or affected by the ITS design intervention (here, surveillance chilling effects in June 2013) can be compared to the study group (page views for privacy-sensitive terrorism related Wikipedia articles). As my results showed, the privacy-sensitive terrorism Wikipedia were impacted (statistically significant drop in June 2013 and trend reversal after that month) while none of the comparator articles showed the same effects. This is consistent with a chilling effect hypothesis.

(f) Salzberg claims that my comparator groups do not corroborate my findings as they are not "proper" controls groups that "exhibit the trend" shown by the terrorism articles before June 2013.[39] Of course, comparator groups identical to the study group are ideal but are often not feasible.[40] In fact, Campbell and Stanley, in their leading 1966 text *Experimental and Quasi-Experiment Designs for Research* that Salzberg cites, recommend that for an ITS study of a "major administrative change" that a researcher use a "similar institution" as a comparator not expected to undergo the change and upon which the same ITS design can be tested.[41] Wagner et al.'s (2002) leading article on ITS design recommends that where an identical control group is not

[37] R. BARKER BAUSELL, *THE DESIGN AND CONDUCT OF MEANINGFUL EXPERIMENTS INVOLVING HUMAN PARTICIPANTS 199* (OUP, 2015) (noting that an ITS design can be "significantly buttressed" by adding one or more comparator group); MARY A. M. ROGERS, *COMPARATIVE EFFECTIVENESS RESEARCH 94* (OUP, 2014).
[38] BAUSELL, *id.* at 200; Wagner et al., *supra* note 2, at 306.
[39] Salzberg Decl. at ¶¶ 33-46.
[40] BAUSELL, *id.* at 199; Wagner et al., *supra* note 2, at 306.
[41] DONALD CAMPBELL & JULIAN STANLEY, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH 55* (Houghton-Mifflin, 1966).

possible, a study may examine a "different but related group . . . not expected to change following the intervention, in the same group of subjects."[42] This is precisely the methodology that I adopted. Since identical or randomly sampled control groups were not possible, I used comparator groups drawn from a "different but related group" of Wikipedia articles drawn from a "different but related" group of DHS key words (security and infrastructure), as well as a set of popular Wikipedia articles. Given that security, infrastructure, and popular Wikipedia articles do not raise privacy concerns, they would not be expected to change after a surveillance chilling effect intervention in June 2013. We would also expect similar viewer audiences for these articles (e.g. someone with an interest in national security would be just as likely to view Wikipedia articles on terrorism as domestic or infrastructure security articles, the comparators in this study). My results showed that they were not impacted, while the privacy-sensitive terrorism-related articles were. This, as noted, was consistent with a surveillance chilling effect hypothesis.

31. **Salzberg's fourth critique:** The model did not take into account that the 48 terror articles chosen based on 2011 DHS list would naturally rise and decline in interest over time. In other words, the 2011 terrorism-related key words would "undoubtedly become stale over time." The same is true for the trends in the comparator article groups. (Salzberg Decl. ¶¶ 61–64.)

32. **My Response:**

(a) The articles in my study were chosen based on keywords associated with "terrorism" that DHS uses to track and monitor social media. Since the media coverage relating to the Snowden revelations framed the issue of NSA surveillance as a matter of national security and terrorism threats, it was logical for me to use DHS keywords associated with "terrorism" to create the Wikipedia articles that represent the sort of articles that users may be chilled from accessing

[42] WAGNER ET AL., *supra* note 2, at 306.

in light of privacy concerns about government surveillance. I used this approach for pragmatic methodological reasons, as there is no pre-determined list reflecting all privacy-sensitive Wikipedia articles from which to draw a random sample. Furthermore, cherry-picking a list of articles relating to certain sensitive topics (like Syria or ISIL, as Salzberg suggests) would be subject to serious selection biases. Using government keyword lists to study government surveillance is not a novel or unprecedented approach;[43] indeed, my methodology is similar to the methodology of an earlier, peer-reviewed study exploring the chilling effects associated with the NSA surveillance in Google search data, which also used these DHS keyword lists.[44]

(b) The fact that some likely privacy-sensitive articles (like Syria and ISIL) that were not included in the DHS keyword list may have recorded higher page views for some period of time during my study does not in any way undermine the overall conclusions for the privacy-sensitive Wikipedia articles examined over the entire 32-month period that I studied.

(c) Salzberg's Declaration offers no evidence for his claim that the "many of the 2011 terrorism-related keywords undoubtedly became stale over time" and thus "page views dropped."

33. **Salzberg's fifth critique:** The page view data examined only extends for 32 months through August 2014, therefore the "results do not and cannot imply that an effect of the June 2013 disclosures persists today, or did so even in 2015." (Salzberg Decl. ¶ 65.)

---

[43] JEDIDIAH R. CRANDALL & MASASHI CRETE-NISHIHATA ET AL., *CHAT PROGRAM CENSORSHIP AND SURVEILLANCE IN CHINA: TRACKING TOM-SKYPE AND SINA UC*, First Monday, July 1, 2013, http://firstmonday.org/ojs/index.php/fm/article/view/4628/3727 [https://perma.cc/M5FJ-T4D5]; JEFFREY KNOCKEL, JEDIDAH CRANDALL & JARED SAIA, *THREE RESEARCHERS, FIVE CONJECTURES: AN EMPIRICAL ANALYSIS OF TOM-SKYPE CENSORSHIP AND SURVEILLANCE*, 16:4 FOCI '11: USENIX Workshop on Free & Open Comm. on Internet (2011), https://www.cs.unm.edu/~crandall/foci11knockel.pdf [https://perma.cc/FH8H-JUBA].

[44] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR*, *IN* CAMBRIDGE UNIVERSITY HANDBOOK ON SURVEILLANCE LAW (David Gray et al. eds., 2017).

34. **My Response:**

(a) The statistically significant trend reversal from increasing monthly views prior to June 2013 to a downward trend, with a monthly reduction in page views afterwards, is indicative of a lasting chilling effect. This is supported by other research on long-term online chilling effects due to public awareness of surveillance.

(b) First, an Massachusetts Institute of Technology (MIT) study on Google search data later published a peer reviewed chapter in the *Cambridge University Handbook on Surveillance Law*, found a statistically significant reductions in privacy-sensitive Google searches after the June 2013 Snowden disclosures about NSA surveillance.[45] The findings, the authors concluded, provided "substantial empirical documentation of a chilling effect," both in the "shorter term" and "in the longer term", that "appeared to be related to increased awareness of government surveillance online."[46]

(c) Second, a 2017 peer reviewed study on Wikipedia editors found evidence, based on qualitative interviews in the spring and summer of 2015, that editors were chilled from certain activities on Wikipedia due to awareness of government surveillance.[47] For example, one Wikipedia editor stated, "for the Edward Snowden page, I have pulled myself away from adding sensitive contributions, like different references, because I thought the name may be traced back to me in some way." The fact that some Wikipedia users have avoided sensitive or controversial topics in 2015, two years after the Snowden revelations in 2013, is consistent with my findings.

---

[45] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR* 1, 3-4 (MIT Sloane Working Paper No. 14380, 2015); MARTHEWS & TUCKER, *supra* note 44.
[46] ALEX MARTHEWS & CATHERINE TUCKER, *GOVERNMENT SURVEILLANCE AND INTERNET SEARCH BEHAVIOR* 1, 3-4 (MIT Sloane Working Paper No. 14380, 2015); MARTHEWS & TUCKER, *supra* note 44.
[47] ANDREA FORTE, NAZANIN ANDALIBI, AND RACHEL GREENSTADT, *PRIVACY, ANONYMITY, AND PERCEIVED RISK IN OPEN COLLABORATION: A STUDY OF TOR USERS AND WIKIPEDIANS*, *in* CSCW 1800 (2017).

(d) Third, a recent 2018 study exploring how journalists have been impacted by "potential surveillance by government," which involved qualitative interviews with American journalists in 2015, found all seven journalists in the study indicated that "their work and lives have changed under a real or perceived threat of mass government surveillance."[48] The author concluded that "participants reported an increased awareness of mass government surveillance" and "[i]n every case, they reported adjusting their behavior to some degree."[49] This is also consistent with my findings.

(e) Fourth, a Pew Research Center survey of 475 adult Americans conducted between November 26, 2014 and January 3, 2015 found that, among the 87% of respondents aware of "government surveillance programs" due to the Snowden revelations, 34% had taken "at least one step to hide or shield their information from the government," including avoiding using "certain terms in online communications."[50] The survey also found 25% changed the patterns of their own use of various online platforms "a great deal" or "somewhat" since the Snowden revelations. These findings from a survey administered in late 2014 and early 2015 are also consistent with my conclusions.

(f) Fifth, a PEN America survey of 520 American writers in October 2013[51] found that 28% of the writers surveyed had "curtailed or avoided" certain online activities due to "fear of surveillance" and another 12% "seriously considered" doing so; 24% "deliberately avoided certain topics in phone or email conversations," and another 9% have "seriously

---

[48] STEPHENSON WATERS, *THE EFFECTS OF MASS SURVEILLANCE ON JOURNALISTS' RELATIONS WITH CONFIDENTIAL SOURCES, DIGITAL JOURNALISM,* 6:10 Digital Journalism 1294, 1310 (2018).
[49] *Id*. at 1310.
[50] LEE RAINIE ET AL., *PEW RES. INTERNET PROJECT*, Americans' Privacy Strategies Post-Snowden 4 (Mar. 16, 2015), http://www.pewinternet.org/files/2015/03/PI_AmericansPrivacyStrategies_0316151.pdf.
[51] FDR GROUP & PEN, AMERICAN CENTER, *CHILLING EFFECTS: NSA SURVEILLANCE DRIVES U.S. WRITERS TO SELF-CENSOR 3–4* (2013), http://www.pen.org/sites/default/files/Chilling%20Effects_PEN%20American.pdf.

considered it"; and 16% have refrained from "conducting Internet searches or visiting websites on topics that may be considered controversial or suspicious," and another 12% have "seriously considered it." These results are consistent with my conclusions as to a long term chilling effect.

(g) Also, as explained above, Salzberg's extended comparison analysis that relies on more recent page view data is fundamentally invalid because it compares across page view definitions—the more recent data includes "mobile" page views, while my study relied on "non-mobile" data.[52]

35. **Salzberg's sixth critique:** Penney's model fails to isolate the "particular effect of public 'awareness' about the NSA Upstream program" from potential other effects of the Snowden disclosures, including increased awareness about other NSA surveillance activities. (Salzberg Decl. ¶ 66.)

36. **My Response:**

(a) In any study of naturalistic changes in human behavior, it will not be possible to isolate the source of all causes and effects on behavior. It is enough for purposes of establishing whether Upstream likely had a chilling effect on Wikipedia users that the reporting on NSA surveillance in June 2013 included multiple references to international Internet communications monitoring, and that general public awareness of NSA surveillance grew due to media coverage after June 2013.[53] As earlier noted, a Pew Research Center survey of adult Americans conducted between November 26, 2014 and January 3, 2015 found 87% of respondents were aware of NSA surveillance programs due to the Snowden revelations.[54]

(b) Furthermore, in-line with the empirical conclusions of my Declaration,

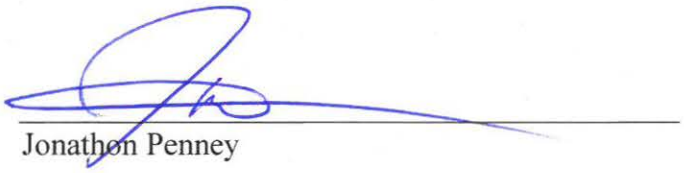
[52] *See supra* ¶ 23.
[53] See ¶¶ 32-33 of my Declaration dated December 18, 2018.
[54] *See supra* ¶ 30(e).

Wikimedia has introduced other evidence establishing the particular chilling effect that awareness of Upstream surveillance had on Wikimedia's readers and contributors. The Declarations of Michelle Paulson and James Alexander describe the chilling effect that Upstream surveillance had on the Wikimedia community at large, particularly among users abroad who engage with the platform concerning privacy-sensitive topics. (*See* Pl.'s Ex. 3, ¶¶ 41, 45, 46; Pl.'s Ex. 4, ¶¶ 4-11.) The Declaration of Emily Temple-Wood, an active Wikimedia community member, further describes first-hand the chilling effect that awareness of Upstream surveillance has had among the community of readers and contributors. (*See* Pl.'s Ex. 6, ¶¶ 20-21.)

(c) Salzberg's critique that the "particular effect" of Upstream cannot be entirely isolated is not actually a methodological critique, but rather, a general observation about a naturalistic studies. However, courts have rejected such challenges when ruling on *Daubert* motions. *See A Woman's Choice-East Side Women's Clinic v. Newman,* 904 F. Supp. 1434 (S.D. Ind. 1995) (upholding naturalistic study against *Daubert* challenge when ruling on preliminary injunction motion).

I declare under penalty of perjury under the laws of the United States that the foregoing is true and correct to the best of my knowledge and belief.

Executed on March 8, 2019 in Halifax, Canada.



Jonathon Penney

APPENDIX

23 Most Privacy Sensitive Article Set Cross Validation Analysis

RAW total monthly views

```
.
. reg HP23RawViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.82
       Model |  3.1487e+11     3  1.0496e+11           Prob > F      =  0.0000
    Residual |  1.8571e+11    28  6.6324e+09           R-squared     =  0.6290
-------------+------------------------------           Adj R-squared =  0.5893
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   81439

------------------------------------------------------------------------------
HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   21383.58   4031.855     5.30   0.000      13124.7    29642.46
intervention |  -224931.2   58212.11    -3.86   0.001    -344173.3   -105689.1
   postslope |  -29367.59   6320.044    -4.65   0.000    -42313.61   -16421.56
       _cons |   349787.6   41314.22     8.47   0.000     265159.3      434416
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   89506.354
Mean Absolute Errors     |   63503.274
Pseudo-R2                |      0.49622
----------------------------------------

. reg HP23RawViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.34
       Model |  2.8498e+11     3  9.4993e+10           Prob > F      =  0.0000
    Residual |  2.1560e+11    28  7.6998e+09           R-squared     =  0.5693
-------------+------------------------------           Adj R-squared =  0.5232
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   87749

------------------------------------------------------------------------------
   HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |    22020.4   4758.837     4.63   0.000     12272.36    31768.43
interventionMAY |  -152067.1   62229.88    -2.44   0.021    -279539.2   -24594.96
   postslopeMAY |  -34404.22   6730.013    -5.11   0.000    -48190.03   -20618.42
          _cons |   345966.8   46015.77     7.52   0.000     251707.7    440225.8
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   100292.59
Mean Absolute Errors     |   71401.914
Pseudo-R2                |      0.37718
----------------------------------------
```

```
. reg HP23RawViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.66
       Model |  2.4087e+11     3  8.0289e+10           Prob > F      =  0.0003
    Residual |  2.5971e+11    28  9.2752e+09           R-squared     =  0.4812
-------------+------------------------------           Adj R-squared =  0.4256
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   96308

------------------------------------------------------------------------------
     HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |    13623.8   5755.506     2.37   0.025     1834.181    25413.42
interventionAPRIL |   65634.32   68033.27     0.96   0.343    -73725.51    204994.1
   postslopeAPRIL |   -34751.7   7473.906    -4.65   0.000    -50061.31    -19442.1
            _cons |   393547.5   52329.76     7.52   0.000     286354.8    500740.1
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   108143.33
Mean Absolute Errors     |   69579.447
Pseudo-R2                |     0.28104
----------------------------------------

. reg HP23RawViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.46
       Model |  2.3801e+11     3  7.9335e+10           Prob > F      =  0.0004
    Residual |  2.6257e+11    28  9.3775e+09           R-squared     =  0.4755
-------------+------------------------------           Adj R-squared =  0.4193
       Total |  5.0057e+11    31  1.6148e+10           Root MSE      =   96837

------------------------------------------------------------------------------
     HP23RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   13450.41   6420.254     2.09   0.045     299.1154     26601.7
interventionMARCH |   88472.13   68403.98     1.29   0.206    -51647.07    228591.3
   postslopeMARCH |  -33519.11   7782.969    -4.31   0.000     -49461.8   -17576.42
            _cons |   394472.2   54666.49     7.22   0.000       282493    506451.4
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   105572.3
Mean Absolute Errors     |   69332.589
Pseudo-R2                |     0.30842
----------------------------------------
```

Average monthly views

```
.
. reg HP23AvgViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.83
       Model |   595220203     3   198406734           Prob > F      =  0.0000
    Residual |   351044760    28  12537312.9           R-squared     =  0.6290
-------------+------------------------------           Adj R-squared =  0.5893
       Total |   946264964    31  30524676.2           Root MSE      =  3540.8

------------------------------------------------------------------------------
HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   929.7206   175.2961     5.30   0.000     570.6429    1288.798
intervention |  -9779.688   2530.932    -3.86   0.001    -14964.07   -4595.308
   postslope |  -1276.853   274.7814    -4.65   0.000    -1839.717   -713.9885
       _cons |   15208.16    1796.25     8.47   0.000     11528.71    18887.61
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   3891.5395
Mean Absolute Errors     |   2760.9415
Pseudo-R2                |     0.49623
----------------------------------------

. reg HP23AvgViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.34
       Model |   538718524     3   179572841           Prob > F      =  0.0000
    Residual |   407546439    28    14555230           R-squared     =  0.5693
-------------+------------------------------           Adj R-squared =  0.5232
       Total |   946264964    31  30524676.2           Root MSE      =  3815.1

------------------------------------------------------------------------------
   HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   957.4059   206.9046     4.63   0.000     533.5811    1381.231
interventionMAY |  -6611.619   2705.628    -2.44   0.021    -12153.85   -1069.391
   postslopeMAY |  -1495.838   292.6072    -5.11   0.000    -2095.217   -896.4595
          _cons |   15042.05   2000.672     7.52   0.000     10943.86    19140.24
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4360.523
Mean Absolute Errors     |   3104.3749
Pseudo-R2                |     0.37718
----------------------------------------

. reg HP23AvgViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.66
       Model |   455331527     3   151777176           Prob > F      =  0.0003
    Residual |   490933436    28    17533337           R-squared     =  0.4812
-------------+------------------------------           Adj R-squared =  0.4256
```

```
       Total |   946264964    31  30524676.2           Root MSE      =  4187.3

------------------------------------------------------------------------------
     HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   592.3357    250.238     2.37   0.025     79.74639    1104.925
interventionAPRIL |   2853.676   2957.952     0.96   0.343    -3205.414    8912.766
   postslopeAPRIL |  -1510.946   324.9507    -4.65   0.000    -2176.577   -845.3148
            _cons |   17110.78   2275.194     7.52   0.000     12450.26    21771.31
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4701.8614
Mean Absolute Errors     |   3025.1117
Pseudo-R2                |     0.28105
----------------------------------------

. reg HP23AvgViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.46
       Model |   449922646     3   149974215           Prob > F      =  0.0004
    Residual |   496342317    28  17726511.3           R-squared     =  0.4755
-------------+------------------------------           Adj R-squared =  0.4193
       Total |   946264964    31  30524676.2           Root MSE      =  4210.3

------------------------------------------------------------------------------
     HP23AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   584.7802   279.1393     2.09   0.045     12.98935    1156.571
interventionMARCH |   3846.882   2974.063     1.29   0.206    -2245.209    9938.974
   postslopeMARCH |  -1457.347   338.3873    -4.31   0.000    -2150.502   -764.1918
            _cons |   17151.08   2376.785     7.22   0.000     12282.45    22019.7
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4590.0594
Mean Absolute Errors     |   3014.3548
Pseudo-R2                |     0.30843
----------------------------------------
```

Median monthly views

```
. reg HP23Median time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   11.65
       Model |  48268345.7     3  16089448.6           Prob > F      =  0.0000
    Residual |  38654921.8    28  1380532.92           R-squared     =  0.5553
-------------+------------------------------           Adj R-squared =  0.5077
       Total |  86923267.5    31  2803976.37           Root MSE      =    1175

------------------------------------------------------------------------------
  HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   268.5098   58.16923     4.62   0.000     149.3555    387.6641
intervention |  -3080.506   839.8499    -3.67   0.001     -4800.86   -1360.151
   postslope |  -337.3348   91.18187    -3.70   0.001    -524.1124   -150.5572
       _cons |   6097.706   596.0572    10.23   0.000     4876.738    7318.674
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1273.658
Mean Absolute Errors     |   1017.3205
Pseudo-R2                |      0.41399
----------------------------------------

. reg HP23Median time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    9.23
       Model |  43222658.1     3  14407552.7           Prob > F      =  0.0002
    Residual |  43700609.4    28  1560736.05           R-squared     =  0.4973
-------------+------------------------------           Adj R-squared =  0.4434
       Total |  86923267.5    31  2803976.37           Root MSE      =  1249.3

------------------------------------------------------------------------------
     HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   279.9441   67.75249     4.13   0.000     141.1594    418.7288
interventionMAY |  -2201.731   885.9788    -2.49   0.019    -4016.576   -386.8856
   postslopeMAY |  -406.6324   95.81649    -4.24   0.000    -602.9035   -210.3612
          _cons |     6029.1   655.1354     9.20   0.000     4687.116    7371.084
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1373.4931
Mean Absolute Errors     |   1147.5357
Pseudo-R2                |      0.32413
----------------------------------------

. reg HP23Median time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    6.36
       Model |  35235938.7     3  11745312.9           Prob > F      =  0.0020
    Residual |  51687328.8    28  1845976.03           R-squared     =  0.4054
-------------+------------------------------           Adj R-squared =  0.3417
```

```
       Total |  86923267.5    31  2803976.37           Root MSE      =  1358.7

------------------------------------------------------------------------------
       HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |      234.75   81.19588     2.89   0.007     68.42777    401.0722
interventionAPRIL |  -483.0456   959.7803    -0.50   0.619    -2449.066    1482.975
   postslopeAPRIL |   -435.625   105.4382    -4.13   0.000    -651.6054   -219.6446
            _cons |      6285.2   738.2428     8.51   0.000     4772.978    7797.422
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1508.2152
Mean Absolute Errors     |   1224.239
Pseudo-R2                |      0.20017
----------------------------------------

. reg HP23Median time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    6.84
       Model |  36767609.2     3  12255869.7           Prob > F      =  0.0013
    Residual |  50155658.3    28  1791273.51           R-squared     =  0.4230
-------------+------------------------------           Adj R-squared =  0.3612
       Total |  86923267.5    31  2803976.37           Root MSE      =  1338.4

------------------------------------------------------------------------------
       HP23Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   146.1165   88.73404     1.65   0.111    -35.64696    327.8799
interventionMARCH |   1425.536   945.4084     1.51   0.143    -511.0457    3362.117
   postslopeMARCH |  -402.6335   107.5681    -3.74   0.001    -622.9767   -182.2903
            _cons |   6757.912   755.5431     8.94   0.000     5210.252    8305.572
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1433.4557
Mean Absolute Errors     |   1161.2393
Pseudo-R2                |      0.26252
----------------------------------------
.
```

21 article set (23 set minus Ammonium Nitrate and Jihad articles)

RAW total monthly views

```
. reg HP21RawViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.38
       Model |  1.5222e+11     3  5.0742e+10           Prob > F      =  0.0000
    Residual |  6.9713e+10    28  2.4898e+09           R-squared     =  0.6859
-------------+------------------------------           Adj R-squared =  0.6522
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   49898

------------------------------------------------------------------------------
HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   14449.77   2470.296     5.85   0.000     9389.599    19509.95
intervention |    -134792   35666.25    -3.78   0.001      -207851   -61732.97
   postslope |  -22040.86   3872.258    -5.69   0.000    -29972.82    -14108.9
       _cons |   284385.7   25313.01    11.23   0.000     232534.4      336237
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   54117.269
Mean Absolute Errors     |   43122.186
Pseudo-R2                |      0.58202
----------------------------------------

. reg HP21RawViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.65
       Model |  1.3902e+11     3  4.6342e+10           Prob > F      =  0.0000
    Residual |  8.2913e+10    28  2.9612e+09           R-squared     =  0.6264
-------------+------------------------------           Adj R-squared =  0.5864
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   54417

------------------------------------------------------------------------------
   HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   14548.22   2951.169     4.93   0.000     8503.022    20593.41
interventionMAY |  -81022.03   38591.54    -2.10   0.045    -160073.2   -1970.848
   postslopeMAY |   -25025.8   4173.583    -6.00   0.000    -33574.99    -16476.6
          _cons |     283795   28536.44     9.95   0.000     225340.8    342249.3
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   60680.401
Mean Absolute Errors     |   47288.848
Pseudo-R2                |      0.47884
```

```
------------------------------------------

. reg HP21RawViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.26
       Model |  1.2601e+11     3  4.2004e+10           Prob > F      =  0.0000
    Residual |  9.5927e+10    28  3.4260e+09           R-squared     =  0.5678
-------------+------------------------------           Adj R-squared =  0.5215
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   58532

------------------------------------------------------------------------------
     HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   10747.09   3497.949     3.07   0.005     3581.866    17912.31
interventionAPRIL |   29848.38   41347.69     0.72   0.476    -54848.53    114545.3
   postslopeAPRIL |  -25421.95   4542.319    -5.60   0.000    -34726.47   -16117.43
            _cons |   305334.8   31803.78     9.60   0.000     240187.7    370481.8
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   64314.986
Mean Absolute Errors     |   43954.927
Pseudo-R2                |     0.41567
----------------------------------------

. reg HP21RawViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.20
       Model |  1.2575e+11     3  4.1916e+10           Prob > F      =  0.0000
    Residual |  9.6190e+10    28  3.4354e+09           R-squared     =  0.5666
-------------+------------------------------           Adj R-squared =  0.5202
       Total |  2.2194e+11    31  7.1593e+09           Root MSE      =   58612

------------------------------------------------------------------------------
     HP21RawViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   10464.43   3885.943     2.69   0.012     2504.441    18424.43
interventionMARCH |   52992.36    41402.4     1.28   0.211    -31816.62    137801.3
   postslopeMARCH |  -24766.06   4710.744    -5.26   0.000    -34415.58   -15116.54
            _cons |   306842.3   33087.61     9.27   0.000     239065.4    374619.1
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   63048.122
Mean Absolute Errors     |   43494.51
Pseudo-R2                |     0.43631
----------------------------------------
```

Average monthly views

```
. reg HP21AvgViews time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.38
       Model |   345191835     3   115063945           Prob > F      =  0.0000
    Residual |   158083216    28  5645829.13           R-squared     =  0.6859
-------------+------------------------------           Adj R-squared =  0.6522
       Total |   503275051    31  16234679.1           Root MSE      =  2376.1

------------------------------------------------------------------------------
HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   688.1005   117.6342     5.85   0.000     447.1377    929.0633
intervention |   -6418.64   1698.409    -3.78   0.001    -9897.673   -2939.608
   postslope |  -1049.597   184.3949    -5.69   0.000    -1427.313    -671.881
       _cons |   13542.04   1205.392    11.23   0.000      11072.9    16011.17
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   2577.0343
Mean Absolute Errors     |   2053.4405
Pseudo-R2                |     0.58203
----------------------------------------

. reg HP21AvgViews time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.65
       Model |   315262322     3   105087441           Prob > F      =  0.0000
    Residual |   188012729    28  6714740.31           R-squared     =  0.6264
-------------+------------------------------           Adj R-squared =  0.5864
       Total |   503275051    31  16234679.1           Root MSE      =  2591.3

------------------------------------------------------------------------------
   HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-----------------------------------------------------------------
           time |   692.7941    140.532     4.93   0.000     404.9274    980.6608
interventionMAY |  -3858.256   1837.694    -2.10   0.045    -7622.601   -93.91028
   postslopeMAY |  -1191.737   198.7422    -6.00   0.000    -1598.842   -784.6318
          _cons |   13513.87    1358.88     9.94   0.000     10730.34    16297.41
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   2889.5437
Mean Absolute Errors     |   2251.8857
Pseudo-R2                |     0.47886
----------------------------------------
```

```
. reg HP21AvgViews time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.26
       Model |   285751156     3  95250385.3           Prob > F      =  0.0000
    Residual |   217523895    28  7768710.54           R-squared     =  0.5678
-------------+------------------------------           Adj R-squared =  0.5215
       Total |   503275051    31  16234679.1           Root MSE      =  2787.2

------------------------------------------------------------------------------
     HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   511.7821   166.5695     3.07   0.005       170.58    852.9843
interventionAPRIL |   1421.438   1968.944     0.72   0.476     -2611.76    5454.636
   postslopeAPRIL |  -1210.601   216.3015    -5.60   0.000    -1653.674   -767.5272
            _cons |   14539.61    1514.47     9.60   0.000     11437.36    17641.86
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   3062.6289
Mean Absolute Errors     |   2093.0978
Pseudo-R2                |     0.41568
----------------------------------------

. reg HP21AvgViews time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.20
       Model |   285155599     3  95051866.3           Prob > F      =  0.0000
    Residual |   218119452    28  7789980.44           R-squared     =  0.5666
-------------+------------------------------           Adj R-squared =  0.5202
       Total |   503275051    31  16234679.1           Root MSE      =  2791.1

------------------------------------------------------------------------------
     HP21AvgViews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------------
             time |   498.3099   185.0451     2.69   0.012     119.2623    877.3575
interventionMARCH |   2523.736   1971.545     1.28   0.211    -1514.791    6562.263
   postslopeMARCH |   -1179.36   224.3213    -5.26   0.000    -1638.862    -719.859
            _cons |   14611.46   1575.602     9.27   0.000     11383.99    17838.94
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   3002.2907
Mean Absolute Errors     |   2071.122
Pseudo-R2                |     0.43632
----------------------------------------
```

Median monthly views

```
. . reg HP21Median time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    5.51
       Model |  20811047.6     3  6937015.87           Prob > F      =  0.0042
    Residual |  35281531.9    28  1260054.71           R-squared     =  0.3710
-------------+------------------------------           Adj R-squared =  0.3036
       Total |  56092579.5    31  1809438.05           Root MSE      =  1122.5

------------------------------------------------------------------------------
  HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   206.3873    55.5731     3.71   0.001     92.55092    320.2236
intervention |  -1514.518   802.3669    -1.89   0.069    -3158.092    129.0565
   postslope |  -282.8123   87.11237    -3.25   0.003    -461.2539   -104.3707
       _cons |   5062.868   569.4548     8.89   0.000     3896.392    6229.343
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1218.2317
Mean Absolute Errors     |   886.83799
Pseudo-R2                |     0.18535
----------------------------------------

. reg HP21Median time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    4.20
       Model |  17413400.8     3  5804466.93           Prob > F      =  0.0142
    Residual |  38679178.7    28  1381399.24           R-squared     =  0.3104
-------------+------------------------------           Adj R-squared =  0.2366
       Total |  56092579.5    31  1809438.05           Root MSE      =  1175.3

------------------------------------------------------------------------------
     HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   185.0985   63.74119     2.90   0.007     54.53063    315.6664
interventionMAY |  -427.3015   833.5242    -0.51   0.612    -2134.698    1280.095
   postslopeMAY |  -313.7162   90.14365    -3.48   0.002    -498.3671   -129.0653
          _cons |     5190.6   616.3479     8.42   0.000     3928.068    6453.132
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1293.5702
Mean Absolute Errors     |   965.87467
Pseudo-R2                |     0.10277
----------------------------------------

. reg HP21Median time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    4.84
       Model |  19161806.3     3  6387268.76           Prob > F      =  0.0077
    Residual |  36930773.2    28  1318956.19           R-squared     =  0.3416
-------------+------------------------------           Adj R-squared =  0.2711
```

```
       Total |  56092579.5    31  1809438.05           Root MSE      =  1148.5

------------------------------------------------------------------------------
       HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |     108.275    68.6335     1.58   0.126    -32.31436    248.8644
interventionAPRIL |    1336.64    811.286     1.65   0.111    -325.2039    2998.484
   postslopeAPRIL |  -297.9931   89.12516    -3.34   0.002    -480.5577   -115.4285
            _cons |   5625.933   624.0241     9.02   0.000     4347.678    6904.189
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1233.6708
Mean Absolute Errors     |   996.36376
Pseudo-R2                |      0.16269
----------------------------------------

. reg HP21Median time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    3.97
       Model |  16749983.3     3  5583327.77           Prob > F      =  0.0177
    Residual |  39342596.2    28  1405092.72           R-squared     =  0.2986
-------------+------------------------------           Adj R-squared =  0.2235
       Total |  56092579.5    31  1809438.05           Root MSE      =  1185.4

------------------------------------------------------------------------------
       HP21Median |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   125.9802   78.58901     1.60   0.120    -35.00206    286.9625
interventionMARCH |   1064.811   837.3191     1.27   0.214    -650.3597    2779.981
   postslopeMARCH |  -282.8265   95.26972    -2.97   0.006    -477.9776   -87.67527
            _cons |   5531.505   669.1613     8.27   0.000     4160.791     6902.22
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1264.4955
Mean Absolute Errors     |   1001.816
Pseudo-R2                |      0.12727
----------------------------------------
```

46 Article Group Cross Validation (48 minus Hamas and Fundamentalism articles)

RAW TOTAL MONTHLY VIEWS

```
reg T46NoFundaorHamasRaw time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   24.30
       Model |  3.2537e+12     3  1.0846e+12           Prob > F      =  0.0000
    Residual |  1.2496e+12    28  4.4628e+10           R-squared     =  0.7225
-------------+------------------------------           Adj R-squared =  0.6928
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.1e+05

------------------------------------------------------------------------------
T46NoFunda~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   37645.21   10458.63     3.60   0.001     16221.69    59068.73
intervention |  -683829.1   151002.1    -4.53   0.000    -993142.9   -374515.3
   postslope |  -60274.25   16394.18    -3.68   0.001    -93856.21   -26692.28
       _cons |    2261895     107169    21.11   0.000      2042370     2481421
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   226571.35
Mean Absolute Errors     |   189268.95
Pseudo-R2                |     0.63861
----------------------------------------



. reg T46NoFundaorHamasRaw time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.61
       Model |  2.9994e+12     3  9.9980e+11           Prob > F      =  0.0000
    Residual |  1.5039e+12    28  5.3711e+10           R-squared     =  0.6660
-------------+------------------------------           Adj R-squared =  0.6303
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.3e+05

------------------------------------------------------------------------------
T46NoFundaorH~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   40060.48   12568.78     3.19   0.004      14314.5    65806.46
interventionMAY |    -500903   164358.1    -3.05   0.005    -837575.4   -164230.7
   postslopeMAY |  -75642.86   17774.94    -4.26   0.000    -112053.2   -39232.55
          _cons |    2247404   121534.3    18.49   0.000      1998452     2496355
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   253115.47
Mean Absolute Errors     |   200434.46
Pseudo-R2                |     0.55147
----------------------------------------
```

```
. reg T46NoFundaorHamasRaw time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.50
       Model |  2.6623e+12     3  8.8744e+11           Prob > F      =  0.0000
    Residual |  1.8410e+12    28  6.5749e+10           R-squared     =  0.5912
-------------+------------------------------           Adj R-squared =  0.5474
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.6e+05

------------------------------------------------------------------------------
T46NoFundaorHam~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |       34170   15323.79     2.23   0.034     2780.649    65559.36
interventionAPRIL |    -203913   181135.6    -1.13   0.270    -574952.6    167126.5
   postslopeAPRIL |   -83616.5   19898.95    -4.20   0.000    -124377.7   -42855.34
            _cons |    2280783   139325.7    16.37   0.000      1995387     2566179
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   280248.05
Mean Absolute Errors     |   226329.8
Pseudo-R2                |      0.45312
----------------------------------------

. reg T46NoFundaorHamasRaw time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.67
       Model |  2.5927e+12     3  8.6423e+11           Prob > F      =  0.0000
    Residual |  1.9106e+12    28  6.8236e+10           R-squared     =  0.5757
-------------+------------------------------           Adj R-squared =  0.5303
       Total |  4.5033e+12    31  1.4527e+11           Root MSE      =  2.6e+05

------------------------------------------------------------------------------
T46NoFundaorHam~w |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   29286.11   17318.75     1.69   0.102    -6189.741    64761.95
interventionMARCH |   265.6691   184520.9     0.00   0.999    -377708.4    378239.7
   postslopeMARCH |  -84909.06    20994.7    -4.04   0.000    -127914.7   -41903.37
            _cons |    2306830   147463.8    15.64   0.000      2004765     2608896
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   281781.73
Mean Absolute Errors     |   236816.42
Pseudo-R2                |      0.44585
----------------------------------------
```

AVERAGE TOTAL MONTHLY VIEWS

```
reg T46NoFundHamasAVG time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   24.30
       Model |  1.5377e+09     3   512561551           Prob > F      =  0.0000
    Residual |   590556612    28  21091307.6           R-squared     =  0.7225
-------------+------------------------------           Adj R-squared =  0.6928
```

```
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  4592.5

------------------------------------------------------------------------------
T46NoFundH~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   818.3775    227.364     3.60   0.001     352.6434    1284.111
intervention |  -14866.27   3282.691    -4.53   0.000    -21590.56   -8141.982
   postslope |  -1310.277   356.3993    -3.68   0.001    -2040.328   -580.2265
       _cons |   49171.72   2329.788    21.11   0.000     44399.37    53944.07
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4925.5268
Mean Absolute Errors     |   4114.587
Pseudo-R2                |      0.63861
----------------------------------------

. reg T46NoFundHamasAVG time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.61
       Model |  1.4175e+09     3   472494167           Prob > F      =  0.0000
    Residual |   710758766    28  25384241.6           R-squared     =  0.6660
-------------+------------------------------           Adj R-squared =  0.6303
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5038.3

------------------------------------------------------------------------------
T46NoFundHama~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   870.8809    273.239     3.19   0.004     311.1762    1430.586
interventionMAY |  -10889.49   3573.063    -3.05   0.005    -18208.58   -3570.406
   postslopeMAY |  -1644.387   386.4183    -4.26   0.000    -2435.929   -852.8448
          _cons |    48856.7   2642.095    18.49   0.000     43444.61    54268.79
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5502.6127
Mean Absolute Errors     |   4357.318
Pseudo-R2                |      0.55146
----------------------------------------



. reg T46NoFundHamasAVG time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.50
       Model |  1.2582e+09     3   419393074           Prob > F      =  0.0000
    Residual |   870062045    28  31073644.5           R-squared     =  0.5912
-------------+------------------------------           Adj R-squared =  0.5474
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5574.4

------------------------------------------------------------------------------
T46NoFundHamasAVG |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   742.8179   333.1326     2.23   0.034     60.42674    1425.209
interventionAPRIL |  -4432.982   3937.811    -1.13   0.270    -12499.22    3633.259
   postslopeAPRIL |   -1817.73   432.5947    -4.20   0.000     -2703.86   -931.5995
```

```
       _cons |   49582.39   3028.882    16.37   0.000     43378.01    55786.77
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   6092.4748
Mean Absolute Errors     |   4920.2718
Pseudo-R2                |     0.45311
----------------------------------------


. reg T46NoFundHamasAVG time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.66
       Model |  1.2253e+09     3   408423359           Prob > F      =  0.0000
    Residual |   902971189    28    32248971           R-squared     =  0.5757
-------------+------------------------------           Adj R-squared =  0.5303
       Total |  2.1282e+09    31  68652944.1           Root MSE      =  5678.8

------------------------------------------------------------------------------
T46NoFundHamasAVG |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   636.6527    376.502     1.69   0.102    -134.5765    1407.882
interventionMARCH |    5.58394   4011.404     0.00   0.999    -8211.405    8222.573
   postslopeMARCH |  -1845.833   456.4154    -4.04   0.000    -2780.758   -910.9087
            _cons |    50148.6   3205.799    15.64   0.000     43581.82    56715.39
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   6125.807
Mean Absolute Errors     |   5148.2146
Pseudo-R2                |     0.44584
----------------------------------------
```

MEDIAN TOTAL MONTHLY VIEWS

```
reg T46NoFundHamasMED time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   12.75
       Model |  76524624.7     3  25508208.2           Prob > F      =  0.0000
    Residual |  56024918.5    28  2000889.95           R-squared     =  0.5773
-------------+------------------------------           Adj R-squared =  0.5320
       Total |   132549543    31  4275791.72           Root MSE      =  1414.5

------------------------------------------------------------------------------
T46NoFundH~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   345.4044   70.02958     4.93   0.000     201.9553    488.8535
intervention |  -3798.599    1011.09    -3.76   0.001    -5869.723   -1727.474
   postslope |   -439.108   109.7733    -4.00   0.000    -663.9684   -214.2476
       _cons |   9221.419   717.5897    12.85   0.000     7751.503    10691.33
------------------------------------------------------------------------------

. cv_regress
```

```
Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1513.5372
Mean Absolute Errors     |   1253.5942
Pseudo-R2                |      0.45732
----------------------------------------

reg T46NoFundHamasMED time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   14.61
       Model |  80886765.8     3  26962255.3           Prob > F      =  0.0000
    Residual |  51662777.4    28  1845099.19           R-squared     =  0.6102
-------------+------------------------------           Adj R-squared =  0.5685
       Total |   132549543    31  4275791.72           Root MSE      =  1358.3

------------------------------------------------------------------------------
T46NoFundHama~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   400.9118   73.66656     5.44   0.000     250.0127    551.8109
interventionMAY |  -3519.988   963.3153    -3.65   0.001     -5493.25   -1546.726
   postslopeMAY |  -529.4309   104.1803    -5.08   0.000    -742.8345   -316.0273
          _cons |   8888.375   712.3218    12.48   0.000      7429.25     10347.5
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1436.5901
Mean Absolute Errors     |   1220.6195
Pseudo-R2                |      0.50736
----------------------------------------



. reg T46NoFundHamasMED time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    9.74
       Model |  67688935.8     3  22562978.6           Prob > F      =  0.0001
    Residual |  64860607.4    28  2316450.27           R-squared     =  0.5107
-------------+------------------------------           Adj R-squared =  0.4582
       Total |   132549543    31  4275791.72           Root MSE      =    1522

------------------------------------------------------------------------------
T46NoFundHamasMED |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   381.7107   90.95623     4.20   0.000     195.3953    568.0261
interventionAPRIL |  -1824.989   1075.153    -1.70   0.101     -4027.34    377.3625
   postslopeAPRIL |  -592.4264   118.1127    -5.02   0.000    -834.3693   -350.4835
            _cons |   8997.181   826.9851    10.88   0.000     7303.179    10691.18
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1645.5625
Mean Absolute Errors     |   1376.5626
Pseudo-R2                |      0.36082
```

```
----------------------------------------


. reg T46NoFundHamasMED time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.37
       Model |    62686297     3  20895432.3           Prob > F      =  0.0004
    Residual |  69863246.2    28  2495115.94           R-squared     =  0.4729
-------------+------------------------------           Adj R-squared =  0.4165
       Total |   132549543    31  4275791.72           Root MSE      =  1579.6

------------------------------------------------------------------------------
T46NoFundHamasMED |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |       349.2    104.726     3.33   0.002     134.6784    563.7216
interventionMARCH |  -326.1035   1115.794    -0.29   0.772    -2611.703    1959.496
   postslopeMARCH |  -609.2341   126.9544    -4.80   0.000    -869.2884   -349.1797
            _cons |   9170.571   891.7101    10.28   0.000     7343.986    10997.16
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1706.3859
Mean Absolute Errors     |   1444.0275
Pseudo-R2                |      0.31438
----------------------------------------
```

44 Article Group Cross Validation (48 minus Hamas Fundamentalism Jihad Ammonium Nitrate articles)

RAW TOTAL MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundRAW time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   23.63
       Model |  2.7271e+12     3  9.0903e+11           Prob > F      =  0.0000
    Residual |  1.0771e+12    28  3.8467e+10           R-squared     =  0.7169
-------------+------------------------------           Adj R-squared =  0.6865
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.0e+05

------------------------------------------------------------------------------
T44NoHamAm~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |    30711.4   9709.946     3.16   0.004     10821.48    50601.32
intervention |  -593689.9   140192.6    -4.23   0.000    -880861.5   -306518.3
   postslope |  -52947.52   15220.61    -3.48   0.002    -84125.52   -21769.52
       _cons |    2196493   99497.33    22.08   0.000      1992682     2400304
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   210080.45
Mean Absolute Errors     |   174432.3
Pseudo-R2                |     0.63222
----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.46
       Model |  2.5266e+12     3  8.4221e+11           Prob > F      =  0.0000
    Residual |  1.2776e+12    28  4.5627e+10           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6282
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.1e+05

------------------------------------------------------------------------------
T44NoHamAmmJi~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |    32588.3   11584.35     2.81   0.009     8858.842    56317.76
interventionMAY |    -429858     151485    -2.84   0.008    -740160.9     -119555
   postslopeMAY |  -66264.44   16382.74    -4.04   0.000    -99822.96   -32705.91
          _cons |    2185232   112015.3    19.51   0.000      1955779     2414685
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   232324.24
Mean Absolute Errors     |   185440.75
Pseudo-R2                |     0.55250
----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionAPRIL postslopeAPRIL
```

```
      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.12
       Model |  2.3522e+12     3  7.8406e+11           Prob > F      =  0.0000
    Residual |  1.4520e+12    28  5.1857e+10           R-squared     =  0.6183
-------------+------------------------------           Adj R-squared =  0.5774
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.3e+05

------------------------------------------------------------------------------
T44NoHamAmmJihF~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   31293.29   13608.94     2.30   0.029     3416.651    59169.93
interventionAPRIL |    -239699   160865.2    -1.49   0.147    -569216.3    89818.35
   postslopeAPRIL |  -74286.75   17672.11    -4.20   0.000    -110486.4   -38087.08
            _cons |    2192570   123734.1    17.72   0.000      1939113     2446028
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   247561.74
Mean Absolute Errors     |   199474.92
Pseudo-R2                |     0.49301
----------------------------------------

. reg T44NoHamAmmJihFundRAW time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.67
       Model |  2.2607e+12     3  7.5358e+11           Prob > F      =  0.0000
    Residual |  1.5434e+12    28  5.5123e+10           R-squared     =  0.5943
-------------+------------------------------           Adj R-squared =  0.5508
       Total |  3.8042e+12    31  1.2272e+11           Root MSE      =  2.3e+05

------------------------------------------------------------------------------
T44NoHamAmmJihF~W |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   26300.13   15565.88     1.69   0.102    -5585.138     58185.4
interventionMARCH |   -35214.1   165845.2    -0.21   0.833    -374932.6    304504.4
   postslopeMARCH |  -76156.01   18869.78    -4.04   0.000      -114809   -37503.01
            _cons |    2219201   132538.7    16.74   0.000      1947707     2490694
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   253436.9
Mean Absolute Errors     |   211313.23
Pseudo-R2                |     0.46859
----------------------------------------
```

AVERAGE MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundAVG time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   23.63
       Model |  1.4086e+09     3   469536182           Prob > F      =  0.0000
    Residual |   556363558    28  19870127.1           R-squared     =  0.7169
-------------+------------------------------           Adj R-squared =  0.6865
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  4457.6

------------------------------------------------------------------------------
T44NoHamAm~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   697.9975   220.6837     3.16   0.004     245.9475    1150.048
intervention |  -13493.04   3186.241    -4.23   0.000    -20019.76   -6966.322
   postslope |  -1203.355   345.9278    -3.48   0.002    -1911.956   -494.7537
       _cons |   49920.14   2261.335    22.08   0.000     45288.01    54552.27
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   4774.6186
Mean Absolute Errors     |   3964.3921
Pseudo-R2                |     0.63221
----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   18.46
       Model |  1.3051e+09     3   435018695           Prob > F      =  0.0000
    Residual |   659916019    28  23568429.3           R-squared     =  0.6642
-------------+------------------------------           Adj R-squared =  0.6282
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  4854.7

------------------------------------------------------------------------------
T44NoHamAmmJi~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |     740.65   263.2848     2.81   0.009     201.3354    1279.965
interventionMAY |  -9769.475   3442.896    -2.84   0.008    -16821.93   -2717.022
   postslopeMAY |  -1506.013    372.341    -4.04   0.000    -2268.719   -743.3073
          _cons |   49664.22   2545.843    19.51   0.000      44449.3    54879.15
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5280.1831
Mean Absolute Errors     |   4214.5736
Pseudo-R2                |     0.55248
----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.12
       Model |  1.2150e+09     3   404983797           Prob > F      =  0.0000
    Residual |   750020712    28    26786454           R-squared     =  0.6183
-------------+------------------------------           Adj R-squared =  0.5774
```

```
       Total |  1.9650e+09    31  63386196.9           Root MSE      =  5175.6

------------------------------------------------------------------------------
T44NoHamAmmJihF~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   711.2107   309.2991     2.30   0.029     77.64032    1344.781
interventionAPRIL |  -5447.554   3656.086    -1.49   0.147    -12936.71    2041.599
   postslopeAPRIL |  -1688.336   401.6453    -4.20   0.000    -2511.069   -865.6025
            _cons |   49831.05   2812.184    17.72   0.000     44070.55    55591.55
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5626.4981
Mean Absolute Errors     |   4533.5744
Pseudo-R2                |     0.49299
----------------------------------------

. reg T44NoHamAmmJihFundAVG time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   13.67
       Model |  1.1677e+09     3   389240217           Prob > F      =  0.0000
    Residual |   797251454    28  28473266.2           R-squared     =  0.5943
-------------+------------------------------           Adj R-squared =  0.5508
       Total |  1.9650e+09    31  63386196.9           Root MSE      =    5336

------------------------------------------------------------------------------
T44NoHamAmmJihF~G |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   597.7121   353.7757     1.69   0.102    -126.9645    1322.389
interventionMARCH |  -799.9311   3769.269    -0.21   0.833    -8520.929    6921.066
   postslopeMARCH |  -1730.808   428.8654    -4.04   0.000    -2609.299    -852.317
            _cons |   50436.37   3012.291    16.74   0.000     44265.97    56606.77
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   5760.0175
Mean Absolute Errors     |   4802.6243
Pseudo-R2                |     0.46858
----------------------------------------
```

MEDIAN MONTHLY VIEWS

```
. reg T44NoHamAmmJihFundMED time intervention postslope

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   15.36
       Model |  57879940.6     3  19293313.5            Prob > F      =  0.0000
    Residual |  35166601.3    28  1255950.05            R-squared     =  0.6221
-------------+------------------------------           Adj R-squared =  0.5816
       Total |  93046541.9    31  3001501.35            Root MSE      =  1120.7

------------------------------------------------------------------------------
T44NoHamAm~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        time |   277.2255   55.48251     5.00   0.000     163.5747    390.8763
intervention |  -3722.905    801.059    -4.65   0.000      -5363.8    -2082.01
   postslope |  -315.3791   86.97037    -3.63   0.001    -493.5298   -137.2283
       _cons |     8611.5   568.5266    15.15   0.000     7446.926    9776.074
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1204.2531
Mean Absolute Errors     |   998.7613
Pseudo-R2                |      0.50971
----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionMAY postslopeMAY

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   20.34
       Model |  63777614.9     3    21259205           Prob > F      =  0.0000
    Residual |    29268927    28  1045318.82            R-squared     =  0.6854
-------------+------------------------------           Adj R-squared =  0.6517
       Total |  93046541.9    31  3001501.35            Root MSE      =  1022.4

------------------------------------------------------------------------------
T44NoHamAmmJi~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
----------------+-------------------------------------------------------------
           time |   336.9338   55.44788     6.08   0.000      223.354    450.5137
interventionMAY |  -3670.966   725.0751    -5.06   0.000    -5156.215   -2185.717
   postslopeMAY |  -404.5265   78.41514    -5.16   0.000    -565.1526   -243.9003
          _cons |    8253.25   536.1555    15.39   0.000     7154.985    9351.515
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1079.2187
Mean Absolute Errors     |   935.57829
Pseudo-R2                |      0.60239
----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionAPRIL postslopeAPRIL

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =   10.04
       Model |  48209351.7     3  16069783.9            Prob > F      =  0.0001
    Residual |  44837190.2    28  1601328.22            R-squared     =  0.5181
-------------+------------------------------           Adj R-squared =  0.4665
       Total |  93046541.9    31  3001501.35            Root MSE      =  1265.4
```

```
------------------------------------------------------------------------------
T44NoHamAmmJihF~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   309.8536   75.62426     4.10   0.000     154.9443    464.7629
interventionAPRIL |   -1926.92   893.9207    -2.16   0.040    -3758.034   -95.80659
   postslopeAPRIL |  -468.0104   98.20312    -4.77   0.000    -669.1704   -266.8505
            _cons |   8406.705    687.585    12.23   0.000     6998.251    9815.159
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1387.7669
Mean Absolute Errors     |   1091.2975
Pseudo-R2                |      0.35549
----------------------------------------

. reg T44NoHamAmmJihFundMED time interventionMARCH postslopeMARCH

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    8.08
       Model |  43173895.2     3  14391298.4           Prob > F      =  0.0005
    Residual |  49872646.7    28  1781165.95           R-squared     =  0.4640
-------------+------------------------------           Adj R-squared =  0.4066
       Total |  93046541.9    31  3001501.35           Root MSE      =  1334.6

------------------------------------------------------------------------------
T44NoHamAmmJihF~D |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
             time |   283.9912   88.48334     3.21   0.003     102.7413    465.2411
interventionMARCH |  -632.2332   942.7373    -0.67   0.508    -2563.343    1298.877
   postslopeMARCH |  -489.7167   107.2642    -4.57   0.000    -709.4374    -269.996
            _cons |   8544.637   753.4085    11.34   0.000      7001.35    10087.92
------------------------------------------------------------------------------

. cv_regress


Leave-One-Out Cross-Validation Results
----------------------------------------
         Method          |    Value
-------------------------+--------------
Root Mean Squared Errors |   1445.5777
Mean Absolute Errors     |   1183.4531
Pseudo-R2                |      0.30089
----------------------------------------
```

List of 48 Terrorism Articles with Privacy Sensitivity Scores (from Survey)

Topic Keyword	Wikipedia Articles	Government Trouble	Browser Delete	Privacy Sensitive	Avoidance	Average
Al Qaeda	http://en.wikipedia.org/wiki/Al-Qaeda	2.20	2.11	2.21	2.84	2.34
terrorism	http://en.wikipedia.org/wiki/terrorism	2.19	2.05	2.16	2.79	2.30
terror	http://en.wikipedia.org/wiki/terror	1.98	1.96	2.01	2.64	2.15
attack	http://en.wikipedia.org/wiki/attack	1.92	1.91	1.92	2.56	2.08
Iraq	http://en.wikipedia.org/wiki/iraq	1.60	1.74	1.76	2.25	1.84
Afghanistan	http://en.wikipedia.org/wiki/afghanistan	1.61	1.71	1.75	2.23	1.83
Iran	http://en.wikipedia.org/wiki/iran	1.62	1.73	1.78	2.25	1.85
Pakistan	http://en.wikipedia.org/wiki/pakistan	1.59	1.71	1.75	2.22	1.82
agro	http://en.wikipedia.org/wiki/agro	1.51	1.80	1.76	2.29	1.84
Environmental terrorism	http://en.wikipedia.org/wiki/Environmental_terrorism	2.20	2.20	2.24	2.92	2.39
Eco terrorism	http://en.wikipedia.org/wiki/Eco-terrorism	2.22	2.20	2.22	2.92	2.39
Conventional weapon	http://en.wikipedia.org/wiki/Conventional_weapon	2.03	2.16	2.07	2.81	2.27
Weapons grade	http://en.wikipedia.org/wiki/Weapons-grade	2.18	2.22	2.17	2.99	2.39
dirty bomb	http://en.wikipedia.org/wiki/Dirty_bomb	2.72	2.55	2.50	3.45	2.81
Nuclear Enrichment	http://en.wikipedia.org/wiki/Nuclear_enrichment	2.22	2.21	2.21	2.92	2.39
Nuclear	http://en.wikipedia.org/wiki/nuclear	1.84	1.97	1.91	2.55	2.07
Chemical weapon	http://en.wikipedia.org/wiki/Chemical_weapon	2.43	2.36	2.39	3.16	2.59
Biological weapon	http://en.wikipedia.org/wiki/Biological_weapon	2.44	2.39	2.39	3.18	2.60
Ammonium nitrate	http://en.wikipedia.org/wiki/Ammonium_nitrate	2.49	2.44	2.26	3.24	2.61
Improvised explosive device	http://en.wikipedia.org/wiki/Improvised_explosive_device	2.82	2.64	2.53	3.46	2.86
Abu Sayyaf	http://en.wikipedia.org/wiki/Abu_Sayyaf	2.02	1.96	1.99	2.57	2.14
Hamas	http://en.wikipedia.org/wiki/hamas	1.90	1.93	1.97	2.49	2.07
FARC	http://en.wikipedia.org/wiki/FARC	1.83	1.88	1.90	2.46	2.02
Irish Republican Army	http://en.wikipedia.org/wiki/Irish_Republican_Army	1.62	1.77	1.83	2.24	1.87
Euskadi ta Askatasuna	http://en.wikipedia.org/wEuskadi_ta_Askatasuna	1.86	1.88	1.88	2.43	2.01
Hezbollah	http://en.wikipedia.org/wiki/hezbollah	1.86	1.90	1.96	2.46	2.05
Tamil Tigers	http://en.wikipedia.org/wiki/Tamil_Tigers	1.76	1.86	1.87	2.39	1.97
PLO	http://en.wikipedia.org/wiki/Palestine_Liberation_Organization	1.77	1.87	1.91	2.42	1.99
Palestine Liberation Front	http://en.wikipedia.org/wiki/Palestine_Liberation_Front	1.81	1.89	1.95	2.47	2.03
Car bomb	http://en.wikipedia.org/wiki/Car_bomb	2.72	2.61	2.50	3.40	2.81
jihad	http://en.wikipedia.org/wiki/jihad	2.15	2.19	2.17	2.89	2.35
Taliban	http://en.wikipedia.org/wiki/taliban	2.06	2.03	2.10	2.70	2.22
Suicide bomber	http://en.wikipedia.org/wiki/Suicide_bomber	2.25	2.31	2.24	2.97	2.44
Suicide attack	http://en.wikipedia.org/wiki/Suicide_attack	2.30	2.36	2.29	3.04	2.50
AL Qaeda in the Arabian Peninsula	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Arabian_Peninsula	2.01	1.98	2.06	2.63	2.17
Al Qaeda in the Islamic Maghreb	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Islamic_Maghreb	2.05	1.98	2.06	2.60	2.17
Tehrik-i-Taliban Pakistan	http://en.wikipedia.org/wiki/Tehrik-i-Taliban_Pakistan	1.96	1.96	1.97	2.59	2.12
Yemen	http://en.wikipedia.org/wiki/yemen	1.60	1.72	1.74	2.18	1.81
Pirates	http://en.wikipedia.org/wiki/pirates	1.44	1.67	1.67	2.10	1.72
Extremism	http://en.wikipedia.org/wiki/extremism	1.64	1.90	1.86	2.40	1.95
Somalia	http://en.wikipedia.org/wiki/somalia	1.50	1.68	1.67	2.12	1.74
Nigeria	http://en.wikipedia.org/wiki/nigeria	1.48	1.66	1.64	2.07	1.71
Political radicalism	http://en.wikipedia.org/wiki/Political_radicalism	1.75	1.91	1.97	2.48	2.03
Al-Shabaab	http://en.wikipedia.org/wiki/Al-Shabaab	1.84	1.89	1.89	2.48	2.03
nationalism	http://en.wikipedia.org/wiki/nationalism	1.48	1.71	1.73	2.20	1.78
Recruitment	http://en.wikipedia.org/wiki/recruitment	1.74	1.90	1.87	2.54	2.01
Fundamentalism	http://en.wikipedia.org/wiki/fundamentalism	1.60	1.79	1.80	2.32	1.88
Islamist	http://en.wikipedia.org/wiki/islamist	1.79	1.89	1.93	2.45	2.45
					MEDIAN	2.08

- 2.08 = Median of the Average Privacy Ratings for the 48 Articles
- 23 Most Privacy Sensitive Article Set includes all articles in the 48 Terrorism Group with combined privacy rating average above the 2.08 median.

Al Qaeda	http://en.wikipedia.org/wiki/Al-Qaeda	2.20	2.11	2.21	2.84	2.34
terrorism	http://en.wikipedia.org/wiki/terrorism	2.19	2.05	2.16	2.79	2.30
terror	http://en.wikipedia.org/wiki/terror	1.98	1.96	2.01	2.64	2.15
attack	http://en.wikipedia.org/wiki/attack	1.92	1.91	1.92	2.56	2.08
Environmental terrorism	http://en.wikipedia.org/wiki/Environmental_terrorism	2.20	2.20	2.24	2.92	2.39
Eco terrorism	http://en.wikipedia.org/wiki/Eco-terrorism	2.22	2.20	2.22	2.92	2.39
Conventional weapon	http://en.wikipedia.org/wiki/Conventional_weapon	2.03	2.16	2.07	2.81	2.27
Weapons grade	http://en.wikipedia.org/wiki/Weapons-grade	2.18	2.22	2.17	2.99	2.39
dirty bomb	http://en.wikipedia.org/wiki/Dirty_bomb	2.72	2.55	2.50	3.45	2.81
Nuclear Enrichment	http://en.wikipedia.org/wiki/Nuclear_enrichment	2.22	2.21	2.21	2.92	2.39
Chemical weapon	http://en.wikipedia.org/wiki/Chemical_weapon	2.43	2.36	2.39	3.16	2.59
Biological weapon	http://en.wikipedia.org/wiki/Biological_weapon	2.44	2.39	2.39	3.18	2.60
Ammonium nitrate	http://en.wikipedia.org/wiki/Ammonium_nitrate	2.49	2.44	2.26	3.24	2.61
Improvised explosive device	http://en.wikipedia.org/wiki/Improvised_explosive_device	2.82	2.64	2.53	3.46	2.86
Abu Sayyaf	http://en.wikipedia.org/wiki/Abu_Sayyaf	2.02	1.96	1.99	2.57	2.14
Car bomb	http://en.wikipedia.org/wiki/Car_bomb	2.72	2.61	2.50	3.40	2.81
jihad	http://en.wikipedia.org/wiki/jihad	2.15	2.19	2.17	2.89	2.35
Taliban	http://en.wikipedia.org/wiki/taliban	2.06	2.03	2.10	2.70	2.22
Suicide bomber	http://en.wikipedia.org/wiki/Suicide_bomber	2.25	2.31	2.24	2.97	2.44
Suicide attack	http://en.wikipedia.org/wiki/Suicide_attack	2.30	2.36	2.29	3.04	2.50
AL Qaeda in the Arabian Peninsula	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Arabian_Penin	2.01	1.98	2.06	2.63	2.17
Al Qaeda in the Islamic Maghreb	http://en.wikipedia.org/wiki/Al-Qaeda_in_the_Islamic_Maghreb	2.05	1.98	2.06	2.60	2.17
Tehrik-i-Taliban Pakistan	http://en.wikipedia.org/wiki/Tehrik-i-Taliban_Pakistan	1.96	1.96	1.97	2.59	2.12

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 3

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 15-cv-00662-TSE

SECOND DECLARATION OF MICHELLE PAULSON

I, Michelle Paulson, declare:

1. I am a resident of San Francisco, California, over the age of eighteen. I have personal knowledge of the facts stated in this declaration, and, if called to testify, I could and would testify competently thereto. I am providing this declaration in my capacity as a former employee of and current consultant to the Wikimedia Foundation, Inc. ("Wikimedia").

2. As explained in my prior declaration in this litigation and the First Declaration of Emily Temple-Wood, Wikimedia has an active and close relationship with its community—which consists of the individuals who read or contribute to the body of knowledge comprising the twelve Wikimedia Projects. Although I focus here on Wikimedia's readers, it is critical to understand that readers and contributors are not static or mutually exclusive categories of users within the Wikimedia community. A Wikimedia user can be a reader at one moment in time, contribute an edit to a Wikipedia page at another, and then return to reading pages. In other words, our users play a variety of different roles as they engage with the Projects.

3. Below, I discuss several additional examples of Wikimedia programs and policies that help maintain the strong relationship between Wikimedia and its readers. These examples also illustrate the ways in which Wikimedia's ability to fulfill its mission depends on its ability to reach and engage readers. Indeed, readers are the beneficiaries of the vast body of human knowledge and free educational content that Wikimedia strives to develop and distribute online. Just as libraries work to enrich reader experiences and to expand their base of patrons in fulfillment of their mission, Wikimedia does the same. The ultimate aim of Wikimedia's work is to attract more readers as well as more contributors to the Projects. Readers are especially important to Wikimedia because, unlike patrons at a library, they also routinely generate content for the Projects—which, in turn, helps to attract even more readers.

4. First, unlike many other websites and platforms, Wikimedia offers all of its readers the opportunity to participate directly in the development of community standards and policies. Readers of the Projects are able to participate in discussions and decision-making that help determine how the community governs itself, and thus the shape of the Projects going forward. These include discussions about matters ranging from the Wikimedia movement's strategy and values to copyright issues and search query structure. *See, e.g.*, First Decl. of Michelle Paulson ¶¶ 9-11.[1]

5. Second, Wikimedia is engaged in a continuous process of consultation with its readers, soliciting their input on a variety of different topics related to the content and operation of the Projects.

[1] *See also, e.g.*, https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2017; https://meta.wikimedia.org/wiki/Values/2016_discussion; https://meta.wikimedia.org/wiki/Copyright_strategy; https://www.mediawiki.org/wiki/User_Interaction_Consultation.

a. Wikimedia assigns various teams to conduct research and surveys to learn about readers' experiences of the Wikimedia Projects. Some of these research projects are specifically designed to help increase and diversify readership. For instance, Wikimedia has a "New Readers" team, which is dedicated to learning about the preferences of readers in underserved countries with respect to Wikimedia content and their reading experiences.

b. Other research projects are designed to gather information to enrich current readers' experiences—for example, a short survey that asks readers, "Why are you reading this article?"

c. Wikimedia's research activities involve significant expenditures, including costs related to survey design, consultants, and foreign travel to interview Wikimedia readers, contributors, and others who have not yet engaged with the Wikimedia projects.

6. Third, Wikimedia devotes considerable resources to maintaining and increasing the engagement and involvement of its readers. Wikimedia has an entire Department dedicated to the engagement of its community, including readers. The Community Engagement Department is responsible for a variety of functions, including facilitating communications between Wikimedia and its community, and assisting Wikimedia and its community with studying and evaluating projects and program activities to help make them more effective. In addition, Wikimedia's Galleries, Libraries, Archives, and Museums program establishes partnerships between Wikimedia and these institutions to help incorporate their content into the Projects. By incorporating images and information from these institutions into Wikimedia pages, Wikimedia seeks to enrich the experiences of current readers and to attract new ones.

7. Fourth, Wikimedia devotes considerable resources to expanding its readership, including by making the content of the Projects available to greater numbers of readers. For example:

a. In the interests of disseminating educational content to a broader readership base, Wikimedia has dedicated substantial resources to the translation of its web pages, to help ensure that robust content is available in multiple languages.

b. Wikimedia's Communications Department runs Wikimedia-awareness campaigns around the world to increase readership. Recent campaigns involved outreach in Mexico and Nigeria. These campaigns include the creation and local dissemination of videos and other advertisements, such as the following video, Superdotada, available at https://www.youtube.com/watch?v=fsCxG0MTrJs.

c. The Wikipedia Zero program, discussed at greater length in my earlier declaration, has been primarily focused on providing free access to Wikipedia readers via mobile devices.

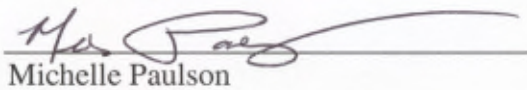
d. Wikimedia engineering teams have dedicated time and resources to ensuring that the Projects are available to readers using older model mobile phones or web browsers that are popular with people in emerging markets. This helps ensure that readers in all locations, even with readers with older devices or slow or limited internet access, can use the Wikimedia Projects.

8. Finally, Wikimedia has advocated for and will continue to zealously advocate for all of its readers' and contributors' free expression and privacy rights, including the rights afforded by the First and Fourth Amendments. As discussed at greater length in my earlier declaration, Wikimedia takes numerous, costly steps to protect the confidentiality of its communications,

including its communications with readers. Through policies, public statements, and guidelines, Wikimedia assures its community—including its readers—that it will reject third-party requests for non-public user information unless it is legally required to disclose that information. In keeping with these assurances, Wikimedia resists third-party demands for information—including information about readers' page views—that are overly broad, unclear, or irrelevant, and it notifies users individually of information requests when legally permitted and it is able to do so.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on March 8, 2018 in Los Angeles, California.



Michelle Paulson

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 4

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 1:15-cv-00662-TSE

SECOND DECLARATION OF TILMAN BAYER

I, Tilman Bayer, declare:

1. I am a resident of San Francisco, California, over the age of eighteen. I have personal knowledge of the facts stated in this declaration, and, if called to testify, I could and would testify competently thereto. I am providing this declaration in my capacity as an employee of the Wikimedia Foundation, Inc. ("Wikimedia").

2. I am a Senior Analyst in Wikimedia's Product Analytics team, and have been a full-time employee of the organization since 2012. My responsibilities include the reporting of pageview statistics and other key web traffic metrics to Wikimedia's executives and board. I hold degrees in mathematics from the University of Cambridge (Certificate of Advanced Study in Mathematics) and the University of Bonn (diploma, equivalent to a Master's degree in the US).

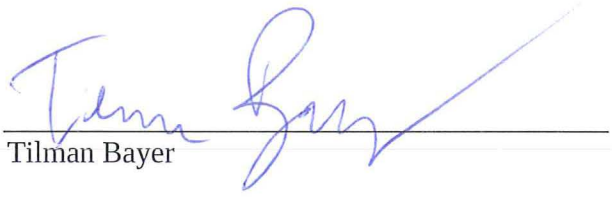
3. As explained in my previous Declaration in support of Wikimedia's Opposition to Defendants' Motion for Summary Judgment, Wikimedia provides technical infrastructure to twelve free-knowledge "Projects" on the Internet. (Pl.'s Ex. 5 ¶¶ 3-4.) Wikimedia users edit specific Wikimedia Project pages via "Edit" pages, which contain a text box with editable versions of the online article. As of March 2018, there have been approximately 3.4 billion edits over the

lifespan of the Wikimedia Projects. (*Id.* ¶¶ 12-13.) Edits are transmitted from users to Wikimedia servers as HTTP and HTTPS communications, part of the previously described *Category 1 – Wikimedia communications with its community members*. (*Id.* ¶ 26.)

4. When a user edits Wikimedia Project pages, the final changes to the master versions of Wikimedia's content databases are stored on Wikimedia's servers located in the United States, not Wikimedia's caching servers located abroad. (*See id.* ¶ 4 (noting country locations of Wikimedia's servers).) For this reason, all edits to Wikimedia Project pages from users located in foreign countries are routed to Wikimedia's servers in the United States.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on March 8, 2019 in San Francisco, California.



Tilman Bayer

Wikimedia Foundation v. NSA
No. 15-cv-0062-TSE (D. Md.)

Plaintiff's Exhibit 5

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY / CENTRAL SECURITY SERVICE, *et al.*, *Defendants*.	No. 15-cv-00662-TSE

SECOND DECLARATION OF EMILY TEMPLE-WOOD

I, Emily Temple-Wood, declare:

1. I am a U.S. citizen and resident of Downers Grove, Illinois, over the age of eighteen. I have personal knowledge of the facts stated in this declaration, and, if called to testify, I could and would testify competently thereto. I am providing this declaration in my capacity as a Wikimedia Foundation, Inc. ("Wikimedia") community member. I am not an employee or contractor of Wikimedia.

I. Background

2. My earlier declaration in this litigation attested to the relationship between Wikimedia and its users; the importance of Wikimedia's non-U.S. readers and contributors to U.S. users like myself; the importance of anonymity to Wikimedia users; and obstacles that other users and I face in bringing suit.

3. As detailed in that declaration, since April 2007, I have been both a reader and an editor of the English Wikipedia. I have created nearly 400 articles on Wikipedia and have edited thousands more.

II. Contributing to Wikimedia While Outside of the United States

4. Throughout my time as a member of the Wikimedia community, I have traveled to numerous countries outside of the United States, including for the purpose of attending Wikimedia conferences abroad.

5. As a U.S. citizen traveling abroad, I have continued to serve as an editor of the English Wikipedia. While outside of the United States, I have edited hundreds of Wikipedia articles, including:

- Cherney incision, https://en.wikipedia.org/wiki/Cherney_incision (edited from Mexico);
- Urethral diverticulum, https://en.wikipedia.org/wiki/Urethral_diverticulum (edited from Mexico);
- X-linked recessive hypoparathyroidism, https://en.wikipedia.org/wiki/X-linked_recessive_hypoparathyroidism (edited from Mexico);
- Lucie Randoin, https://en.wikipedia.org/wiki/Lucie_Randoin (edited from Mexico);
- Birt-Hogg-Dubé syndrome, https://en.wikipedia.org/wiki/Birt%E2%80%93Hogg%E2%80%93Dub%C3%A9_syndrome (edited from Hong Kong);
- Wikipedia talk:WikiProject Women scientists: Difference between revisions, https://en.wikipedia.org/w/index.php?title=Wikipedia_talk:WikiProject_Women_scientists&diff=prev&oldid=567822352 (edited from Hong Kong); and
- Endometrial cancer, https://en.wikipedia.org/wiki/Endometrial_cancer (edited from the United Kingdom).

6. While traveling abroad, I have edited Wikipedia articles from my pseudonymous account.

7. While traveling abroad, I have removed sensitive items within Wikipedia pages from public view in my capacity as an English Wikipedia administrator.

8. I have also witnessed dozens of individuals whom I believe to be U.S. citizens or permanent residents edit Wikipedia articles while they, too, are outside of the United States.

9. I intend to continue traveling abroad on at least an annual basis. During these travels, I also intend to continue editing Wikipedia articles.

10. As a Wikipedia user, I am concerned about government surveillance, including Upstream surveillance of my communications while abroad with Wikimedia servers in the United States. *See also* First Declaration of Emily Temple-Wood ¶ 20.

III. Upstream Surveillance Impairs My and Other U.S. Users' Ability to Exchange Information with Wikimedia's Foreign Readers and Editors

11. As detailed in my earlier declaration, my interest in contributing to Wikipedia is based in part on my ability to reach an international audience, and I also read and benefit from the contributions of non-U.S. users on a wide array of topics.

12. NSA surveillance, including Upstream surveillance, threatens the anonymity and privacy of Wikimedia community members, and that threat discourages individuals from engaging with the Wikimedia Projects.

13. The loss of foreign readers and editors is a direct detriment to me and to the Wikimedia community at large. When foreign users are less willing to read Wikimedia Project pages, I and other community members lose potential readers of our original content and other contributions to the Projects. When foreign users are less willing to contribute to Wikimedia Project pages, I and other community members are unable to read and benefit from their contributions.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on March 8, 2019, in Downers Grove, Illinois.



Emily Temple-Wood

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 16

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	)	No. 1:15-cv-0662 (TSE)

(THIRD) DECLARATION OF DR. HENNING SCHULZRINNE

Dr. Henning Schulzrinne, for his (third) declaration pursuant to 28 U.S.C. § 1746, deposes and says as follows:

INTRODUCTION AND SUMMARY

1. I am the Julian Clarence Levi Professor of Computer Science at Columbia University in New York, New York. I previously submitted two declarations in this case, dated November 12, 2018 and February 15, 2019. I submit this third declaration at the request of the United States Department of Justice to address the conclusions reached by Mr. Scott Bradner in his reply declaration filed on March 8, 2019, including Mr. Bradner's assessment of conclusions reached in my February 15 declaration. My background and qualifications in the fields of computer science, electrical engineering, and digital communications technology; the sources of information I considered in arriving at the conclusions stated in this case (apart from those cited herein); and my compensation for my services in this matter, are all stated in my prior declarations.

2. For the reasons I detail herein, it remains my conclusion that the hypothesis advanced in this case by plaintiff Wikimedia Foundation ("Wikimedia"), that the National Security Agency ("NSA"), in the course of conducting Upstream collection, must as a matter of

technological necessity be intercepting, copying, and reviewing at least some of Wikimedia's electronic communications that traverse the Internet, is incorrect. Based on what is publicly known about the NSA's Upstream collection technique, the NSA in theory could be conducting this activity, at least as Wikimedia conceives of it, in a number of technically feasible, readily implemented ways that could avoid NSA interaction with Wikimedia's online communications. Nothing stated in Mr. Bradner's reply declaration, or in Wikimedia's sur-reply, alters that conclusion.

3. I also adhere to the conclusions reached in my second declaration, (Second) Declaration of Dr. Henning Schulzrinne (hereinafter, "Second Decl."). In his first declaration Mr. Bradner concluded (i) that the NSA "most likely" uses his copy-all-then-scan configuration to conduct Upstream collection, and (ii) that it is "implausible" that the NSA employs a filter-then-copy-and-scan approach, using whitelisting and blacklisting techniques, such as I described in my first declaration; and (iii) that that even if the NSA uses one or more of the techniques I describe, it is still "virtually certain" that the NSA copies and scans at least some of Wikimedia's communications. *See generally* Declaration of Scott Bradner (hereinafter, "Bradner Decl."). In my second declaration, I explained that none of these conclusions has a foundation in Internet technology or engineering, and instead are based principally on assumptions Mr. Bradner makes about the NSA's practices and priorities, its resources and capabilities, and its Upstream surveillance targets, matters about which Mr. Bradner has no specialized knowledge or information. These conclusions also remain unaltered by Mr. Bradner's reply declaration, or Wikimedia's sur-reply.

4. In his second declaration, Mr. Bradner takes a somewhat different approach. The central thesis of Mr. Bradner's second declaration is that the use of traffic-mirroring techniques (that is, whitelisting and blacklisting) to filter the communications traversing a monitored link before communications of interest are copied and then scanned for selectors, would "conflict[ ]" with "the government's own descriptions" of the Upstream program. Reply Declaration of Scott Bradner (hereinafter "Bradner Reply Decl.") ¶¶ 6, 12, 30. By the "government's own

descriptions," Mr. Bradner means, first, a single statement contained in an 80-page opinion issued by the Foreign Intelligence Surveillance Court ("FISC") in October 2011, regarding the NSA's acquisition of wholly domestic "about" communications, Bradner Reply Decl., App'x P, and second, a single remark in the nearly 200-page report of the Privacy and Civil Liberties Oversight Board on the NSA's Section 702 surveillance program (hereinafter, "PCLOB Section 702 Report"), regarding the NSA's goal of "comprehensively" acquiring communications to or from its targets, Bradner Reply Decl., App'x F.

5. As I already discussed in my prior declaration, the use of a filter-then-copy-and-scan approach to Upstream collection would be entirely consistent with both of these statements. Second Decl. ¶¶ 56-58. The conflict perceived by Mr. Bradner with the FISC's statement arises not from any technical grounds but his own speculative interpretation of that statement, and his assumption that a remark made by the FISC nearly eight years ago still reflects the technical realities of Upstream collection today. The supposed conflict that Mr. Bradner attempts to seize on with the PCLOB's remark ignores the difference between a statement of aspirations on a printed page and the real-world challenges of designing, constructing, deploying, maintaining, and paying for the collection systems required to implement the kind of Upstream collection process that Mr. Bradner envisions. Moreover, he misconstrues the PCLOB's description of the NSA's prior collection of "about" communications as an aim of the program, whereas the PCLOB clearly described "about" collection as "byproduct" of the NSA's efforts to acquire communications sent to or from its Upstream surveillance targets.

6. Mr. Bradner also opines that a filter-then-copy-and-scan approach "conflicts with other technical and practical necessities of conducting" Upstream surveillance. Bradner Reply Decl. ¶¶ 7, 61-112. As I also discuss below, these so-called "technical and practical necessities" are based on the unsupported conclusions Mr. Bradner draws from the FISC and PCLOB statements, as well as many of the same non-technical, speculative assumptions made in his previous declaration about the NSA's surveillance practices and priorities, resources and

capabilities, and its targets, about which Mr. Bradner has no actual knowledge. As such, they still provide no basis in Internet technology or engineering for the conclusions he reaches.

7. Mr. Bradner remarks that I do not offer evidence that the NSA is actually using whitelisting and/or blacklisting techniques in the course of Upstream surveillance that would avoid interaction with Wikimedia's communications. Bradner Reply Decl. ¶¶ 6, 57-58. The same can be said of Mr. Bradner, of course, that he has offered no evidence—only speculation—that the NSA conducts Upstream surveillance in the manner that he describes. As I have explained before, I did not attempt in my previous declarations to reach conclusions about how the NSA actually conducts Upstream surveillance, because I was not asked by the Department of Justice to opine on that question, because the question implicates operational details of Upstream surveillance that remain classified, and because it would have required that I engage in the same sort of speculation as Mr. Bradner, concerning the NSA's actual surveillance practices, capabilities, and targets, about which neither I nor Mr. Bradner has any specialized knowledge or information. *See* Second Decl. ¶ 3. I do not engage in such speculation now, either.

8. As was the case with my first two declarations, in reaching the conclusions stated herein I have not considered, nor have I been provided with, any classified or other non-public information concerning Upstream surveillance.

MR. BRADNER'S CONCLUSION THAT WHITELISTING AND BLACKLISTING WOULD "CONFLICT" WITH THE FISC'S 2011 STATEMENT REGARDING THE COLLECTION OF WHOLLY DOMESTIC "ABOUT" COMMUNICATIONS LACKS A FOUNDATION IN INTERNET TECHNOLOGY AND ENGINEERING

9. As Mr. Bradner observes, in an October 2011 opinion concerning the legal implications of the NSA's collection of so-called multi-communication transactions ("MCTs"), the FISC stated that:

> Indeed, the government readily concedes that NSA will acquire a wholly domestic "about" communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA or is routed through a foreign server. See June 1 Submission at 29.

Bradner Reply Decl., App'x P at 45.

10. In his previous declaration, Mr. Bradner cited this statement in support of a very different conclusion than the one he offers now. Earlier, he cited the FISC's statement as evidence only that the NSA, at least at some monitored links, was not using "an IP filter to eliminate wholly domestic [communications] before" copying and scanning. Bradner Decl. ¶ 294; *see* Second Decl. ¶ 56. Mr. Bradner did not cite this conclusion as a ground for viewing his copy-all-then-scan configuration as "most likely," or a filter-then-copy-and-scan approach as "implausible." *See* Bradner Decl. ¶¶ 282-89, 366-67. Nevertheless Wikimedia, in its summary judgment opposition brief, argued that that the acquisition of some wholly domestic communications, even at a so-called "international Internet link," is inconsistent with the use of the traffic-mirroring techniques I have described, even though Mr. Bradner had not made any such assertion in his first declaration filed at that time. *See* Second Decl. ¶ 56. I explained, therefore, why use of whitelisting and blacklisting techniques would be consistent with the acquisition of at least some wholly domestic "about" communications, Second Decl. ¶¶ 56-58, a conclusion that Mr. Bradner does not dispute, *see* Bradner Reply Decl. ¶¶ 36, 111.

11. Only now, in his second declaration, does Mr. Bradner make an argument that implementing a filter-then-copy-and-scan approach would be inconsistent with the collection of wholly domestic about communications, as referred to in the FISC's opinion. Bradner Reply Decl. ¶¶ 32-45. Indeed, he is emphatic that the FISC's statement would have to be "false" for the NSA to employ a filter-then-copy-and-scan approach. *Id.* ¶ 154(1). This is the case, according to Mr. Bradner, because the use of whitelists or blacklists to filter out certain communications traversing a monitored link could filter out at least some wholly domestic about communications, containing targeted selectors, that the NSA might otherwise acquire. *Id*. ¶ 44. That outcome, Mr. Bradner states, "would not be consistent with the FISC's statement that *all* 'about' communications 'will' be acquired." *Id*. ¶ 42 (emphasis mine).

12. But the FISC did not state that "all" wholly domestic about communications crossing a monitored international Internet link will be acquired by the NSA. It said that "a wholly domestic 'about" communication" will be acquired if it crosses such a link, or is routed through a

foreign server. Mr. Bradner arrives at the conclusion that whitelisting and blacklisting would be inconsistent with the FISC's statement by imputing to the FISC a term, "all," that the FISC did not use. His contention that a filter-then-copy-and-scan approach would not be consistent with the acquisition of wholly domestic about communications is not based on a disagreement over Internet technology or engineering, but on his current reading of the FISC's 2011 statement, one that he did not propose in his first declaration, *see* Bradner Decl. ¶ 296.

13. There are numerous reasons why the statement in the FISC's 2011 opinion does not support Mr. Bradner's conclusion about how Upstream collection must operate today. First, as noted, the FISC simply does not use the term "all" in its statement regarding the collection of wholly domestic about communications.

14. Second, Mr. Bradner does not take into account the context in which the FISC made this remark. The paragraph in which the statement appears concerned the NSA's inability to prevent the acquisition of certain wholly domestic communications, and the Government's suggestion that these acquisitions resulted from a "failure" of the NSA's "technical means." Bradner Reply Decl., App'x P at [45]. The FISC was unwilling to accept that explanation, finding no reason to conclude that the collection of wholly domestic communications was attributable to malfunctions or failures in the NSA's collection equipment. This was the point at which the FISC remarked that the Government had conceded that a wholly domestic about communication would be collected if it crossed a monitored international Internet link, or was routed through a foreign server. In other words, it appears that when the FISC made this remark it was explaining that acquisitions of wholly domestic about communications would occur as the result of technical limitations in the equipment's normal operation (rather than as the result of a malfunction), and was not making a statement about the scope or completeness of those acquisitions.

15. Third, and telling, in an earlier portion of the FISC's opinion where it was discussing the same phenomenon—the NSA's inability to prevent the acquisition of at least some wholly domestic communications—the FISC stated that due to this limitation on the NSA's abilities, the "NSA *may* acquire wholly domestic communications." Bradner Decl., App'x P at 35 n.34

(emphasis mine). Mr. Bradner's conclusion, that according to the FISC the NSA obtained all wholly domestic about communications crossing a monitored link, is thus further undermined by the FISC's clear statement describing the NSA's acquisition of wholly domestic communications as a possibility ("may acquire"), rather than a certainty in all cases.[1]

16. Mr. Bradner also suggests that the FISC must have been "as precise as it possibly could be" when describing the acquisition of wholly domestic about communications, based on the "multiple hearings" it held and the "multiple submissions" it received from the Government before issuing its October 2011 opinion. Bradner Reply Decl. ¶ 38. Of course, this assumption is not a technical basis on which to reach conclusions about how Upstream surveillance is conducted. And I do not understand how Mr. Bradner could know exactly what was said by the Government in those submissions, what was addressed during those hearings, or in what depth, all of which I am advised by the Department of Justice remain classified, in whole or in substantial part. In the end, however, if the FISC meant to say that the NSA would acquire *all* wholly domestic communications crossing a (hypothetically) monitored international Internet link, then the most precise way of expressing that thought would have been to use the word "all." The FISC did not do so.

17. On this subject I observe finally that the FISC's statement cited by Mr. Bradner was made in October 2011, based on a June 1, 2011 submission by the Government, Bradner Reply Decl., App'x P, at 45, nearly eight years ago. Even if the FISC meant that in June 2011 the NSA would acquire all wholly domestic "about" communications crossing an international Internet

[1] Although in this earlier passage the FISC referred to the collection of "wholly domestic communications," rather than "wholly domestic about communications," as in the later passage, there is no reason to believe that it meant to suggest that the collection of wholly domestic communications in general (including both "to/from" and "about" communications) was only a possibility, but at the same time that collection of all wholly domestic "about" communications was a certainty. At least as described in the PCLOB Section 702 Report, the underlying causes for acquisition of wholly domestic communications are similar for both cases: once a communication (for example, an email), makes it past any IP filters, the scanning mechanism would be just as likely to pick up wholly domestic "to" or "from" communications that contain targeted selectors as "about" communications containing targeted selectors. *See* PCLOB Section 702 Report at 10, 37, 84-86, 123, 124, 144-45.

link theoretically monitored by the NSA, and it could be inferred, therefore, that the NSA was not then employing whitelist or blacklist filters, it cannot be taken for granted that the situation has remained the same since then. The growth of Internet traffic, even since 2011, has been enormous. Between 2012 and 2016 alone, the volume of traffic (as measured in used international bandwidth by Telegeography, Inc.) more than quadrupled from 100 terabits per second to over 400 terabits per second[2]. *See* https://blog.telegeography.com/shaping-the-global-wholesale-bandwidth-market (figure 1). Given the growth in international Internet traffic since 2011, the possibility cannot be ignored that the NSA at some point might have adopted a form of whitelisting or blacklisting at monitored links to reduce the technical and logistical challenges and costs of processing large volumes of traffic, as discussed in my second declaration, Second Decl. ¶¶ 20-21. Neither Mr. Bradner nor I can know which is the case, of course, but it cannot be taken for granted that the assumption made by Mr. Bradner is the correct one.

18. As I have noted, Mr. Bradner does not disagree with my conclusion that the use of whitelisting and blacklisting would be consistent with the acquisition of at least some wholly domestic "about" communications at international Internet links that hypothetically might be monitored by the NSA. *See* Bradner Reply Decl. ¶¶ 36, 111. Therefore, because it is speculative to suggest (i) that the FISC meant to say that the NSA would collect *all* wholly domestic "about" communications if they crossed a monitored international Internet link in 2011, and (ii) even if that was the FISC's meaning in 2011, that the NSA would continue to collect all wholly domestic about communications at such links years afterward, Mr. Bradner lacks a technical basis on which to conclude that whitelisting and blacklisting would be inconsistent with the FISC's statement.

19. In sum, I find no reason, and certainly none based in Internet technology or engineering, to conclude that whitelisting or blacklisting at international Internet links that the NSA may in theory be monitoring would be inconsistent with the 2011 FISC statement cited by Mr. Bradner.

[2] This traffic volume includes all international traffic, not just traffic entering and exiting the United States. The growth rates, however, have been similar across all major geographies.

THE GOAL OF "COMPREHENSIVELY" ACQUIRING TARGETS' COMMUNICATIONS IS NOT A TECHNICAL BASIS ON WHICH TO CONCLUDE THAT THE NSA DOES NOT EMPLOY WHITELISTING OR BLACKLISTING TECHNIQUES

20. Mr. Bradner next opines that a filter-then-copy-and-scan approach to Upstream collection process would be "incompatible" with the goal of "comprehensively acquir[ing] communications that are sent to or from [the NSA's] targets," as stated on pages 10 and 123 of the PCLOB Section 702 Report. Bradner Reply Decl. ¶¶ 46-54.[3] He goes so far as to say that the PCLOB's statement would have to be "false" for the NSA to conduct Upstream collection using the traffic-mirroring techniques I have described. *ID.* ¶ 154(2). As I have explained previously, Mr. Bradner offers no technical basis for concluding as a matter of concrete reality that the NSA "must be" copying and reviewing all communications crossing a monitored link, Bradner Reply Decl. ¶ 29, simply because the NSA might wish in the abstract to obtain as many of its targets' communications as it can. Second Decl. ¶¶ 71-75.

21. Specifically, I explained that one cannot assume away, based on a stated aim of comprehensiveness, the many technical, logistical, and financial hurdles, and competing mission priorities, that would stand in the way of designing, constructing, deploying, and maintaining the kind of collection systems envisioned by Mr. Bradner. Second Decl. ¶ 73. (See also paragraphs 20-21 of my second declaration, in which I describe some of the technical and logistical challenges of implementing Mr. Bradner's copy-all-then-scan approach.) Therefore, one cannot draw meaningful conclusions about the technical details of the NSA's Upstream collection systems with nothing more to go on than an abstract goal of comprehensiveness. Second Decl. ¶ 74.

22. Mr. Bradner does not dispute these practical realities. He himself states that "the NSA must operate in the real world and deal with the technical and operational limitations

[3] I note that in contrast the PCLOB describes the collection of "about" communications as a "byproduct" of the NSA's efforts to collect communications sent to or from its targets. PCLOB Section 702 Report at 10, 123. The PCLOB's description of "about" collection as a "byproduct" rather than an objective of Upstream collection tends to rebut Mr. Bradner's interpretation of the FISC's October 2011 statement to mean that the NSA must have configured Upstream collection in such a way as to acquire "all" wholly domestic about communications on a monitored link. *See also* paragraphs 47-49, below.

inherent in the Internet." Bradner Reply Decl. ¶ 10. Yet he continues to disregard these same realities when he attempts to draw technical conclusions about how Upstream surveillance "must be" conducted from a single statement by the PCLOB attributing a goal of "comprehensiveness" to the NSA.

23. Mr. Bradner remarks that the use of whitelists and blacklists to reduce the technical and logistical difficulties and costs of Upstream collection is "incompatible" with the goal of completeness, Bradner Reply Decl. ¶ 51, but they are entirely compatible with reducing the formidable technical and logistical burdens and costs of processing large volumes of communications traffic, and Mr. Bradner has no way of knowing, and certainly no technical way of determining, whether such "real world" constraints have convinced or compelled the NSA to compromise the goal of completeness. It is, at bottom, the hard realities of achieving Mr. Bradner's vision of Upstream collection that may have proven incompatible with that goal, just as they are incompatible with a bare assumption that the NSA has succeeded in implementing that vision simply because it wants to.

24. Mr. Bradner opines that the single term "comprehensively" provides an "appropriate basis on which to explain the technological implementation" of Upstream collection, because "the PCLOB used the term 'comprehensively' to explain the need for the NSA's specific technological implementation" of the program. Bradner Reply Decl. ¶ 48. But the PCLOB's Section 702 Report, being an unclassified public report about a highly classified foreign-intelligence gathering activity, contains no specific technical detail about how Upstream collection is conducted. It certainly includes no technical detail at the level required to support the conclusions reached about Upstream collection by Mr. Bradner, except as a matter of speculation and conjecture.

25. "If wishes were horses," as they say, then one could simply assume that the NSA had the technical, logistical, and financial wherewithal, consistent with all its other mission requirements, to deploy an Upstream collection architecture capable of comprehensively acquiring every single one of its targets' online communications, just because it would like to do

so. And perhaps it has succeeded in doing so. I (like Mr. Bradner) do not know. But the assumption has no basis in the "real world" of Internet technology and engineering, Bradner Reply Decl. ¶ 10, in which the NSA must operate.

MR. BRADNER'S OPINION THAT WHITELISTING AND/OR BLACKLISTING WOULD "CONFLICT" WITH "TECHNICAL AND PRACTICAL NECESSITIES" OF UPSTREAM COLLECTION RESTS ON SPECULATION RATHER THAN A BASIS IN INTERNET TECHNOLOGY AND ENGINEERING

26. In addition to the inferences he draws from his interpretation of the FISC's 2011 statement, and the PCLOB's use of the term "comprehensive" to describe Upstream's objective, Mr. Bradner also attempts to reach conclusions based on what he calls "technical and practical necessities" that make clear, in his view, that the NSA is copy and scanning at least some of Wikimedia's communications. As I discuss below, these "technical and practical necessities" turn out again to be assumptions by Mr. Bradner about the NSA's surveillance practices and priorities, resources and capabilities, and the nature and behavior of its Upstream surveillance targets, mixed with the unsupported inferences he draws from the FISC and PCLOB statements discussed above. They supply no basis in Internet technology and engineering for concluding that the NSA "most likely" uses a copy-all-then-scan configuration to conduct Upstream surveillance, or that a filter-then-copy-and-scan approach is "implausible."

Whitelisting IP Addresses of Interest

27. Developing and maintaining whitelists: In Mr. Bradner's view whitelisting would be unworkable for purposes of Upstream collection because it would require "knowing in advance all of the IP addresses that might be used by each of the NSA's targets," Bradner Reply Decl. ¶ 68, and maintaining "comprehensive" information on where they will be, what sites they communicate with, and the protocols they use, Bradner Reply Decl. ¶ 69. *See also id.* ¶ 88. These conclusions lack a non-speculative technical basis.

28. As a technical matter, the NSA would not need to know all of its targets' IP addresses or gather comprehensive information on their whereabouts, the sites they visit, or the types of online communications in which they engage, before it could whitelist the IP addresses that the agency already knows about. As pointed out in the Government's reply brief, the NSA

could acquire information about its targets' IP addresses from communications acquired under the Section 702 PRISM program (*see* PCLOB Section 702 Report (Bradner Decl., App'x F) at 7, 33-34), Executive Order 12,333, prior Upstream acquisitions, information obtained from other U.S. intelligence agencies, and other sources. Reply Brief in Support of Defendants' Motion for Summary Judgment at 12-13. Mr. Bradner does not dispute this. Presumably communications acquired from these sources would also provide insight into particular websites of interest that targets visit, and their preferred modes of communication, thus allowing for further enhancement of the NSA's whitelist(s).

29. Mr. Bradner suggests no technological reason why it would not be possible for the NSA to conduct Upstream surveillance in this fashion. Doing so would not mean that the NSA "is only interested in the people and processes it already knows about and that it has decided to actively ignore everything else." Bradner Reply Decl. ¶ 89. Nothing about the use of whitelists would prevent the NSA from learning about new potential targets of interest, or new information about the communications of its existing targets, either from ongoing Upstream collection or other intelligence sources (in Mr. Bradner's words, discovering which streetlights to look under for your keys, *id.* ¶ 70), and then updating its whitelists accordingly.

30. There would be no need whatsoever for the NSA to know in advance all the information Mr. Bradner refers to unless one works backward from the conclusion, as Mr. Bradner does, that the NSA's acquisition of its targets' communications must be comprehensive, *see* Bradner Reply Decl. ¶¶ 69, 88, based on the PCLOB's passing remark about the NSA's goals. As I have discussed, the PCLOB's report does not provide a technical justification for concluding that the NSA's acquisition of its targets' communications is comprehensive in fact, or drawing inferences based on that premise about how Upstream surveillance is conducted.

31. Mr. Bradner observes that whitelists would have to be updated as targets were added or removed, or changed their locations or modes of communications. Bradner Reply Decl. ¶ 85. This is true, of course, but how often that would have to occur Mr. Bradner does not say,

because he cannot know, and he certainly offers no reason why updating whitelists as needed would be beyond the NSA's capabilities.

32. Target numerosity: Mr. Bradner next repeats his argument that the NSA's targets are too numerous to make the development and maintenance of whitelists of their IP addresses practical. Bradner Reply Decl. ¶¶ 75-76; *see* Bradner Decl. ¶ 366(d). I have already explained that in so arguing Mr. Bradner is making speculative assumptions about the number, nature, and communications habits of the NSA's Upstream targets, about which Mr. Bradner has no information. First Decl. ¶¶ 45-48. He does not maintain otherwise. Instead, he offers a hypothetical in place of facts: if the NSA had 1,000 Upstream targets in 2011, when according to the FISC it acquired 26 million communications using its Upstream collection technique, then on average each target would have had to engage in at least 26,000 online communications that year. Bradner Reply Decl. ¶ 76. But if we used another randomly chosen number of Upstream targets, say 5,000, then the number of communications each target would have had to send or receive in 2011 drops to 5,200, or approximately 14 per day. That is a trivial number considering (i) that "communications" can include not only such media as email but also all the individual HTTPS (or HTTP) requests and responses made during a single visit to a website, and (ii) that the NSA's Upstream targets could include organizations as well as individuals, organizations employing dozens or hundreds of persons capable of generating thousands of communications per day. Mr. Bradner has given no factual, technical basis for concluding that the number of the NSA's Upstream targets would make whitelisting unworkable.

33. Target mobility: Mr. Bradner also repeats his arguments that NSA use of whitelists is not "remotely possible" because its targets may move around, and as a result their IP addresses could change; because targets may use intermediary communications services, such as virtual private networks (VPNs, which remove the targets' IP address from packet headers during certain legs of their journeys across the Internet); and because targets may use multiple Internet service providers (ISPs), that assign different IP addresses to their subscribers. Bradner Reply Decl. ¶¶ 77-78, 82. In my view, Mr. Bradner overstates the matter. To take just one example, if the

NSA were seeking to track a targeted individual's email, determining the source and destination IP addresses to whitelist could be automated readily. An exhaustive list of SMTP destination IP addresses for any email address can be looked up online via DNS (MX records), and the list of source IP addresses can be derived from the DNS SPF entries. These IP addresses would not be affected by user mobility, and would only depend on the sender or receiver email address, not the user's current location.

34. But again, most fundamentally, Mr. Bradner is making assumptions about the nature, mobility, and communications practices of the NSA's targets. Only the NSA knows the extent to which its targets' IP addresses change due to their mobility, the extent to which they use VPNs, or multiple ISPs, to communicate, and therefore whether whitelisting would be impractical for purposes of meeting its intelligence-collection needs. Whitelisting could not be considered technologically impossible, however, unless one started from the premise that the NSA must be "comprehensively" acquiring every single one of its targets' communications without fail, and worked backward from there. As I have explained, Mr. Bradner has offered no justification, certainly no technological justification, for adopting that premise.

35. Further, as I note in my second declaration, to the extent a target moves from place to place within a given geographic area, the NSA could whitelist a set of IP addresses, rather than just a single address, associated with the geographic region where the target is believed to be located. Second Decl. ¶ 47. Mr. Bradner does not dispute this, but points out that using a range of IP addresses could increase the chances that communications to or from Wikimedia could be copied and scanned. Bradner Reply Decl. ¶¶ 80, 82. This is true in the abstract, but Mr. Bradner cites no information about the extent to which the NSA might find it necessary to whitelist ranges of IP addresses—how many, how large, in what geographic areas—in order to reliably monitor its targets' communications. As a result he can only speculate whether whitelisting ranges of IP addresses would result in the NSA copying and scanning Wikimedia communications.

36. Combined whitelisting and blacklisting to selectively acquire web communications: I previously observed that if it wished the NSA could, as a technological matter, simultaneously whitelist the IP addresses of particular websites, webmail services, and/or chatrooms of interest while blacklisting all other HTTP and HTTPS traffic, and thus obtain access to web communications of interest without necessarily copying and scanning Wikimedia's. Second Decl. ¶¶ 35, 36(b), 37. In response, Mr. Bradner remarks that "many" websites and other web-based services are now making use of content distribution networks ("CDNs"), which can have different and changing IP addresses around the world. Bradner Reply Decl. ¶ 84. Yet again, the extent to which websites, webmail services, and chatrooms of interest to the NSA (if any) are using CDNs for their communications, and the extent to which their use of CDNs would make it difficult or unworkable for the NSA to use whitelists to track its targets' communications, are matters known to the NSA, but about which Mr. Bradner has no information, and can only speculate.

37. Whitelisting by protocol: Mr. Bradner also points out (i) that if the NSA used protocol-based whitelisting (as opposed to whitelisting by IP addresses), it could miss communications that it ideally it might want to review, Bradner Reply Decl. ¶ 72, and (ii) that the NSA could not in fact be using whitelists to exclude all communications using web protocols (HTTP and HTTPS) while at the same time collecting at least some web communications, *id.* ¶ 73. These are points that Mr. Bradner has raised before, Bradner Decl. ¶ 366(f), (g), and that I addressed in my second declaration, Second Decl. ¶¶ 34-35, 36(b), 37. For the reasons already explained in my second declaration, these observations by Mr. Bradner supply no basis in Internet technology or engineering for concluding that whitelisting by the NSA would be implausible.

38. I do not mean to suggest that developing and maintaining whitelists for purposes of NSA Upstream collection would necessarily be "easy," or that "the NSA could get by with a very simple set of whitelist rules." *See* Bradner Reply Decl. ¶¶ 70, 85, 88. But neither I nor Mr. Bradner can know how difficult or easy it would be without far more detailed information about the number, nature, and communications habits of the NSA's Upstream targets. I do not

presume, as does Mr. Bradner, that the NSA would be incapable of developing and maintaining sufficiently reliable whitelists to meet its intelligence needs. Mr. Bradner has presented no basis in Internet technology or engineering for so assuming.

Blacklisting by IP Address or Protocol

39. Mr. Bradner states that "[t]here are multiple reasons" why the use of blacklists to filter out communications before they are copied or scanned in the Upstream process "would be incompatible with the public descriptions of the NSA's upstream collection program," Bradner Reply Decl. ¶ 92, but then proceeds to offer none. Instead he contends (i) that blacklisting Wikimedia IP addresses so as to prevent copying and scanning Wikimedia communications is "improbable," and (ii) that blacklisting Wikimedia IP addresses would not "guarantee" that Wikimedia communications are not copied and scanned during the Upstream collection process. *Id.* ¶¶ 93-101. Both are points that Mr. Bradner has raised before, that I have already addressed, and that still do not constitute non-speculative, technical grounds for deeming it improbable or implausible that the NSA might blacklist Wikimedia communications.

40. Blacklisting high-volume websites, including Wikimedia's: Following the observations in my first declaration that the NSA could blacklist Wikimedia IP addresses to prevent communications to and from Wikimedia from being copied and scanned, First Decl. ¶¶ 78-87, Mr. Bradner responded that he found it "basically inconceivable" and "totally unbelievable" that the NSA would sift through millions of websites to decide which to monitor and which not. Bradner Decl. ¶ 367(a). I then explained, in my second declaration, that I had no such extreme measure in mind, but rather, the trivial task of creating a blacklist of numerous high-volume popular websites of potentially low interest to the NSA, such as Wikimedia's, to eliminate unwanted volumes of communications that would otherwise have to be processed. Second Decl. ¶¶ 39-41.

41. Now Mr. Bradner takes the opposite tack, arguing that it is "very unlikely" that the NSA would decide to "specifically blacklist Wikimedia communications to reduce the load" on its collection apparatus, in light of the relatively small percentage of inter-regional Internet capacity

that Wikimedia communications traffic represents. Bradner Reply Decl. ¶ 96. I note first that a more enlightening statistic would be the percentage of the NSA's processing capacity that Wikimedia traffic represents, but Mr. Bradner has no way of knowing that information any more than he could know, rather than speculate, whether the NSA might find reason to blacklist Wikimedia communications or not. But it is also important to note that Mr. Bradner has once again missed the point of my observations. I did not suggest, in my second declaration, a scenario in which the NSA specifically singles out Wikimedia sites for blacklisting. I pointed out that Wikimedia's websites would naturally fall on any blacklist of high-volume, popular websites, of perhaps low interest to the NSA, that the NSA might assemble in order reduce (potentially by as much as 90 percent or more) the technological, logistical, and financial burdens of processing large volumes of unwanted web traffic. Second Decl. ¶ 41. Whether or not the NSA actually does so I do not know, but Mr. Bradner offers no technological basis for dismissing the possibility as improbable.

42. Hypothetical copying and scanning of blacklisted Wikimedia communications: Second, Mr. Bradner again draws attention to three hypothetical scenarios in which communications to or from Wikimedia would be copied and scanned during the Upstream collection process, even if the NSA had blacklisted communications containing Wikimedia IP addresses. Bradner Reply Decl. ¶¶ 97-101; *see* Bradner Decl. ¶ 367(a). These involved transmission of a Wikimedia communication within a multi-communication transaction (MCT) across an international Internet link, transmission of an email to Wikimedia from abroad using a U.S.-based email service, and visits to Wikimedia websites from abroad using a U.S.-based VPN service. I explained in my prior declaration that in each of these scenarios four (or in the third scenario, five) conditions would have to be met before a communication to or from Wikimedia would be copied or scanned, in each case rendering that possibility a matter of speculation. Second Decl. ¶¶ 77-85.

43. Regarding the initial three (or four) conditions required before each scenario could come to pass, Mr. Bradner summarily asserts that these conditions "would likely be frequently

met," yet offers no supporting data or explanation to demonstrate why that would be so. Instead, he focuses his attention principally on the last condition that would have to be met in each scenario before a Wikimedia communication could be copied or scanned—that the communication not be blacklisted for other reasons. Mr. Bradner dismisses as "far-fetched" the possibility that communications of the kind he posited in his three scenarios would be blacklisted, based on conjecture about the value that the NSA might attach to them. Bradner Reply Decl. ¶¶ 99-101. But he does not address the reasons given in my prior declaration to expect that the communications he describes likely would be encrypted, and perhaps blacklisted by the NSA, therefore, if it lacked the ability to decipher them. Second Decl. ¶¶ 80, 82, 84. In any event, if the initial three (or four) conditions in each scenario are not met, whether or not the final condition is met would be of no consequence. And as noted above, Mr. Bradner gives no reason to expect that the occurrence of the initial three (or four) conditions in each scenario would be anything but speculative.

Additional Points Concerning Whitelisting and Blacklisting

44. Mr. Bradner completes his discussion of whitelisting, blacklisting, and the filter-then-copy-and-scan approach generally with additional points, most already made in his prior declaration that I will now address.

45. "Blind spots": Mr. Bradner returns to his earlier remark that if the NSA blacklisted particular types of communications by port or protocol number, then doing so would leave "blind spots" in its collection that "[s]ophisticated" targets could "easily probe" to discover and evade collection of their communications. Bradner Reply Decl. ¶¶ 103-07; *see* Bradner Decl. ¶ 366(b), (e). On this point I observed previously that Mr. Bradner had not explained what targets could "probe," or how, to discover these so-called blind spots, the level of sophistication required, or on what basis he presumed that the NSA's Upstream targets possess the needed sophistication. Second Decl. ¶ 32. At bottom, I further observed, Mr. Bradner could only speculate whether the creation of "blind spots" would be of such genuine concern to the NSA as to dissuade it from utilizing whitelisting or blacklisting techniques. *Id.* ¶ 33.

46. In his second declaration, Mr. Bradner adds to his earlier conjecture by supposing that unspecified foreign intelligence services that may or may not be Upstream targets could test a protocol (type of communication) they suspect the NSA is not monitoring by communicating "actionable" intelligence ("such as the identity of a foreign agent") over that protocol and then observing whether any responsive action is taken. Bradner Reply Decl. ¶ 103. He vaguely suggests that whitelists or blacklists shared with telecommunication service providers could be at unspecified risk of "hacking," without considering additional security measures that could be taken to mitigate those risks. *Id*. ¶ 105. Typically, however, carriers use segregated networks to manage their routers and switches and encrypt network management information. There have been no indications that I am aware of that such carrier management networks have been breached. In short, neither of these suggestions constitutes a non-speculative basis in Internet technology or engineering to support Mr. Bradner's conclusions. And he overlooks the fundamental point I made previously, that only the NSA knows whether it considers these to be genuine risks that it would be unprepared to take in order to implement a filter-then-copy-and-scan approach to Upstream collection. Second Decl. ¶ 33.

47. <u>"About" communications</u>: In his second declaration Mr. Bradner suggests for the first time that whitelisting or blacklisting would be "entirely inconsistent" with the NSA's now-discontinued acquisition of "about" communications. Bradner Reply Decl. ¶ 108; *see also id.* ¶ 78. (As I noted in my second declaration, this was an argument advanced by Wikimedia in its legal brief, but not by Mr. Bradner in his first declaration. Second Decl. ¶ 49.) The position now taken by Mr. Bradner is based on a complete misunderstanding of "about" collection as described in the PCLOB Section 702 report.

48. Mr. Bradner acknowledges, as I have explained, that even if the NSA were to employ whitelists and/or blacklists to implement a filter-then-copy-and-scan approach to Upstream collection, it would still be possible to acquire at least some "about" communications, that is, communications neither to nor from a target but which refer to a selector (e.g., an email address) associated with the target. *See* Second Decl. ¶ 51; Bradner Reply Decl. ¶ 111. Mr.

Bradner's point is that the use of whitelists or blacklists would be incompatible with the *comprehensive* collection of "about" communications. For example, he states (twice) that developing and maintaining whitelists for acquisition of "about" communications "would be impossible to do . . . for a program *meant to capture* the communications of unknown non-targets about targets." Bradner Reply Decl. ¶¶ 69-70 (emphasis mine). In the same vein, he later reasons that "to set up [a] whitelist filter the NSA would have to know in advance which non-targets' IP addresses to whitelist … *in order to find the 'about' communications.*" Bradner Reply Decl. ¶ 110 (emphasis mine). He ends with the conclusion that, using a filter-then-copy-and-scan model, "the only way the NSA *could reliably capture* about communications would be to whitelist all non-wholly domestic communications." Bradner Reply Decl. ¶ 112 (emphasis mine).

49. The premise of Mr. Bradner's argument, however, is incorrect. At least as publicly described by the PCLOB Section 702 Report, in the very passage reproduced by Mr. Bradner in his declaration, the NSA's collection of "about" communications was

> an inevitable *byproduct* of the government's efforts to comprehensively acquire communications that are sent to or from its targets.

Bradner Reply Decl. ¶ 48 (quoting PCLOB Section 702 Report at 10) (emphasis mine). That is to say, while the goal of Upstream collection as described by the PCLOB is to comprehensively acquire communications sent "to or from" the NSA's foreign-intelligence targets, the program was not likewise "meant to capture," Bradner Reply Decl. ¶¶ 69-70, "reliably" find, *id.* ¶¶ 110, 112, or otherwise "comprehensively acquire" "about" communications. *See also* PCLOB Report at 84-86, 123, 124, 144, 145. The incidental collection of only some "about" communications, even if the NSA were using a filter-then-copy-and-scan configuration as described in my second declaration, would be entirely consistent with the PCLOB's description of "about" collection as a byproduct rather than an objective of Upstream collection. For example, if a whitelist included a specific email server, the NSA could scan for targeted email addresses in the whitelisted communications, and would, as described in the PCLOB Section 702 Report, occasionally acquire "about" communications that referred to the targeted email address in addition to

communications to or from the email address. Thus, the "about" communication would be captured because of the way the scanning is implemented. And being incidental, "about" collection would not have required advance knowledge of all non-targets' IP addresses.

50. Collection of web communications, including encrypted HTTPS communications: Mr. Bradner again addresses the reasons why he believes the NSA is collecting web (i.e., HTTP and/or HTTPS) communications, Bradner Reply Decl. ¶¶ 130-36, 154(3)(c)(i); *see* Bradner Decl. ¶¶ 314-15m 366(f), focusing in particular on his earlier opinions (i) that blacklisting the HTTP and HTTPS protocols to prevent the copying and scanning of all web communications "would leave a very large hole in the NSA's coverage." Bradner Reply Decl. ¶ 135; *see* Bradner Decl. ¶ 366(f), (g), and (ii) that encrypted communications collected by the NSA (if any) include, specifically, HTTPS communications, Bradner Reply Decl. ¶¶ 137-39; *see* Bradner Decl. ¶¶ 325, 366(g).[4] Mr. Bradner largely ignores the observation, in my second declaration, that the NSA could use a combined whitelisting/blacklisting technique to block NSA access to all HTTP and HTTPS communications except those to or from IP addresses included on a whitelist containing the addresses of websites, chatrooms, and/or webmail services of intelligence interest. In this way the NSA could obtain access to HTTP and HTTPS communications of interest, while excluding all others, including, hypothetically, Wikimedia's. Second Decl. ¶¶ 35, 36(b), 37.

51. Mr. Bradner opines that this technique "would also leave very large holes in the NSA's coverage," Bradner Reply Decl. ¶ 136; *see also id.* ¶ 139, but whether the "holes" would be so unacceptably large as to motivate the NSA to copy and scan all HTTP and HTTPS communications (including, therefore, Wikimedia's) is a matter implicating the NSA's surveillance priorities, resources, and capabilities, about which Mr. Bradner has no information and can only

[4] I note in passing that Mr. Bradner expresses skepticism at my suggestion that the NSA could collect encrypted communications under its Section 702 authority using its PRISM acquisition method, as opposed to Upstream collection, because providers assisting in PRISM collection "will frequently have direct access to the user's unencrypted communications." Bradner Reply Decl. ¶ 139. But numerous online file-storage and webmail services now store their users' files and/or messages in encrypted formats, either by default or upon request.

speculate. Mr. Bradner also suggests that combined whitelisting and blacklisting of HTTP and HTTPS communications "would also be contrary to the aim of the 'about' collection program." Bradner Reply Decl. ¶ 136. As I discussed, however, in paragraph 49, above, there is no evidence that the NSA ever engaged in an "'about' collection program." The PCLOB Section 702 Report does not state that "about" collection was ever an "aim" of Upstream collection, merely the "byproduct" of efforts to collect communications to and from designated targets.

52. Relative complexity of the copy-all-then-scan and filter-then-copy-and-scan approaches: Mr. Bradner describes his copy-all-then-scan approach as a "simpl[er], mo[re] reliable, and easi[er] to operate architecture" than a filter-then-copy-and-scan configuration, Bradner Reply Decl. ¶ 116, because in his view, "the need to constantly reconfigure the [provider's router or switch] with updated blacklists and whitelists would create the risk of misconfiguration or overloading." *Id.* ¶ 118; *see also id*. ¶ 124.

53. To begin, it is not unusual for a provider to reconfigure its routers and switches on a routine basis to meet the evolving data transmission needs of a dynamic commercial customer base. Whether it would require more frequent reconfiguration to update an NSA whitelist or blacklist is a matter of conjecture that would depend on how frequently the NSA requested such updates, based on how often it makes additions to or deletions from its target list, and how often its targets change their modes of communication. It is also a matter of speculation, as I have observed previously, whether the NSA's targets (or, more precisely, their associated IP addresses), are so numerous that loading a whitelist would run the risk of overloading a router's or switch's processing capacity. Second Decl. ¶ 25. These are all matters about which Mr. Bradner apparently has no information. Moreover, carriers usually implement change-management protocols in which new router configurations are tested in laboratory settings before they are loaded, in order to mitigate any risk of misconfiguration or overloading.

54. The question Mr. Bradner also leaves unanswered is whether the risks he cites would outweigh the daunting technical and logistical difficulties and financial burdens that would complicate implementation of his "simpl[er]" and "easi[er]" approach, as discussed in paragraphs

20-21 of my second declaration. Mr. Bradner does not take issue with my description of these hurdles (with the exception of the marginal observation that an opto-electronic device is not an "esoteric" piece of equipment). Bradner Reply Decl. ¶ 119. When considered from a broader perspective, the asserted simplicity of Mr. Bradner's copy-all-then-scan configuration as compared to a filter-then-copy-and-scan approach becomes less apparent, to say the least, and which way the scales tip in the NSA's eyes is a matter about which Mr. Bradner, once again, can only speculate.

55. Sharing sensitive information with a provider: Mr. Bradner also repeats a point made in his prior declaration that, "in [his] opinion," the NSA would not want to implement a filter-then-copy-and-scan approach because it would require sharing with provider personnel the sensitive information about NSA targets and the scope of its surveillance that would be contained in whitelists and blacklists. Bradner Reply Decl. ¶ 126. This is a point I have already addressed. Second Decl. ¶ 18. Mr. Bradner acknowledges, as I observed previously, that the NSA shares information about its targets with Internet service providers in order to conduct PRISM collection, but suggests that in his estimation the information that would have to be shared with a provider in a whitelist or blacklist is even more sensitive. Bradner Reply Decl. ¶ 126. Whether or not that is so (a matter about which Mr. Bradner has no apparent expertise), it remains the case that the extent to which the NSA would be willing (or find it necessary) to share classified information with an assisting provider in order to conduct Upstream surveillance is a matter about which Mr. Bradner has no specialized knowledge or information.

56. Proposed "channel mirroring": Mr. Bradner also makes a passing reference to a configuration not previously proposed by him, in which a provider would configure its router or switch to copy all communications just on particular channels (circuits), designated by the NSA, that cross a monitored link. Bradner Reply Decl. ¶ 121. He describes that configuration, too, as very simple, static, and involving fewer disclosures of sensitive information to non-government personnel, *id.*, but it is not so simple a picture as Mr. Bradner paints.

57. In brief, due to carrier practices including link aggregation, among others, it is unclear how the NSA could know which optical channels would be carrying the communications of potential targets, at least without a deep understanding of the assisting carrier's routing architecture and configuration. In addition, because of a phenomenon referred to as traffic failover, traffic can move from one logical link to another, so this configuration is unlikely to be stable and unvarying. As a result, the NSA would have to convey to the carrier which IP addresses it would like to monitor, so that the carrier can map these addresses onto logical links and their optical channels. Thus, the NSA would have to convey just as much information to the carrier as with a filter-then-copy-and-scan configuration. Furthermore, the mirroring capability of at least some common routers and switches is limited to a small number of interfaces. Thus, only a small fraction of the router input or output ports could be monitored at any time if all the traffic, unfiltered, is to be copied to these ports.

58. U.K. Section 8(4) collection: Mr. Bradner returns to this subject to prove a point that I did not take great issue with in my second declaration: that the brief filed by the U.K. government in the European Court of Human Rights describing its "Section 8(4)" collection program includes references to "intercept[ing]"--which could be taken to mean copying—"the entire contents of a [circuit]" before the communications stream is filtered, and the remainder then scanned for targets' communications. Bradner Reply Decl. ¶ 140-48; *see* Second Decl. ¶¶ 61-63.[5] Notably, Mr. Bradner does not take issue with the conclusions I previously drew from the U.K. governments' filings: (i) that the U.K. government documents describe a process of filtering to winnow out communications deemed to lack significant intelligence value before communications are scanned for targets' selectors, Second Decl. ¶ 62, and (ii) that even if all communications on a circuit were copied first, the copying and initial filtering could be conducted

[5] Oddly, Mr. Bradner appears to find fault that I based my discussion of the Section 8(4) collection program on one of the same documents he relied on his declaration (the U.K. government brief), and that I failed to cite a document, *not* appended to his first declaration, that he himself did not cite previously, and that he himself now cites only for the first time in his reply declaration. *See* Bradner Reply Decl. ¶ 144; Second Decl. ¶ 61 (citing Bradner Decl., App'x EE).

by the service provider, so that only those communications meeting the filter criteria, rather than all communications, would pass into the government's (whether the GCHQ's, or hypothetically, the NSA's) control. Second Decl. ¶ 64.[6]

THE SCENARIOS ENVISIONED BY MR. BRADNER IN WHICH AT LEAST SOME WIKIMEDIA COMMUNICATIONS WOULD BE COPIED AND SCANNED, EVEN IF THE NSA EMPLOYED TRAFFIC-MIRRORING TECHNIQUES, ARE SPECULATIVE AND CONJECTURAL

59. Finally, I address Mr. Bradner's repeated contention that the filter-then-copy-and-scan configuration using whitelists and/or blacklists would not "guarantee" that the NSA avoids all interaction with Wikimedia communications during the Upstream collection process. Bradner Reply Decl. ¶ 57; *see also id.* ¶¶ 30, 64, 97-101, 115, 154(5). In support of this contention, Mr. Bradner identifies a number of hypothetical scenarios in which Wikimedia communications would be copied and scanned if all of the necessary conditions were met for these scenarios to come to pass. I have already discussed most of the scenarios outlined by Mr. Bradner above, and in my second declaration. I address the scenarios newly conceived of by Mr. Bradner below. Whether any of them has occurred or would ever come to pass is a matter of speculation for which Mr. Bradner gives no evidence.

60. Whitelisting by IP address: First, Mr. Bradner points to a situation in which a user of a whitelisted IP address communicates with Wikimedia, and the communication traverses an international Internet link (hypothetically) monitored by the NSA. Bradner Reply Decl. ¶ 154(5)(a)(i); *see also id.* ¶ 154(5)(b)(iii). I acknowledged in both my first and second declarations the theoretical possibility that Wikimedia communications of this kind could be copied and scanned during Upstream surveillance, even if the NSA used a whitelisting technique

[6] Mr. Bradner wonders why I also addressed the U.S. Government's cyber-defense system known as Einstein 2.0, when he mentioned it only "in passing" in his first declaration. Bradner Reply Decl. ¶ 150. As I explained in my second declaration, I addressed the implications of Einstein 2.0 because Wikimedia, in its legal brief, attempted to rely on Einstein 2.0 as "corroboration" for Mr. Bradner's conclusions, even though Mr. Bradner himself had not done so, Second Decl. ¶ 66, and does not do so now. For all intents and purpose, Mr. Bradner now disregards my explanation of the reasons why conclusions about the Upstream collection process cannot be drawn from Einstein 2.0. *Compare* Second Decl. ¶¶ 68-69 to Bradner Reply Decl. ¶¶ 151-53.

that otherwise excluded Wikimedia's communications. Second Decl. ¶ 43; First Decl. ¶ 81. But this scenario, like the blacklisting scenarios posited by Mr. Bradner, is conjectural. *See* paragraphs 42-43, above; Second Decl. ¶¶ 78-85. Mr. Bradner cites no evidence of the number or geographic locations of persons using whitelisted IP addresses who communicate with Wikimedia, and of course could not do so without knowing the composition of whitelists (hypothetically) employed by the NSA. There is no basis, therefore, on which to conclude that communications between Wikimedia and persons using whitelisted IP addresses would cross every international Internet link to and from the United States (as Wikimedia claims of its communications generally), or, for that matter, that they would cross one or more links (if any) that happen to be monitored by the NSA. In addition, the communications in question must not themselves be blacklisted, as might be the case if they were encrypted, *see* Second Decl. ¶¶ 80, 82, 84, rendering this scenario even more uncertain.

61. Blacklisting by IP address: Mr. Bradner next mentions three scenarios in which he states that blacklisting by IP address would not "guarantee" that the NSA would avoid all interaction with Wikimedia communications. Bradner Reply Decl. ¶ 154(5)(b). The first of these, involving the enclosure of a Wikimedia communication within an MCT, Bradner Reply Decl. ¶ 154(5)(b)(i), is the same as the first of the three scenarios he posited in his first declaration, Bradner Decl. ¶ 367(b)(1), which I have already discussed in my second declaration, Second Decl. ¶¶ 78-80, and again in paragraphs 42-43, above. The second scenario, involving the passage of a Wikimedia communication through an intermediary service such as a VPN, or an email server, Bradner Reply Decl. ¶ 154(5)(b)(ii), is a generic restatement of the second and third scenarios hypothesized in Mr. Bradner's first declaration, Bradner Decl. ¶ 367(b)(2)-(3), which I discussed in my second declaration, Second Decl. ¶¶ 81-84, and again above in paragraphs 42-43. The third blacklisting scenario suggested in Mr. Bradner's second declaration, Bradner Reply Decl. ¶ 154(5)(b(iii), is a repeat of the whitelisting scenario he referred to in paragraph 154(5)(a)(i) of his second declaration, and that I addressed in paragraph 60, above.

62. Port or protocol blacklisting: The last two scenarios hypothesized by Mr. Bradner concern blacklisting by port or protocol (type of communication) rather than by IP address. Bradner Reply Decl. ¶ 154(5)(c). He envisions two situations in which a Wikimedia communication, even if using a blacklisted protocol, would still be copied and scanned during Upstream surveillance if it crossed a monitored link and either (i) it were enclosed in an MCT using a different protocol, one not blacklisted by the NSA, or (ii) it passed through an intermediary (such as an email server) that, likewise, used a different protocol not blacklisted by the NSA.

63. Mr. Bradner gives no examples of either kind of supposed communication, in which a communication using one protocol is transported within another communication using a different protocol. Nor does he give evidence of how frequently such supposed communications could be expected to occur, or under what circumstances. Consequently, the occurrence of such a communication in the first place is itself a matter of speculation. An even greater degree of speculation would then be required to imagine (i) that the protocol used by the enclosing MCT or the intermediary service is not blacklisted by the NSA, (ii) that the IP addresses assigned to the communication within the MCT, or by the intermediary service, are not excluded to due whitelisting, and (iii) that the communication happened to cross an international Internet link monitored by the NSA (if any).

64. For all of the reasons I have explained herein (paragraphs 42-43, 60-63, above), and in my second declaration, Second Decl. ¶¶ 78-85, whether and when any of the scenarios envisioned by Mr. Bradner might come to pass at a particular international Internet link that happened to be monitored by the NSA (if any), such that the NSA would copy and scan communications of Wikimedia's, is a matter of speculation.

CONCLUSION

65. For the reasons I discuss above and in my first two declarations, it remains my opinion that, based on what is publicly known about the NSA's Upstream collection technique, the NSA in theory could be conducting this activity, at least as Wikimedia conceives of it, in a

number of technically feasible, readily implemented ways that could avoid NSA interaction with Wikimedia's online communications.

66. While I offer no opinion on the likelihood that the NSA does or does not, in fact, employ these techniques, I have previously examined, and now re-examined, the bases of Mr. Bradner's opinions (i) that the NSA, in conducting Upstream surveillance, "most likely" copies, reassembles, and scans for selectors all communications packets traversing an international Internet link that is monitored by the NSA (if any); (ii) that it is "implausible" that the NSA uses the traffic-mirroring techniques (white- and blacklisting) described in my first declaration; and (iii) that even if the NSA uses one or more of the techniques I described, it is still "virtually certain" that the NSA copies and scans at least some of Wikimedia's communications. I still conclude that these opinions lack a non-speculative foundation in Internet technology and engineering.

67. My opinions are unaltered by the statements referred to by Mr. Bradner in the FISC's October 2011 opinion and the PCLOB Section 702 Report. As I explained above, the use of traffic-mirroring techniques to implement a filter-then-copy-and-scan approach to Upstream collection would be entirely consistent with both statements. My opinions are unaltered, as well, by the so-called "technical and practical necessities" discussed by Mr. Bradner in his reply declaration. With few exceptions, they are simply reiterations of the same grounds given for the conclusions reached in his second declaration. They, too, are principally based on speculation about the NSA's surveillance practices and priorities, its capabilities and resources, and the number, nature, and communications practices of its Upstream surveillance targets, and lack a non-speculative foundation in Internet technology and engineering.

I declare of penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief.

Executed in New York, New York on March 22, 2019.



HENNING G. SCHULZRINNE

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, Plaintiff, v. NATIONAL SECURITY AGENCY, *et al.*, Defendants.	Civil Action No. 1:15-cv-00662-TSE

Exhibit 17

SECOND DECLARATION OF DR. ALAN J. SALZBERG

Dr. Alan Salzberg, for his second declaration pursuant to 28 U.S.C. § 1746, deposes and says as follows:

I. Introduction

1. I am the Principal (and owner) of Salt Hill Statistical Consulting. I previously submitted a declaration in this case, dated February 14, 2019 ("Salzberg February Declaration"). My February declaration commented on the "Declaration of Jonathon Penney" ("Penney December Declaration"), which was submitted in December 2018. I submit this second declaration at the request of the United States Department of Justice in response to the "Reply Declaration of Jonathon Penney" ("Penney Reply"), which was submitted on March 8, 2019. I have previously submitted my resume describing my background and qualifications in statistical sampling, analysis, and review for government and industry, as well as information regarding prior testimony and fees.

2. This report proceeds as follows. In the next section, I summarize my findings. In the third section, I detail those findings. In the fourth section, I set forth my conclusions. Finally, I have included an appendix with a program log showing the results of additional analyses.

II. Summary of Findings

3. In my February Declaration I addressed the deeply flawed model presented by Dr. Jonathon Penney in his December Declaration. Specifically, in summary I previously found as follows:[1]

   A. "The methodology used in the Penney Declaration—which purportedly shows an upward trend in page views of certain articles posted on Wikipedia through May 2013, followed by an abrupt drop and downward trend in views of those articles beginning in June 2013—is deeply flawed, inappropriate, and likely biased."

   B. "The Penney Model simply assumes that a single change occurred in June 2013, rather than letting the data identify the timing and number of changes in trends that occurred. Even though there is no consistent trend in the data, the design of the Penney Model will create the appearance that the data contain just one inflection point. And, because of its design—even though changes in trend occurred before these June 2013 disclosures—the Penney Model will find that the disclosures caused them."

   C. "Contrary to the hypothesis presented in the Penney Declaration, analysis of page views for the 48 individual articles in the privacy-sensitive group do not show a rising trend followed by an immediate and sustained drop in June 2013."

   D. "With the one exception of removing the article on Hamas, the Penney Declaration does no analysis or adjustment for factors (such as world events) affecting these individual article page views. Instead, the Penney Declaration inappropriately aggregates the vastly different page view data for individual articles, with the result that these individual differences in page views are masked."

[1] Salzberg February Declaration, paragraph 4.

E. "Even at that aggregate level, I find that the hypothesized peak in page views of "privacy-sensitive" articles in May 2013 does not exist, and the hypothesized upward and then downward trends in views of privacy-sensitive articles before and after June 2013, respectively, do not exist."

F. "Extended data through 2018 regarding page views of the privacy-sensitive articles do not indicate a long-term decline in page views from pre-June 2013 levels."

G. "A proper control dataset would exhibit similar page view behavior prior to June 2013. The comparison datasets used in the Penney Declaration do not and are thus inappropriate controls."

H. "The Penney Declaration analysis ends in July 2014. No data are presented that shed any light on whether page views at the time the Amended Complaint was filed in 2015 (or thereafter) were affected by Upstream. In other words, even if the purported effect and trends were a correct conclusion for the data examined (and they are not), the Penney Declaration analysis does not and cannot show that the effect continued years after the study ended."

I. "Even if a chilling effect occurred in June 2013, there are no data analyzed in the Penney Declaration that show any effect was due specifically to "public awareness of" the specific NSA surveillance program challenged here (known as Upstream surveillance) rather than possible inaccuracies, if any, about the program reported in the press, disclosures about other NSA programs, disclosures about other surveillance programs (e.g., surveillance by Britain), or other, unrelated events of June 2013."

4. As discussed in detail below, the Penney Reply does not raise any valid critiques of my original findings, and the additional analyses in the Penney Reply do not bolster the flawed model presented in the Penney December Declaration. In addition, the Penney Reply does not propose a new model that corrects the flawed model presented in the Penny December Declaration, and the slight modifications attempted do not address any of the issues I raised. Therefore, my findings and conclusions set forth in my February declaration remain unchanged.

III. Details of Findings

5. The Penney Reply begins with seven critiques of my analyses, in paragraphs 4 through 23 of the Penney Reply, and goes on to respond to my critiques in paragraphs 25-36. I reviewed all of the Penney Reply and in this declaration I organize my responses by topic, so as not to be repetitive. In particular, this section proceeds with the following six subsections:

A. Overview of the Incorrect Assumptions Made in the Penney Reply;

B. Spurious Statistical Conclusions from the Penney Model are Partly Due to Aggregation of the Article View Data;

C. The Penney Reply's Additional Analyses Fail to Address the Flaws in the Penney Model;

D. Data Beyond Time Period 2014 Show Article Views at About 2012 through 2014 Levels, Even When Earlier Data is Corrected for Mobile Views;;

E. Omitted Variable Bias of the Penney Model Cannot be Solved by Deleting Valid Data; and

F. The Penney Model's Failure to Isolate the Effect of Awareness of NSA's Upstream Program.

A. Overview of the Incorrect Assumptions Made in the Penney Reply.

6. Before responding to the specific claims of the Penney Reply, I first address some false assumptions the Penney Reply made regarding my critiques.

7. First, while my review of the disaggregated data provides an important, simplified explanation of many of the flaws of the Penney Model, the flaws of that model remain whether that model is applied to the aggregated or the disaggregated data. I am not suggesting that the application of the deeply flawed Penney Model to each of the 48 articles, individually, would be appropriate, nor am I suggesting that there could never be theoretical circumstances where the data could be aggregated without presenting the deeply flawed and misleading results that the Penney Model presented here.

8. Second, my February report provided no wholesale critique of the so-called ITS "Interrupted Time Series" designs or of regression models in general. My critiques instead relate to the particular methods Dr. Penney employed and the underlying data used in the Penney December Declaration.

9. Third, as I pointed out a number of times in my February Report,[2] I do not present an alternate model of page views, but I do use a number of examples and perform analyses that demonstrate the flaws in the Penney Model. Statements in the Penney Reply regarding "alternatives" that I suggest are therefore misleading.

10. Fourth, while the Penney Reply is correct in that much of my analysis uses "visual inspection" as an aid to understanding the issues with the Penney Model, I also perform statistical tests and point out many specific flaws in the Penney December models. As with the issue of aggregation and disaggregation, I am not advocating one or the other, but rather, doing both. A simple visual review of the data using charts and graphs, such as I the one performed, has long been considered a fundamental component to developing correct statistical models.

B. Spurious Statistical Conclusions from the Penney Model are Partly Due to Aggregation of the Article View Data

11. As I stated in my original declaration, a review of the disaggregated data leads to the conclusion that there is no May 2013 peak or steep drop beginning in June 2013, contrary to the Penney December Declaration's conclusion.[3] The aggregated data do not show a May 2013 peak either, but rather an April 2013 peak,[4] followed by declines beginning in May 2013. This means that both the disaggregated data and the aggregated data show that the drop in page views begins *before* the June 2013 disclosures.

[2] In my February Declaration, for example, in paragraphs 52 (note 31) and 53, I explicitly state that I am not proposing an alternative model.
[3] Salzberg February Declaration, paragraphs 11-17.
[4] As stated, while the Boston Marathon bombings are one possible reason for an April peak, such a simple model is far from adequate for many of the same reasons that the Penney Model is inadequate, but at least in the April peak model, the drop in page views comes after the purported cause and not before the purported cause.

12. The Penney Reply responds to this critique by erroneously claiming that my disaggregated review should be disregarded because aggregation is appropriate here. This somewhat misses the point, which is that neither the disaggregated data nor the aggregated data support Dr. Penney's conclusions. Nonetheless, I reviewed the literature that the Penney Reply cites in support of its claim. The literature cited does not support the Penney Reply's claim.[5] The articles cited are general articles on ITS designs rather than articles espousing aggregation.[6] Moreover, the four reasons cited in the Penney Reply paragraph four are not specific to aggregated data (they apply equally to disaggregated data).

13. In paragraph 26b, the Penney Reply acknowledges that there is "there is no single determinative method or factor to decide whether an aggregated or disaggregated analysis of data is appropriate." As I stated, by aggregating all the data prior to analysis, there is no possibility of correcting for any article-specific differences in the data or exploring whether there are important differences among article views that need to be accounted for in any model. This leads to a biased model and erroneous claims of statistical significance where, as here, such differences exist.[7] The idea of not reviewing and understanding the disaggregated data, and discarding information by inappropriately aggregating that data, is anathema to scientists, because such ignorance often leads to false conclusions. That review typically includes graphical analysis, because, as one statistician put it: "[g]raphics reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations."[8]

14. The Penney Reply argues that that my use of simple graphs to provide a visual inspection of the disaggregated data should be disregarded, in part, because a "visual inspection of data . . . can often be misleading," a point he makes with a quotation of one of the great proponents of graphical analysis, Dr. Howard Wainer.[9] Dr. Wainer, however, is not saying that graphs should not be used; he is only saying to be careful that they are not used in a misleading manner.[10] Ironically, by ignoring the disaggregated data and aggregating dissimilar page views to tell a

---

[5] Specifically, in the footnotes for paragraphs 4-5, the Penney Reply identifies several sources that Dr. Penney claims supports his use of aggregation in this circumstance. The only citation that even appears to support aggregation, in this type of situation, is not from a paper or textbook but from a PowerPoint presentation by Emma Beard which appears to have been presented at a conference in London (*see* footnotes 3, 4, and 21 in the Penney Reply). I reviewed the PowerPoint presentation and it presents no reasoning or data to support the claims (nor is it obvious that the author even made such claims regarding a simple regression model like the one in the Penney December Declaration). Additionally, unlike a scholarly article, a PowerPoint presented at a conference is typically accompanied by an oral portion of the presentation that may provide additional context or present the point differently than the language on the printed slides). In short, none of the cited source materials in the Penney Reply alter my conclusion that in this instance the use of aggregated data is inappropriate and misleading.

[6] The Penney Reply in paragraph 26e, takes issue with my terming the data "panel data" and not "time series" data. Panel data is a form of time series data, as the introduction to the text and chapter on panel data in my source make clear. See Wooldridge, Jeffrey M., Introductory Econometrics, A Modern Approach, 5th Edition, 2012, South-Western Cengage Learning, p. 10 and 448.

[7] See Salzberg February Report, paragraphs 56-60, for example.

[8] Tufte, Edward R., The Visual Display of Quantitative Information, Graphics Press LLC, 2001, p. 13. Also, p. 9 of the same text states that: "Often the most effective way to describe, explore, and summarize a set of numbers – even a very large set – is to look at pictures of those numbers. Furthermore, of all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful."

[9] Penney Reply, paragraph 3 footnote 1.

[10] Quote from Penney Reply, paragraph 3 footnote 1 (quoting Howard Wainer).

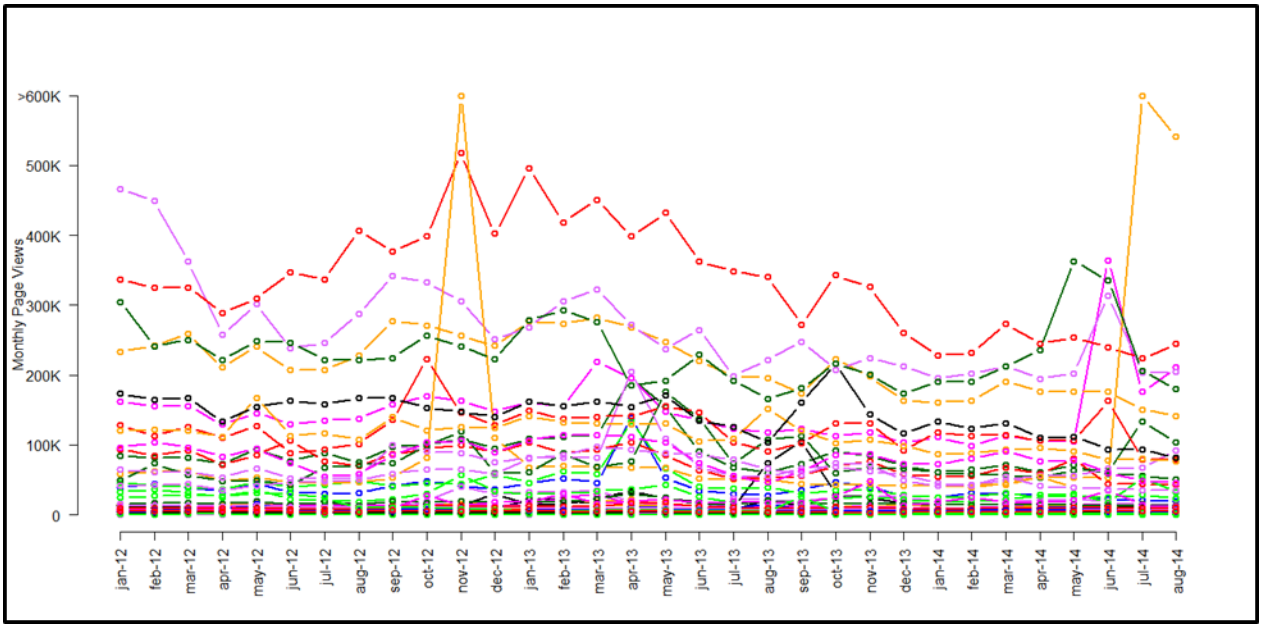
misleading story, Dr. Penney has created precisely the type of misleading graphs that Dr. Wainer is warning against. As Yale statistician Edward Tufte says: "[a]ggregations by area can sometimes mask and even distort the true story of the data"[11] and "[a]ggregations over time may also mask relevant detail and generate misleading signals."[12] Tufte concludes:"[i]f in doubt, graph the detailed underlying data to assess the effects of aggregation."[13] As I explained in my first Declaration (Paragraphs 18-26), Figure 2 of the Penney December Declaration is misleading because it inappropriately aggregates the data and shows a suggestive regression line, while obscuring the fact that the decline was not as indicated.[14]

15. I created and included (as Appendix IV to my first Declaration) graphs of each of the 48 articles' page views individually, so that all the data is available to view in a clear graphical form. I invited (and invite) review of each of those graphs. The only reasonable conclusion from a review of those graphs is that the effect supposedly found in the Penney December Declaration is spurious. I also included graphs of the aggregated data (*see* paragraphs 18-26 of my February Declaration), and those graphs also do not indicate a May 2013 peak. I did not leave out anything or "cherry-pick," contrary to what the Penney Reply states in paragraphs 11, 12, and 32(a).[15]

16. The Penney Reply claims that disaggregation adds "noise" to the data, "both visual and statistical," and points to my first graph showing all 48 articles in a single figure.[16] I showed all the data in a single figure (as well as in 48 separate figures in Appendix IV) because it provides important context and a comparison point to Dr. Penney's aggregated plot, which artificially smooths the differences.[17] As a reminder, my Figure that includes all 48 articles is below.

[11] Tufte, Edward R., Visual Explanations, Graphics Press LLC, 1997, p. 35.
[12] Ibid, p. 36.
[13] Ibid, p. 37.
[14] Gelman, Andrew and Zelizer, Adam, "Evidence on the deleterious impact of sustained use of polynomial regression on causal inference," Research and Politics, January-March 2015, also cited in the Penney Reply, is also clear that graphical analysis is recommended.
[15] Penney Reply, paragraph 12.
[16] Penney Reply, paragraph 5.
[17] This method, of putting all the data into a single plot, is done in so-called spark graphs, examples of which can be found in Tufte, Edward R., Beautiful Evidence, Graphics Press LLC, 2006. p. 47-63.

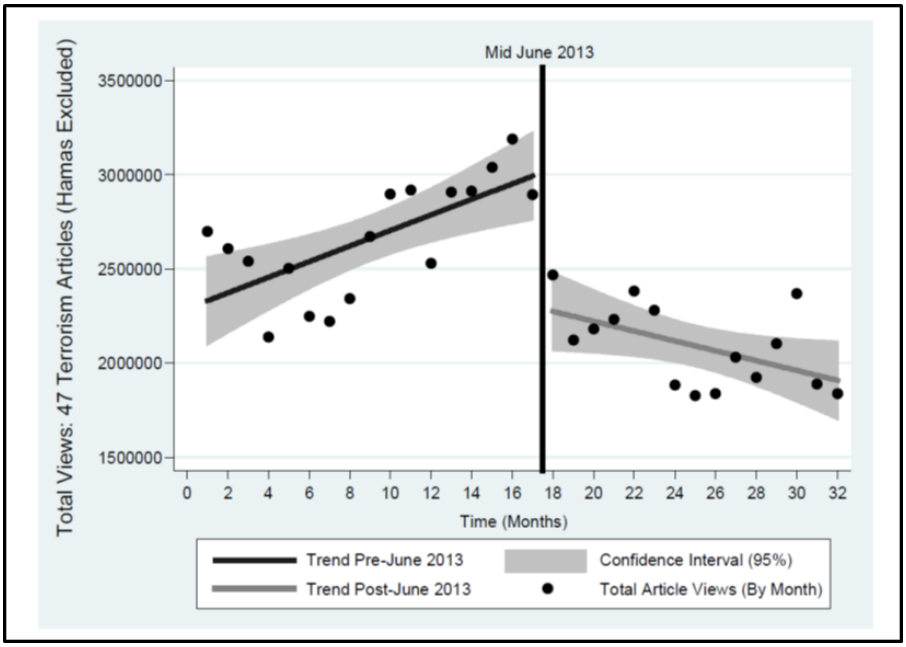
Figure 1: Page Views for all 48 Articles Considered to Suffer from a Chilling Effect beginning in June 2013



17. In contrast, Penney's Figure 2 from his December Declaration, shown below, misleadingly indicates a simple up and down movement that is belied by the individual data in Figure 1, above. The same is true for the confidence intervals drawn on Penney's Figure 2 graph itself, as I pointed out in my February declaration.[18]

[18] See Salzberg February Declaration, paragraph 20.

Figure 2: Penney December Declaration Aggregate Figure Masks Individual Differences



18. The contrast between the simple, disaggregated view and Penney's misleading aggregate view should have led Dr. Penney to question whether his aggregated model masks systematic differences at the article-level. As I stated above, I also provided each plot individually, so the reader can see what is behind the total picture of page-views shown in my first figure.

19. The Penney Reply is also wrong when it categorically states that disaggregation adds noise. The Penney Reply concerns that a disaggregated model will not allow for estimation of an "aggregate level inference about large scale NSA surveillance effects"[19] are misplaced. If the same naive model is run on both datasets, the estimated effect in the disaggregated model is exactly the same as the estimated effect in the aggregate model.[20] The statistical significance of these effects will also be the same if the disaggregation only adds noise to the model, and I show this fact through a simulation.[21]

20. However, if the disaggregated data reveal systematic differences in the data, in that the individual articles' page views do not tell the same or even a similar story as the aggregated data, then the naive model needs to be modified in order to avoid bias, whether run on aggregate or individual article data. To further support the analysis I already performed showing the model is over-simplified and perhaps mis-specified, I performed a statistical test to determine whether the

[19] Penney Reply, paragraph 26b suggests that because the question regards aggregate differences the aggregated data must be used.

[20] This fact is shown in the Appendix to this Declaration, and can be observed by noting the coefficient estimates for the Penney Model as shown in my Appendix. In the Appendix, I run the Penney Model on the averages and run the same model on the individual articles. The estimated effects (model coefficients) are exactly the same.

[21] I have included in the Appendix a simulation that shows the results of running the Penney Model on aggregated and disaggregated data are the same when the errors are statistical noise. This includes not only the regression coefficients (which will be the same whether the difference are due to noise or not, as explained above) but also the standard errors (i.e., the statistical significance) of those coefficients.

differences by article are mere noise or systematic.[22] I found, with high statistical significance, that the differences among articles are systematic (the statistical results are in the Appendix to this Declaration). This means that the model used in the Penney Reply is incorrect, *whether using the aggregated or the disaggregated data.* Only by accounting for the article-level, seasonal, and other differences can a valid model or set of models be produced. Furthermore, the model's estimates show increased error when calculated in disaggregated form.[23] This fact confirms my conclusions in my February report.[24] Because the differences are systematic and not mere "noise," the aggregation produces a result with inflated statistical significance.[25]

21. In reviewing some of the specific examples I cited to explain the fact that aggregating the data masks differences in the articles, the Penney Reply re-explains some analyses and runs additional models, but none address the issues I raised.[26] The Penney Reply presents Figures 2A and 2B, which purported show an "Increase until June 2013 and then a Sharp Drop-off."[27] This labeling is wrong. The increase is only through April, with a drop off in May and a continuation of that drop in June. This fact can be seen in Penney Reply's own Figures 2A, 2B, 3A, and 4 of the Penney Reply. Each shows an April and not a May peak, and a May and not a June start to the drop in page views. As I explain in my first Declaration, the fact that the drop in page views began *before* the June 2013 disclosures does not support Dr. Penney's conclusion that the June 2013 disclosures caused the drop in page views, and violates a basic tenet of causal models (i.e., a cause cannot occur after an effect).[28]

22. The models using the data in Figures 2, 3, and 4 of the Penney Reply suffer from the same problems as the original model in the Penney December Declaration.[29] The Penney Reply

[22] See Salzberg February Report, paragraphs 55-60 for my comments regarding the over-simplified model and omitted variable bias.

[23] As shown in my appendix, attached here, in some cases the claimed effects are not statistically significant. In other cases the statistical significance is weaker. These are further indications that the article differences are not mere noise. As my simulation (in the Appendix attached here) shows, when differences are based on mere noise, the statistical significance of the coefficients for the effects will remain unchanged when running the model on aggregated versus disaggregated data.

[24] See Salzberg February Report, paragraph 48-50 and 55-60.

[25] This is due to omitted variable bias, among other factors. I pointed this out in my February Report, paragraph 56. I do not attempt to correct for the omitted variable bias by adding additional variables, and therefore the disaggregated model is also incorrect.

[26] These re-analyses and the Penney Reply's commentary on them is found in Penney Reply, paragraphs 6-22 and paragraphs 26, 28, and 30.

[27] Penney Reply, Figures 2A and 2B.

[28] For two examples of such spurious inferences that ascribed a later cause to an earlier effect, see a source cited in the Penney Reply: McCleary, Richard, McDowall, David, and Bartos, Bradley J., Design and Analysis of Time Series Experiments, Oxford University Press, 2017. The examples are portrayed in this text in Figure 5.15 (explained on p. 214-215) and Figure 7.1 (explained on p. 275-276), and involve "interventions" and data with similarities to the data analyzed in the Penney December Declaration.

[29] The Penney Reply inexplicably discards its high-privacy group of 31 articles in favor of a new high privacy group of 23 articles for Figure 4 and some accompanying analyses. The Penney December Declaration already determined (perhaps also arbitrarily) a 31-article set that is highly privacy sensitive and this new set of 23 is a subset of those articles. Of course, re-running the same model on datasets that are nearly the same will produce results that are nearly the same, and proves nothing.

analysis ignores the large and obvious effect of events of April 2013 in its analysis of "improvised explosive device," "dirty bomb," "car bomb," and "ammonium nitrate."[30]

23. The only graph that the Penney Reply shows that appears to have a peak in May is Figure 3b (page views for so-called "normalized" Ammonium Nitrate), but that supposed "peak" is artificially created because the Penney Reply manipulated the graph to remove the April peak and replace it with the average of the March and May.[31] Removing such outliers and replacing them with averages in this way is against the practice of statisticians in general. Outlier handling is discussed in detail in an article the Penney Reply cites (at footnote 8), and this article says such adjustment is only appropriate for *error* outliers.[32] Here, the data points for Ammonium Nitrate page views are not errors and so removing the correct data point and replacing it with an average is inappropriate.[33]

C. The Penney Reply's Additional Analyses Fail to Address the Flaws in the Penney Model

24. Paragraphs 18 and 28 of the Penney Reply assert that no assumption is made in the Penney Model concerning a May peak. However, the Penney Model hypothesis is a single trend line through May 2013, and then a second line, starting in a potentially different place. The assumption is a single point of inflection, and that point is a peak in May and a drop off beginning in June.[34] While it is correct that the model can find that there is no peak at all in the data, my point is that no other month is modeled as a possibility, and that if the data goes up and down, the model finding a June peak will be statistically significant even though the peak did not occur in May and the drop did not begin in June.

25. The Penney Reply in paragraph 28 criticizes my demonstration, using a polynomial model, that the peak did not occur in May and says such an approach is biased, citing a scholarly article.[35] That article refers to higher order polynomials (which I did not use) and, even for higher order polynomials, the article does not say that such models are biased, only that they may not reduce bias.[36] Indeed, as shown in the quote below, the article brings up the same issues that I do with

---

[30] While the Boston Marathon bombings did not use ammonium nitrate and were not a "dirty bomb," this does not mean they may not have been a reason for a huge uptick in page views. Some news articles (for example https://www.theatlantic.com/technology/archive/2013/04/new-boston-bomb-parts-photos/316183/) discussed the possibility of ammonium nitrate being used. Even if the Boston Marathon bombings had nothing to do with the April uptick in page views, the complete exclusion of any cause of those changes biases the Penney Model, as I have explained.

[31] Penney Reply, paragraph 14 and footnote 8. See page 11 of the Penney Reply for the graph of Ammonium Nitrate views without April data deleted and replaced with the average of March and May 2013.

[32] The article is Aguinis, Herman, Gottfredson, Ryan K., and Joo, Harry, "Best-Practice Recommendations for Defining, Identifying, and Handling Outliers," Organizational Research Methods, 16(2), 2013, p. 270-301.

[33] Neither Dr. Penney nor I have suggested that the change in views in ammonium nitrate in April 2013 was due to an error in the archives used to collect the data.

[34] Penney December Declaration, paragraph 23, describes the design as testing for a "decrease in level and trend" beginning in June 2013.

[35] The article, cited in paragraph 28(b), footnote 32 of the Penney Reply, is Gelman, Andrew and Zelizer, Adam, "Evidence on the deleterious impact of sustained use of polynomial regression on causal inference," Research and Politics, January-March 2015

[36] Gelman, Andrew and Zelizer, Adam, "Evidence on the deleterious impact of sustained use of polynomial regression on causal inference," Research and Politics, January-March 2015, p. 5.

respect to simplistic linear models, saying that modeling higher polynomial effects does not necessarily fix those issues:

> "the higher-order polynomial has the effect of slightly modifying and improving the fit of the natural linear model. In criticizing the use of high-degree polynomials in RD [RD stands for Regression Discontinuity—the issue theorized in the Penney December Declaration] adjustments, we are not recommending global linear adjustments as an alternative...We recommend that any RD analysis include a plot such as Figure 1 showing data and the fitted model, and that users be wary of any resulting inferences based on fits that don't make substantive sense."[37]

26. In other words, plotting the data is recommended, and the authors are not recommending that a simple linear model is better than a polynomial one. Indeed, they preface that discussion specifically with:

> "Our point here is not to argue that the linear model is correct...Our point is rather that the headline claim, and its statistical significance, is highly dependent on a model choice that may have a data-analytic purpose, but which has no particular scientific basis. Figure 1 indicates to us that neither the linear nor the cubic nor any other polynomial model is appropriate here. Instead, there are other variables not included in the model which distinguish the circles in the graph."[38]

27. I include these extended quotes because despite the Penney Reply's misinterpretation, the article is useful in that it points out the very issue of spurious statistical significance and omitted variable bias that is at the heart of my critiques of the Penney Model in the first place.

28. Next, Paragraphs 19 through 22 of the Penney Reply describes a series of analyses of the single peak May model against other single peak models, concluding that the June model (with a May peak) is better than the others. These analyses are flawed in numerous ways.

29. First and most importantly, the entire exercise is based on a mischaracterization of my critique that implicitly assumes I am proposing a model with an April peak. I merely stated that a *naive* model such as the Penney Model could also be used to "prove" an April peak, meaning that such an analysis could also lead to spurious statistical significance. None of the Penney Reply analyses question this fact. I am not proposing that the data experienced a single change that caused the trend to abruptly reverse after April 2013 (a peak in that month and a decline thereafter). As I have stated numerous times, the data do not indicate a single change model is appropriate, whether that single change is in June 2013 or in some other month.

30. Second, in paragraph 19 of the Penney Reply, Dr. Penney attempts to complete a cross-validation analysis that uses three data sets for each of these article sets. However, two of the three models proposed in paragraph 19 of the Penney Reply, the "total page view" model and the "average total page view" model, are exactly the same statistically.[39] The total page view is simply the average page views multiplied by the number of articles. These two models are equivalent,

[37] Ibid, p. 6.
[38] Ibid, p. 3-4.
[39] Penney Reply, paragraph 19.

statistically, since regression models are invariant to changes in units.[40] For example, suppose we were trying to predict how far a person can jump according to their height in feet, and we ran a regression model that predicted someone who is 6 feet tall can jump 10 feet on average. If we use the same data but run the regression model based on inches, that new model would predict that someone who is 72 inches tall can jump 120 inches on average – in other words, the prediction is unchanged except for the expression in inches instead of feet.

31. The same is the case with running one model on the total and a second on the average, as is done in the Penney Reply (the results of which are summarized in the Penney Reply, paragraph 22). The model is unchanged but one is in terms of averages and one is in terms of totals. Therefore, the estimates for the model run on totals will be 23 times the estimates for the model run on the averages (for the Penney Reply model that has 23 articles). Thus, for example, in the Appendix to the Penney Reply showing the "23 Most Privacy Sensitive Article Set Cross Validation Analysis" (page 41), the coefficient for the variable *time* for the total model is shown to be 21,383.58. Two pages later (page 43), the same coefficient for the variable *time* in the average model is 929.72, which is exactly 21,383.58 divided by 23. The summary statistics like the t-statistic, which is 5.30, are also exactly the same.[41] The Root Mean Square Error and Mean Absolute Errors highlighted for the total model are 89,506.35 and 63,503.27 (on page 41), which, when divided by the 23 articles considered, is equal to the highlighted totals of 3,891.54 and 2760.94 shown for the average model for the highlighted RMSE and Mean Absolute Error, respectively, shown in the attachments to the Penney Reply (on page 43).[42]

32. Thus, while the Penney Reply asserts that there are 48 models (3 models by 4 datasets by 4 change points), there are really only 32 (2 models by 4 datasets by 4 change points). The four datasets also largely overlap, since the 46 article dataset includes all 44 articles in the 44 article dataset, which includes all 23 articles in the 23 article dataset, which includes all 21 articles in the 21 article dataset. In addition, the four months modeled are adjacent, meaning the regression models are very similar (this was part of my original point that the specification of the change point does not make much difference). In other words, though the Penney Reply asserts there are 48 separate models, there are only 32, and most of the 32 are highly related to one another and must produce similar results.

33. Third, the Penney Reply's use of cross validation is misplaced and performed incorrectly. In part the Penney Reply employs a "cross validation analysis."[43] This approach, which the Penney Reply uses to delete different time periods one at a time, is improper for time series models, in which the data points are related to one another.[44] In addition, the Penney Reply's cross

[40] See, for example, Wooldridge, Jeffrey M., Introductory Econometrics, A Modern Approach, 5th Edition, 2012, South-Western Cengage Learning, p. 40-41.
[41] The r-squared and the p-values are also exactly the same.
[42] There is a slight difference due to rounding or less than 1 for each of the figures.
[43] Penney Reply, paragraph 19.
[44] This is because the data in the cross validation set, or the data "left out", is not independent of the other data. See for example, Bergmeir, Christopher, and Benitez, Jose M., *"On the use of cross-validation for time series predictor evaluation,"* Information Sciences, 2012, 192-213. This paper discusses some of the fundamental problems with traditional cross-validation in time series, primarily in Sections 3.3 and 3.4. Also, see David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig and Carsten F. Dormann, *"Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography 40: 913-929 (913-925 in particular), 2017.*

validation's purpose is to compare the June model to models with a different change point. None of the Penney Reply's cross validation analyses compare the simple single-change model to a model that accounts for other factors or otherwise corrects for omitted variables. Therefore, the Penney Reply's use of cross-validation to compare models and the attempt to show a May Peak model is better than an April peak model or other models are mere distractions that are not related to my criticism.

D. Data Beyond 2014 Show Article Views at About 2012 through 2014 Levels, Even When Earlier Data is Corrected for Mobile Views

34. In my February Declaration I pointed out that the extended view of page view data also does not indicate any long term decline. The Penney Reply, in paragraphs 23 and 34(g), responds to point out that my extended data includes mobile use while the original data presented in the Penney December Declaration did not. To address this "apples to oranges" comparison, I therefore adjusted the 2012 to 2014 data to account for mobile usage. As I describe below, this adjustment has no effect on my conclusions.

35. I also considered the effect that the non-inclusion of mobile usage and the lack of adjustment of that increasing usage had on the Penney Model. I find that such exclusion and lack of adjustment are additional flaws in the Penney Model. Specifically, the Penney December Declaration data excluded mobile page views from the terror and control data sets.[45] If these views were a constant percentage of total views, such an exclusion would likely not affect the Penney Model. However, as I explain below, there was a dramatic increase in mobile web access from January 2012, the first month of data included in the Penney December Declaration analysis, to August 2014, the last month included.

36. The data provided with the Penney December Declaration (but not used in the Penney December Declaration or the Penney Reply) indicates that in January 2012, mobile views accounted for about 12% of total page views.[46] By the end of the study period, that figure was 32%. In other words, the Penney December Declaration's exclusion of mobile views had an increasingly downward bias on total page views. This is yet another bias that affects the Penney Model, and, by not accounting for mobile visits, the Penney Model is biased toward finding an effect and toward finding a larger effect. This bias is a result of the fact that for later data the model excluded more views than for earlier data.[47]

37. In terms of my graphs of extended data as compared to earlier data, the data prior to August 2014 would be higher with mobile data. My graphs included the data as originally provided with the Penney December Declaration, which did not include mobile data for the terror articles. Using

[45] It may be that mobile views were not available, in which case an adjustment, like the one I made, could have been made; or the Penney Model could have included a factor that accounts for such usage.
[46] This is based on the difference between the global English page views non-mobile and the total global English page views, and is consistent with this article https://techcrunch.com/2016/11/01/mobile-internet-use-passes-desktop-for-the-first-time-study-finds/.
[47] While it may seem that simply using desktop views only would not cause a bias to the results, this notion is not correct. To the extent that 1) mobile use was growing during the period, and 2) individuals were using mobile instead of (rather than in addition to) desktop views, the desktop views would be depressed in the latter part of the period and thus bias the results. This has occurred to such an extent that an increasing number of people rely exclusively on mobile access. See, for example, https://techcrunch.com/2016/11/01/mobile-internet-use-passes-desktop-for-the-first-time-study-finds/.

the Penney December Declaration's global article view dataset, which provides total views as well as total views excluding mobile, I adjusted the page views for the terror articles from January 2012 through August 2014 to account for mobile views.[48] The graphs below, showing the extended average and median page views with mobile page views factored in, are consistent with my earlier graphs of the extended data in that they indicate there was no downward trend after June 2013.[49]

38. Average and median page views appear to decline some in mid-2017 but views in 2015 and 2016 appear to be at or above 2012 through 2014 levels. It is also notable in these longer data series that there are clear peaks around the times of major U.S. or European terror attacks, adding further evidence that any reasonable model would account for such attacks (and of course the Boston Marathon bombings occurred very close to the time of the alleged drop due to the Snowden disclosures).

[48] This rough adjustment is undoubtably inaccurate but captures the magnitude and pattern of the mobile views.
[49] The adjustment results in an increase in article views for each month from January 2012 through August 2014, with the amount of increase depending on the share of total Wikipedia views that were mobile.

Figure 3: Average Page Views, Adjusting Data Before 2015 to Factor in Mobile Page Views

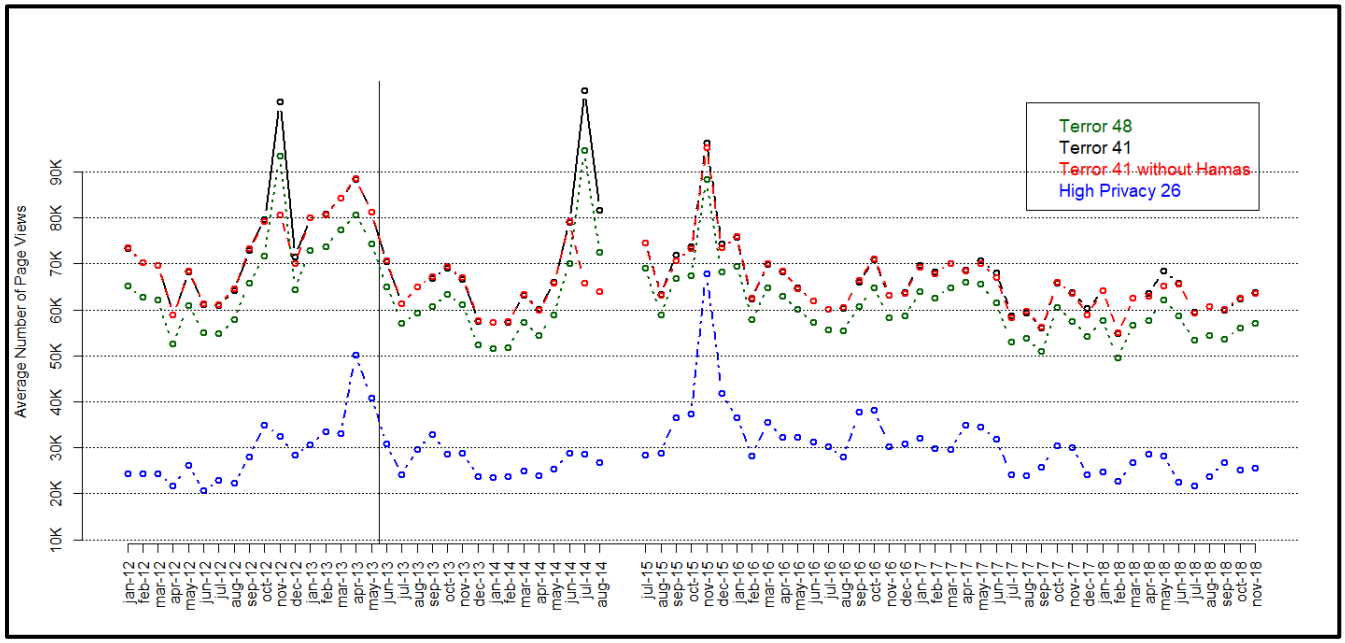
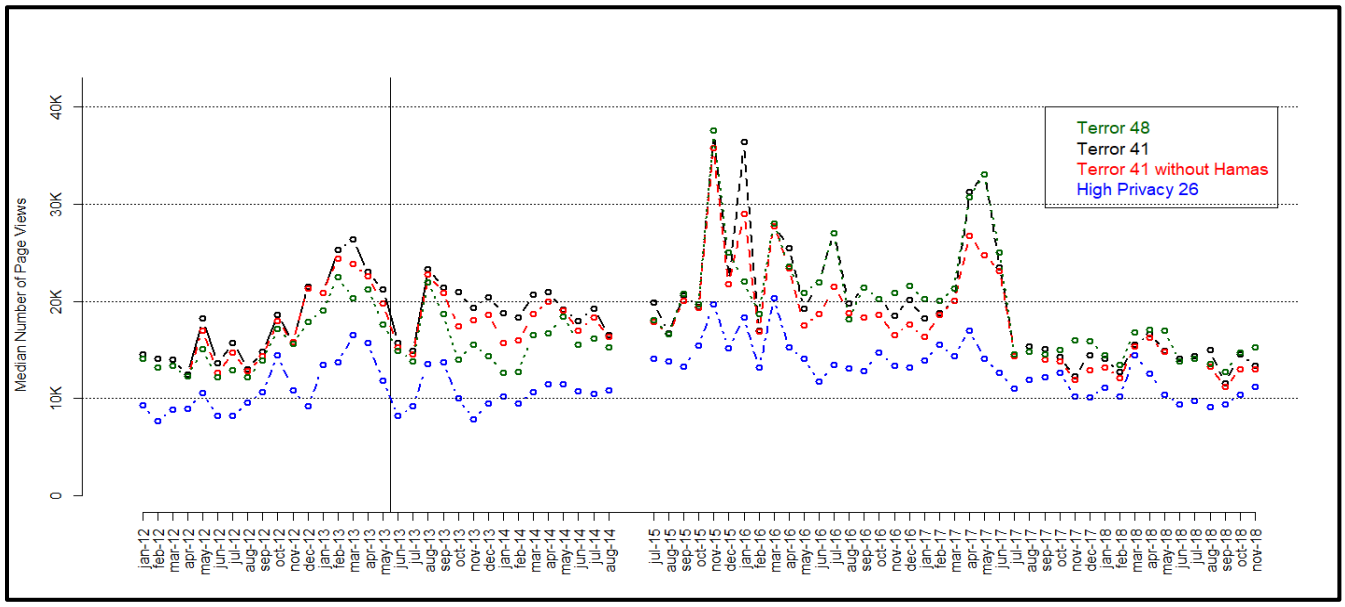


Figure 4: Median Page Views, Adjusting Data Before 2015 to Factor in Mobile Page Views



39. The Penney Reply cites some studies that purport to support the idea that the Penney December Declaration conclusions would continue beyond August 2014, but the Penney Reply neither considers (nor produced in this case) the data underlying those other studies. Even if those studies were to be based on a solid scientific and statistical grounds (and I cannot evaluate whether this is true without the underlying data), they only claim to offer conclusions applicable

to (at the latest) 2015.[50] Moreover, only one of the studies Dr. Penney cites in his reply appears to look at web data, rather than interview answers, and that study is from a working paper that was not published in a scientific journal and it expressly states it only includes data from 2013, and thus does not include any extended data. In any case, there is no way for me to evaluate the validity of those results, because I was not provided the data and it is not publicly available.

40. I do note that one article[51] cited in the Penney Reply footnote 44 adjusts for additional variables and appears to find a smaller (and not statistically significant) effect in terms of drops in searches. This finding is consistent with omitted variable bias I outlined in my first Declaration with respect to the Penney December Declaration.[52]

E. Omitted Variable Bias of the Penney Model Cannot be Solved by Deleting Valid Data.

41. In my February declaration, I pointed out a number of omitted variables that cause bias to the estimates made in the Penney December Declaration. These variables include ones associated with seasonality, individual differences in articles, and news events (the Boston Marathon bombings in particular).[53] The Penney Reply leaves these largely unaddressed but does assert it controls for seasonality because it includes more than one year of data before and after June 2013.[54] However, despite having sufficient data (barely), the Penney Model makes no correction for seasonality and includes no analysis that shows there is not such an effect. I showed such seasonal changes appear in this data and they are statistically significant.[55] In other words, though there was sufficient data, and that data shows statistically significant seasonal effects, the Penney December Declaration ignored seasonality. Wikimedia acknowledged these effects during the deposition of its designee, James Alexander: "global user base, especially in English Wikipedia, tends to have a bit of a dip during the summer, just because there are people out of school, and a lot of people use it in school or when they are studying."[56] Curiously, the Penney Reply, paragraph 30a, states that there is "no basis to expect large seasonal effects with these page views." This statement is speculation that flies in the face of the qualitative and statistical evidence.

42. The Penney declaration correctly states that "in a naturalistic study outside the experimental context, it is not possible to control for all confounding factors."[57] However, the Penney December Declaration corrects for no confounding factors. As one recent author put it: "Obviously, one cannot include in a regression every variable that might conceivably be relevant. But when a factor has a reasonable chance of being important, to exclude it from the modeling is to risk substantial distortion."[58] The Penney Reply re-asserts that the comparator datasets help

[50] Penney Reply, paragraph 34.
[51] Section 3.2 of the article Marthews, Alex, and Tucker, Catherine, "Government Surveillance and Internet Search Behavior," February 17, 2017, found at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564 .
[52] See p. 38 in Marthews and Tucker for lack of statistical significance. For my discussion of omitted variable bias, see Salzberg February Declaration, paragraphs 55-60.
[53] Salzberg February Declaration, paragraphs 55-60.
[54] Penney Reply, paragraph 30a.
[55] Salzberg February Declaration, paragraph 57.
[56] Deposition of Wikimedia designee, James Alexander, April 12, 2018, p. 145.
[57] Penney Reply, paragraph 30e.
[58] Barnett, Arnold I., Applied Statistics: Models and Intuition, Dynamic Ideas LLC, 2015, p. 582.

control for confounding factors, but this is not correct for article-specific factors and is not true when the comparator data is not comparable, as I showed in my February Declaration.[59]

43. In some of the re-analyses in the Penney Reply, articles or time periods are deleted and the Penney Model is re-run.[60] It may be that the Penney Reply does this to address particular examples of the data not following the Penney Model. However, as I said above, these re-analyses do not support the results any more than the original analysis in the Penney December Declaration. Furthermore, by deleting data that tends to disprove the Penney Model and then re-running that data rigs the results toward adoption of the flawed Penney Model.

44. The Penney Reply seems to misinterpret my remarks concerning the staleness of a 2011 DHS list.[61] I was not commenting on the objective nature of the selection, but rather that any list gets stale over time, and the list here used is no exception. For that reason, the static list has no mechanism to update the key articles and therefore a natural decline occurs. The same was not true for the comparator list of popular articles. Because the determination of which articles were popular was made after the time period studied in the Penney December Declaration, articles such as Deaths in 2014 -- which had virtually no page views in 2012 -- were part of the list.[62] On the other hand, a group like ISIL/ISIS, which gained prominence in 2014, was not on the 2011 list, as I pointed out.[63]

F. The Penney Model's Failure to Isolate the Effect of Public Awareness about the NSA Upstream Program

45. My sixth critique, discussed in my February declaration, is that "there are no data or statistical analysis offered that indicate such an effect [an abrupt decline in page views] was due to awareness of the specific NSA program at issue here rather than other related or unrelated events of June 2013."[64] The Penney Reply acknowledges that "in any study of naturalistic changes in human behavior, it will not be possible to isolate the source of all causes and effects on behavior" and that my critique is "a general observation about a *[sic]* naturalistic studies."[65] While this is correct, the Penney December Declaration analysis does not adjust for *any* of those causes, even the obvious ones like seasonality that affect summer page views.

46. Furthermore, the fact that the Penney Model may have been doomed from the start in terms of isolating the effect it intended to prove is not a reason for accepting the model; rather, it is a reason for rejecting it. Despite the passage of nearly six years since the Snowden disclosures, the Penney Reply does not cite a single study published in a peer-reviewed scientific journal that demonstrates the particular effect or even any chilling effect on Internet usage due to awareness of the actual operation of NSA programs.

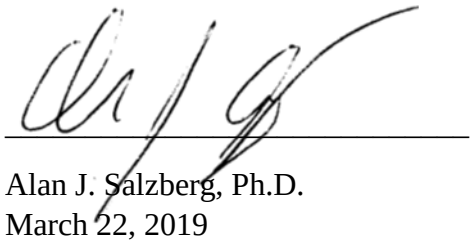
---

[59] Salzberg February Declaration, paragraph 32-46.
[60] In Figure 3b and its explanation in the Penney Reply, the key month of April 2013 is deleted. In Figure 4 and accompanying analyses in the Penney Reply, eight of the original 31 high-privacy articles are deleted for reasons that are unclear to me and unstated in the Penney Reply.
[61] Penney Reply, paragraphs 31 and 32, refer to this critique.
[62] Salzberg Paragraph 64 and database showing 26 most popular articles, which accompanied the Penney December Declaration.
[63] Salzberg Declaration, paragraph 63.
[64] Salzberg Declaration, paragraph 66.
[65] Penney Reply paragraph 36a and 36c, for the first and second quoted material, respectively.

IV. Conclusion

47. In conclusion, my original critiques, detailed in my February Declaration are unchanged by the Penney Reply. In short, the analysis in the Penney December Declaration and the Penney Reply fail to show that public awareness of the Snowden revelations regarding the NSA Upstream program caused any drop in page views of Wikipedia articles.

I declare of penalty of perjury that the foregoing is true and correct to the best of my knowledge and belief. Executed in New York, New York on March 22, 2019.



Alan J. Salzberg, Ph.D.
March 22, 2019

Appendix: Stata Program Log

The following log shows the results of the analysis I performed and described in this declaration. The program was run using Stata, Version 14.

```
      name:  <unnamed>
       log:
D:\clients_2018\DOJ_Wiki_NSA\programsdata\penneyreply\regression_effects_20190318.log
  log type:  text
 opened on:  19 Mar 2019, 11:01:40

. clear

.
. insheet using orig48long.csv
(23 vars, 3,504 obs)

. drop if artnames=="Hamas"
(73 observations deleted)

. save orig48long, replace
file orig48long.dta saved

. keep if monthindex<=32
(1,927 observations deleted)

. save orig48long32, replace
file orig48long32.dta saved

.
. *
. * Simulation that shows no difference in agg v. disagg if same model is run and
issue is just noise
. *
. use orig48long32, clear

. drop if artnames=="Hamas"
(0 observations deleted)

. drop if monthindex>32
(0 observations deleted)

. * run regression to get forecast error
. * no need to show output (but will show output of this for a different purpose
below)
. regress pageviews monthindex intervention postslope, noheader notable

. predict pviewmont
(option xb assumed; fitted values)

. predict sf, stdf

. * simulate data with same forecast error and run regression on disagg
. sort artnum monthindex

. isid artnum monthindex

. * set rndnum seed so can be replicated
. set seed 20190318
```

```
. gen errsim=rnormal(0,sf)

. replace pviewmont=pviewmont+errsim
(1,504 real changes made)

. regress pviewmont monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =     1,504
-------------+----------------------------------   F(3, 1500)      =      3.35
       Model |  6.2583e+10         3  2.0861e+10   Prob > F        =    0.0185
    Residual |  9.3476e+12     1,500  6.2317e+09   R-squared       =    0.0067
-------------+----------------------------------   Adj R-squared   =    0.0047
       Total |  9.4102e+12     1,503  6.2609e+09   Root MSE        =     78941

------------------------------------------------------------------------------
   pviewmont |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   568.2692   570.0658     1.00   0.319    -549.9417     1686.48
intervention |  -11618.79   8230.636    -1.41   0.158    -27763.56    4525.991
   postslope |  -1155.076    893.594    -1.29   0.196    -2907.903      597.75
       _cons |   51521.53   5841.437     8.82   0.000     40063.28    62979.78
------------------------------------------------------------------------------

. * now aggregate, and see that regression standard errors and pvalues are about the
same
. * coeffcients are exactly the same except for rounding because they do not depend on
simulation
. * the Root mean square error is about rmse of disagg model * sqrt(47), or about 7
times as high as mean
. collapse (mean) pviewmont , by( monthindex intervention postslope)

. regress pviewmont monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      3.18
       Model |  1.3316e+09         3   443853226   Prob > F        =    0.0392
    Residual |  3.9062e+09        28   139508526   R-squared       =    0.2542
-------------+----------------------------------   Adj R-squared   =    0.1743
       Total |  5.2378e+09        31   168961239   Root MSE        =     11811

------------------------------------------------------------------------------
   pviewmont |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   568.2691   584.7501     0.97   0.339    -629.5372    1766.075
intervention |  -11618.78   8442.648    -1.38   0.180    -28912.77    5675.196
   postslope |  -1155.076   916.6119    -1.26   0.218    -3032.671    722.5181
       _cons |   51521.53   5991.905     8.60   0.000     39247.67    63795.39
------------------------------------------------------------------------------


. *
. * END Simulation
. *
.
.
. use orig48long32, clear

. drop if artnames=="Hamas"
(0 observations deleted)

. *
. * large changes in standard errors and stat. sign. with removal of a single
observation is another sign of a poor model
```

```
. *
. preserve

. keep if highprivind==1
(512 observations deleted)

. collapse (median) pageviews, by( monthindex intervention postslope highpriv)

. regress pageviews monthindex intervention postslope if highpriv==1

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      2.98
       Model |  13595332.7         3  4531777.56   Prob > F        =    0.0482
    Residual |  42544868.8        28   1519459.6   R-squared       =    0.2422
-------------+----------------------------------   Adj R-squared   =    0.1610
       Total |  56140201.5        31  1810974.24   Root MSE        =    1232.7

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   123.6005   61.02594     2.03   0.052    -1.405487    248.6065
intervention |  -1336.267   881.0953    -1.52   0.141    -3141.109    468.5747
   postslope |  -189.8362   95.65985    -1.98   0.057    -385.7865    6.114118
       _cons |   6285.478   625.3298    10.05   0.000     5004.548    7566.408
------------------------------------------------------------------------------

. restore

. preserve

. * possible error since recruitment and fundamentalism have exact same page views
nearly every month
. * thus show results without as well as with
. drop if artnames=="Recruitment" | artnames=="Fundamentalism"
(64 observations deleted)

. keep if highprivind==1
(480 observations deleted)

. collapse (median) pageviews, by( monthindex intervention postslope highpriv)

. regress pageviews monthindex intervention postslope if highpriv==1

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      4.55
       Model |  9185572.85         3  3061857.62   Prob > F        =    0.0102
    Residual |  18850621.1        28   673236.47   R-squared       =    0.3276
-------------+----------------------------------   Adj R-squared   =    0.2556
       Total |    28036194        31  904393.355   Root MSE        =    820.51

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   93.56127   40.62129     2.30   0.029     10.35233    176.7702
intervention |  -1379.924    586.492    -2.35   0.026    -2581.298   -178.5493
   postslope |  -117.9791     63.675    -1.85   0.074    -248.4115    12.45319
       _cons |   6070.125   416.2444    14.58   0.000     5217.487    6922.763
------------------------------------------------------------------------------

. restore

. preserve
```

```
. collapse (median) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =      8.18
       Model |  84545042.9         3    28181681   Prob > F        =    0.0005
    Residual |    96510363        28  3446798.68   R-squared       =    0.4670
-------------+----------------------------------   Adj R-squared   =    0.4098
       Total |   181055406        31  5840496.96   Root MSE        =    1856.6

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   374.8799    91.9132     4.08   0.000     186.6043    563.1556
intervention |  -3299.076   1327.047    -2.49   0.019    -6017.408   -580.7433
   postslope |  -535.3763   144.0765    -3.72   0.001    -830.5036    -240.249
       _cons |   9601.022     941.83    10.19   0.000     7671.771    11530.27
------------------------------------------------------------------------------

. restore

. * now without possibly error data
. drop if artnames=="Fundamentalism" | artnames=="Recruitment"
(64 observations deleted)

. collapse (median) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =     17.36
       Model |  72354244.3         3  24118081.4   Prob > F        =    0.0000
    Residual |  38905201.2        28  1389471.47   R-squared       =    0.6503
-------------+----------------------------------   Adj R-squared   =    0.6129
       Total |   111259446        31  3589014.37   Root MSE        =    1178.8

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   315.9363   58.35724     5.41   0.000     196.3969    435.4757
intervention |  -4298.331   842.5644    -5.10   0.000    -6024.246   -2572.416
   postslope |  -342.8184   91.47658    -3.75   0.001    -530.1997   -155.4371
       _cons |   8841.338   597.9838    14.79   0.000     7616.424    10066.25
------------------------------------------------------------------------------

. *
. * demonstrate that errors are correlated with articles, meaning disggregation or
some type of adjustment is needed
. * Also shows that stat significance does not exist for overall data
. use orig48long32, clear

. regress pageviews monthindex intervention postslope if artnames!="Fundamentalism" &
artnames!="Recruitment"

      Source |       SS           df       MS      Number of obs   =     1,440
-------------+----------------------------------   F(3, 1436)      =      3.37
       Model |  6.7546e+10         3  2.2515e+10   Prob > F        =    0.0178
    Residual |  9.5866e+12     1,436  6.6759e+09   R-squared       =    0.0070
-------------+----------------------------------   Adj R-squared   =    0.0049
       Total |  9.6541e+12     1,439  6.7089e+09   Root MSE        =     81706

------------------------------------------------------------------------------
```

```
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   752.6646   603.0018     1.25   0.212    -430.1942    1935.523
intervention |  -14970.22   8706.167    -1.72   0.086    -32048.39    2107.947
   postslope |  -1179.932   945.2219    -1.25   0.212    -3034.096    674.2313
       _cons |   49658.62    6178.93     8.04   0.000     37537.93    61779.32
------------------------------------------------------------------------------

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =     1,504
-------------+----------------------------------   F(3, 1500)      =      3.86
       Model |  7.4228e+10         3  2.4743e+10   Prob > F        =    0.0091
    Residual |  9.6056e+12     1,500  6.4037e+09   R-squared       =    0.0077
-------------+----------------------------------   Adj R-squared   =    0.0057
       Total |  9.6798e+12     1,503  6.4403e+09   Root MSE        =     80023

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   881.2874   577.8797     1.53   0.127    -252.2505    2014.825
intervention |  -14757.81   8343.452    -1.77   0.077    -31123.88    1608.265
   postslope |  -1436.449   905.8423    -1.59   0.113    -3213.301    340.4031
       _cons |   48705.37   5921.504     8.23   0.000     37090.07    60320.68
------------------------------------------------------------------------------

. predict residual1, residual

. * stat sign correlation between articles and residuals mean model is insufficient
(see p-value and f-statistic)
. anova residual1 artnum

                          Number of obs =      1,504    R-squared     =  0.9258
                          Root MSE      =    22124.8    Adj R-squared =  0.9234

                  Source | Partial SS         df         MS        F    Prob>F
              -----------+----------------------------------------------------
                   Model |  8.892e+12         46   1.933e+11    394.91  0.0000
                         |
                  artnum |  8.892e+12         46   1.933e+11    394.91  0.0000
                         |
                Residual |  7.132e+11      1,457   4.895e+08
              -----------+----------------------------------------------------
                   Total |  9.606e+12      1,503   6.391e+09

. * note same coefficients in agg results
. collapse (mean) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =     24.85
       Model |  1.5793e+09         3   526437311   Prob > F        =    0.0000
    Residual |   593272771        28  21188313.2   R-squared       =    0.7269
-------------+----------------------------------   Adj R-squared   =    0.6977
       Total |  2.1726e+09        31  70083377.5   Root MSE        =    4603.1

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   881.2874   227.8862     3.87   0.001     414.4836    1348.091
intervention |  -14757.81   3290.232    -4.49   0.000    -21497.54   -8018.073
   postslope |  -1436.449    357.218    -4.02   0.000    -2168.177    -704.721
```

```
       _cons |   48705.37   2335.139    20.86   0.000     43922.06    53488.69
------------------------------------------------------------------------------

.
. *
. * show art is also stat sign for 31 high privacy
. use orig48long32, clear

. keep if highpriv==1
(512 observations deleted)

. regress pageviews monthindex intervention postslope if artnames!="Fundamentalism" &
artnames!="Recruitment"

      Source |       SS           df       MS      Number of obs   =       960
-------------+----------------------------------   F(3, 956)       =      4.20
       Model |  1.3198e+10         3  4.3994e+09   Prob > F        =    0.0058
    Residual |  1.0017e+12       956  1.0478e+09   R-squared       =    0.0130
-------------+----------------------------------   Adj R-squared   =    0.0099
       Total |  1.0149e+12       959  1.0582e+09   Root MSE        =     32369

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   823.6276   292.5762     2.82   0.005     249.4618    1397.793
intervention |  -8112.912   4224.229    -1.92   0.055    -16402.74    176.9203
   postslope |  -1145.897   458.6213    -2.50   0.013    -2045.918   -245.8766
       _cons |    14796.3   2998.014     4.94   0.000     8912.854    20679.75
------------------------------------------------------------------------------

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =       992
-------------+----------------------------------   F(3, 988)       =      5.18
       Model |  1.6582e+10         3  5.5273e+09   Prob > F        =    0.0015
    Residual |  1.0532e+12       988  1.0660e+09   R-squared       =    0.0155
-------------+----------------------------------   Adj R-squared   =    0.0125
       Total |  1.0698e+12       991  1.0795e+09   Root MSE        =     32650

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   918.8429   290.3169     3.16   0.002     349.1343    1488.552
intervention |  -8179.243   4191.609    -1.95   0.051    -16404.72    46.23536
   postslope |  -1340.458   455.0798    -2.95   0.003    -2233.492   -447.4243
       _cons |   15198.27   2974.863     5.11   0.000     9360.491    21036.04
------------------------------------------------------------------------------

. predict residual1, residual

. * stat sign correlation between articles and residuals mean model is insufficient
(see p-value and f-statistic)
. anova residual1 artnum

                          Number of obs =        992    R-squared     =  0.8558
                          Root MSE      =    12570.6    Adj R-squared =  0.8513

                  Source | Partial SS         df         MS        F    Prob>F
              -----------+----------------------------------------------------
                   Model |  9.014e+11         30   3.005e+10    190.14  0.0000
                         |
                  artnum |  9.014e+11         30   3.005e+10    190.14  0.0000
                         |
```

```
    Residual |  1.519e+11        961   1.580e+08
-------------+------------------------------------------------------------
       Total |  1.053e+12        991   1.063e+09

. * note same coefficients in agg results
. collapse (mean) pageviews, by( monthindex intervention postslope)

. regress pageviews monthindex intervention postslope

      Source |       SS           df       MS      Number of obs   =        32
-------------+----------------------------------   F(3, 28)        =     20.87
       Model |   534899840         3   178299947   Prob > F        =    0.0000
    Residual |   239215204        28  8543400.16   R-squared       =    0.6910
-------------+----------------------------------   Adj R-squared   =    0.6579
       Total |   774115045        31  24971453.1   Root MSE        =    2922.9

------------------------------------------------------------------------------
   pageviews |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  monthindex |   918.8429   144.7056     6.35   0.000     622.4269    1215.259
intervention |  -8179.243   2089.266    -3.91   0.001    -12458.91   -3899.577
   postslope |  -1340.458     226.83    -5.91   0.000    -1805.099   -875.8181
       _cons |   15198.27   1482.791    10.25   0.000     12160.91    18235.63
------------------------------------------------------------------------------

.
. log close
      name:  <unnamed>
       log:
D:\clients_2018\DOJ_Wiki_NSA\programsdata\penneyreply\regression_effects_20190318.log
  log type:  text
 closed on:  19 Mar 2019, 11:01:40
```

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION,	)	
Plaintiff,	)	
	)	
v.	)	**Case No. 1:15-cv-662**
	)	
NATIONAL SECURITY AGENCY/	)	
CENTRAL SECURITY SERVICE, *et*	)	
***al.*,**	)	
Defendants.	)	

MEMORANDUM OPINION

Plaintiff, Wikimedia Foundation ("Wikimedia"),[1] challenges the legality of the National Security Agency's ("NSA") Upstream surveillance data gathering efforts, one of a series of recent cases challenging the constitutionality of the NSA's surveillance programs.[2] According to the Director of National Intelligence ("DNI"), Upstream surveillance is a surveillance program authorized pursuant to § 702 of the Foreign Intelligence Surveillance Act ("FISA") that involves the targeted collection of non-U.S. persons' international Internet communications by the NSA.[3] Wikimedia alleges that the NSA has intercepted, copied, and collected Wikimedia's Internet

---

[1] This action was originally brought by nine organizations, including Wikimedia, that communicate over the Internet. The other eight organizations were dismissed at the threshold because those organizations lacked Article III standing. *See Wikimedia Found. v. Nat'l Sec. Agency*, 857 F.3d 193, 216–17 (4th Cir. 2017) (affirming in part *Wikimedia Found. v. Nat'l Sec. Agency*, 143 F. Supp. 3d 344 (D. Md. 2015)).

[2] *See Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1144 (2013) (involving a facial challenge to Section 702 of the Foreign Intelligence Surveillance Act); *Obama v. Klayman*, 800 F.3d 559 (D.C. Cir. 2015) (involving a challenge to the NSA's bulk collection of telephone metadata produced by telephone companies); *Am. Civil Liberties Union v. Clapper*, 785 F.3d 787 (2d Cir. 2015) (involving a challenge to the NSA's bulk telephone metadata collection program); *Jewel v. Nat'l Sec. Agency*, No. C 08–04373 (N.D. Cal. April 25, 2019), *appeal docketed*, No. 19–16066 (9th Cir. May 21, 2019) (involving a challenge to the NSA's interception of Internet communications); *Schuchardt v. Trump*, 2019 WL 426482 (W.D. Pa. Feb. 4, 2019), *appeal docketed*, No. 19–1366 (3d Cir. Feb. 14, 2019) (involving a challenge to the NSA's interception of Internet communications through the PRISM surveillance program).

[3] *See* Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2.

communications pursuant to the Upstream surveillance program and that such interception, duplication, and collection exceeds the NSA's authority under FISA and violates Wikimedia's rights under the First and Fourth Amendments of the Constitution.

At issue in this matter is defendants' motion for summary judgment. Defendants argue that judgment must be entered in their favor because Wikimedia, the only remaining plaintiff, lacks Article III standing. Defendants also argue that even if a genuine dispute of material fact exists as to Wikimedia's standing, the state secrets doctrine precludes further litigation of Wikimedia's standing, and thus requires entry of judgment in defendants' favor.

Before analyzing the parties' arguments on the issue of Article III standing and the state secrets doctrine, however, it is important to address briefly three topics: (i) the definition of Upstream surveillance and the statutory authority for the NSA's Upstream surveillance program, (ii) the procedural history of this case, and (iii) the undisputed factual record developed by the parties. After addressing these three preliminary topics, which frame all of the analysis that follows, the pertinent summary judgment standard is set forth, and the parties' arguments are analyzed under that standard. For the reasons that follow, Wikimedia has failed to establish that it has Article III standing sufficient to survive summary judgment, and further litigation of this matter is precluded by the state secrets doctrine. Accordingly, this case must be dismissed, and judgment must be entered in favor of defendants.

I.

To begin with, it is necessary to define Upstream surveillance, the NSA program at issue in this litigation, and to clarify what is meant by the term Upstream surveillance as that term is used in this litigation. The NSA conducts Upstream surveillance pursuant to § 702 of FISA, 50 U.S.C. § 1881a. The government has acknowledged that it conducts § 702 surveillance through

two programs, namely the Upstream and PRISM programs.[4] In PRISM surveillance, the government acquires communications directly from a United States-based Internet Service Provider ("ISP"). *See* PCLOB 702 Report, at 33. In contrast, the acquisition of communications via Upstream surveillance does not occur "with the compelled assistance of the United States ISPs, but instead with the compelled assistance…of the providers that control the telecommunications backbone over which communications transit."[5] *Id.* at 35. Thus, Upstream collection, unlike PRISM collection, "does not occur at the local telephone company or email provider with whom the targeted person interacts." *Id.* Instead, the collection of communications for Upstream surveillance "occurs 'upstream' in the flow of communications between communication service providers." *Id.* Only the Upstream surveillance program is at issue in this case.

As noted, the government contends that its Upstream surveillance program is conducted pursuant to FISA § 702. Specifically, § 702 permits the Attorney General and the DNI to authorize jointly, for up to one year, foreign-intelligence surveillance targeted at non-U.S. persons located abroad,[6] if the Foreign Intelligence Surveillance Court ("FISC")[7] approves the government's written certification demonstrating that the intended surveillance complies with statutory

---

[4] *See* Privacy and Civil Liberties Oversight Board, *Report on the Surveillance Program Operated Pursuant to Section 702 of the Foreign Intelligence Surveillance Act* 7 (2014) ("PCLOB 702 Report"), *available at* https://www.pclob.gov/library/702-Report-2.pdf.

[5] The telecommunications or Internet "backbone" is the network of high-capacity fiber-optic cables, switches, and routers operated by telecommunications service providers that facilitates both domestic and international communication via the Internet. This backbone primarily consists of a network of fiber-optic cables, including terrestrial cables that link areas across the U.S. and transoceanic cables that link the U.S. to the rest of the world.

[6] Importantly, the statute expressly prohibits the intentional targeting of any persons known at the time of acquisition to be in the United States or any U.S. person reasonably believed to be located outside the United States. 50 U.S.C. § 1881a(b). Section 702 does allow the government, however, to intercept communications between a U.S. person inside the United States and a foreigner located abroad who has been targeted by intelligence officials. *See id.* § 1881a(a)–(b).

[7] FISC, a tribunal composed of eleven federal judges designated by the Chief Justice of the U.S. Supreme Court, is charged with the review of applications for electronic surveillance. *See* 50 U.S.C. § 1803(a).

requirements.[8] To approve such a certification, the FISC must determine that the government's targeting procedures are reasonably designed:

> (i) to ensure that acquisition "is limited to targeting persons reasonably believed to be located outside the United States," 50 U.S.C. § 1881a(j)(2)(B)(i);
>
> (ii) to prevent the intentional acquisition of wholly domestic communications, *id.* § 1881a(j)(2)(B)(ii);
>
> (iii) to "minimize the acquisition and retention, and prohibit the dissemination, of nonpublicly available information concerning unconsenting United States persons consistent with the need of the United States to obtain, produce, and disseminate foreign-intelligence information," *id.* § 1801(h)(1); *see id.* § 1881a(j)(2)(C); and
>
> (iv) to ensure that the procedures "are consistent with…the [F]ourth [A]mendment," *id* § 1881a(j)(3)(A).[9]

In effect, FISC approval of government surveillance pursuant to § 702 means that the FISC has found that the surveillance comports with the statutory requirements and the Constitution.

The recent release of public reports and declassification of some FISC opinions have revealed additional details regarding the collection of communications pursuant to § 702. After the FISC approves a § 702 certification, the NSA designates "targets," which are non-U.S. persons located outside the United States who are reasonably believed to possess or receive, or are likely to communicate, foreign-intelligence information designated in the certification.[10] The NSA then attempts to identify "selectors," namely the specific means by which the targets communicate,

---

[8] The government must certify that a significant purpose of the acquisition is to obtain foreign intelligence information and that the acquisition will be conducted in a manner consistent with the Fourth Amendment and the targeting and minimization procedures required by statute. 50 U.S.C. § 1881a(b), (g).

[9] In addition, following the passage of the FISA Amendments Reauthorization Act of 2017, the FISC must now also find that the government's querying procedures meet the statutory requirements and are consistent with the Fourth Amendment. *Id.* § 1881a(j)(2)(D); (j)(3)(A). These provisions have been cited to the version of § 1881a in effect since January 18, 2018. All of these provisions are identical to those in the version of § 1881a effective between June 2, 2015 and January 18, 2018, but the provisions are now located within § 1881a(j) rather than § 1881a(i).

[10] PCLOB 702 Report, at 41–46.

such as email addresses or telephone numbers.[11] Importantly, selectors cannot be key words (*e.g.*, "bomb") or targets' names (*e.g.*, "Bin Laden"); rather, selectors must be specific communication identifiers.[12] The government then may issue a § 702 directive to a U.S. telecommunications service provider requiring it to assist the government in acquiring communications involving those selectors.[13]

As for the actual collection of communications containing these targeted selectors, the government has described the Upstream surveillance collection process as follows:

> [C]ertain Internet transactions transiting the Internet backbone network(s) of certain electronic communication service provider(s) are filtered for the purpose of excluding wholly domestic communications[,] and are then scanned to identify for acquisition those transactions [that contain communications] to or from . . . persons targeted in accordance with the applicable NSA targeting procedures; only those transactions that pass through both the filtering and the scanning are ingested into Government databases.

Defs.' Br. 4 (quoting Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2).[14] Thus, the Upstream surveillance collection process involves three steps—(1) filtering, (2) scanning, and (3) ingesting. As this description shows, although the government has disclosed some information about Upstream surveillance in declassified documents and

---

[11] NSA Director of Civil Liberties and Privacy Office Report, *NSA's Implementation of FISA Section 702* 4 (2014), *available at* https://www.nsa.gov/Portals/70/documents/news-features/press-room/statements/NSAImplementationofFISA70216Apr2014.pdf.

[12] *Id.*; PCLOB 702 Report, at 32–33, 36.

[13] 50 U.S.C. § 1881a(i); PCLOB 702 Report, at 32–33.

[14] Prior to April 2017, Upstream collection included Internet communications "that were to, from ***or about*** (i.e., containing a reference in the communication's text to) a selector tasked for acquisition under Section 702." FISC Mem. Op. & Order, at 16 (April 26, 2017) (emphasis in original), *available at* https://www.dni.gov/files/documents/icotr/51117/2016_Cert_FISC_Memo_Opin_Order_Apr_2017.pdf. According to the PCLOB 702 Report, under the Upstream surveillance program that included "about" collection, "a communication between two third parties might be acquired because it contains a targeted email address in the body of the communication." PCLOB 702 Report, at 119. As of March 2017, however, the NSA ceased "about" collection entirely, which a FISC judge concluded "should substantially reduce the acquisition of non-pertinent information concerning U.S. persons pursuant to Section 702." FISC Mem. Op. & Order, at 23, 25 (April 16, 2017).

unclassified reports, most technical details of the Upstream surveillance process remain classified. *Wikimedia Found. v. Nat'l Sec. Agency*, 857 F.3d 193, 202 (4th Cir. 2017) (citing *Jewel v. Nat'l Sec. Agency*, 810 F.3d 622, 627 (9th Cir. 2015)).

II.

With this statutory framework and definition of Upstream surveillance in mind, it is appropriate to turn to the procedural history of this case. On June 22, 2015, Wikimedia, along with eight other organizations,[15] filed the Amended Complaint in this suit, challenging the legality of the NSA's Upstream surveillance program. The Amended Complaint alleges that Upstream surveillance (i) exceeds the scope of the government's authority under § 702, (ii) violates Article III, (iii) violates the First Amendment, and (iv) violates the Fourth Amendment and requests (i) a declaration that Upstream surveillance violates the Constitution and § 702 and (ii) an order permanently enjoining the NSA from conducting Upstream surveillance. On August 6, 2015, defendants moved to dismiss the Amended Complaint, arguing that plaintiffs lacked Article III standing. On October 23, 2015, defendants' motion was granted on the ground that plaintiffs' allegations were too speculative to establish Article III standing. *Wikimedia Found. v. Nat'l Sec. Agency,* 143 F. Supp. 3d 344, 356 (D. Md. 2015), *aff'd in part, vacated in part, and remanded by*, 857 F.3d 193 (4th Cir. 2017).

Thereafter, plaintiffs appealed, and the Fourth Circuit affirmed in part, vacated in part, and remanded the case for further consideration. *Wikimedia Found.,* 857 F.3d at 200. Specifically, the Fourth Circuit vacated the finding that Wikimedia lacked standing, but affirmed the finding that the other plaintiffs lacked standing. *Id.* The Fourth Circuit concluded that Wikimedia had

[15] These original plaintiffs included the National Association of Criminal Defense Lawyers, Human Rights Watch, Amnesty International USA, Pen American Center, Global Fund for Women, the Nation magazine, the Rutherford Institute, and the Washington Office on Latin America.

established standing sufficient to survive a facial challenge to the Amended Complaint based on the "Wikimedia Allegation", namely the allegation "that the sheer volume of [Wikimedia's] communications makes it virtually certain that the NSA has intercepted, copied, and reviewed at least some of [Wikimedia's] communications[,]" "even if the NSA conducts Upstream surveillance on only a single [I]nternet [backbone] link." *Id.* at 202, 209 (internal quotation marks and citation omitted). Three factual allegations, accepted as true as required at the motion to dismiss stage, made the Wikimedia Allegation plausible: (i) "Wikimedia's communications almost certainly traverse every international [Internet] backbone link connecting the United States with the rest of the world[,]" (ii) "the NSA has confirmed that it conducts Upstream surveillance at more than one point along the [I]nternet backbone[,]" and (iii) "the government, for technical reasons[,] . . . must be copying and reviewing all the international text-based communications that travel across a given [Internet backbone] link upon which it has installed surveillance equipment." *Id.* at 210–11 (internal quotation marks and citations omitted).

Importantly, the Fourth Circuit rejected the "Dragnet Allegation", that is the allegation "that[,] in the course of conducting Upstream surveillance[,] the NSA is intercepting, copying, and reviewing substantially all text-based communications entering and leaving the United States, including" those of the nine plaintiffs. *Id.* at 202 (internal quotation marks and citation omitted). Plaintiffs alleged the following facts in support of the Dragnet Allegation: (i) "the NSA has a strong incentive to intercept communications at as many [Internet] backbone chokepoints as possible, and indeed must be doing so at many different [Internet] backbone chokepoints," (ii) "the technical rules governing online communications make this conclusion especially true," and (iii) "a *New York Times* article asserts that the NSA is temporarily copying and then sifting through the contents of what is apparently most e-mails and other text-based communications that cross the

[U.S.] border." *Id.* at 213 (internal quotation marks and citations omitted). The Fourth Circuit concluded that the Dragnet Allegation failed to establish standing because it did "not contain enough well-pleaded facts entitled to the presumption of truth." *Id.* at 200. As such, although Wikimedia pled sufficient facts to establish standing at the motion to dismiss stage, the other plaintiffs did not. *Id.* at 200. Thus, Wikimedia is the only remaining plaintiff.

On remand, an Order issued on October 3, 2017 directing the parties to conduct a limited five-month period of jurisdictional discovery. *See* ECF Nos. 117, 123. Both sides took depositions and served requests for written discovery and production of documents. Defendants objected to 53 of Wikimedia's 84 discovery requests on the ground that responses to the requests would reveal classified information protected by the common law state secrets privilege and related statutory privileges. Thereafter, the DNI formally asserted the state secrets privilege and the statutory privilege set forth in 50 U.S.C. § 3024(i)(1).[16] Defendants stated that the information Wikimedia sought, if disclosed, reasonably could be expected to result in exceptionally grave damage to U.S. national security.[17] Wikimedia subsequently moved to compel production of the documents. On August 20, 2018, an Order and Memorandum Opinion issued, concluding that defendants satisfied the procedural requirements necessary to invoke the state secrets privilege, that the information sought to be protected qualified as privileged under

---

[16] Defendants also submitted a classified declaration from George C. Barnes, the Deputy Director of the NSA. The classified declaration provided additional detail about the harm to national security that would be caused by disclosure of the information contained in Wikimedia's discovery requests.

[17] The DNI's and the NSA's assertions of privilege encompassed seven categories of information: (i) individuals or entities subject to Upstream surveillance; (ii) operational details of the Upstream collection process such as the technical details concerning methods, processes, and devices employed (including the design, operation, and capabilities of the devices); (iii) locations (and nature of the locations) at which Upstream surveillance is conducted; (iv) the specific types or categories of communications either subject to or acquired in the course of the Upstream collection process; (v) the scope and scale on which Upstream collection has or is now being conducted; (vi) the NSA's cryptanalytic capabilities or limitations; and (vii) additional categories of classified information encompassed within numerous FISC opinions and orders. *See* DNI Decl. ¶¶ 18, 21–47.

the state secrets doctrine, and that therefore, Wikimedia's motion to compel must be denied. *Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 790 (D. Md. 2018). Accordingly, the parties continued jurisdictional discovery, limited to information not protected by the state secrets privilege.

Defendants now seek summary judgment on the ground that Wikimedia lacks Article III standing to contest the legality of the NSA's Upstream surveillance program, or alternatively, that if there is a genuine issue of material fact as to the three essential elements of the Wikimedia Allegation articulated in the Fourth Circuit's remand order, the state secrets doctrine operates to preclude further litigation of Wikimedia's standing and thus requires entry of judgment in defendants' favor.

III.

Summary judgment is appropriate only where there are no genuine disputes of material fact. Rule 56, Fed. R. Civ. P. Accordingly, the material facts as to which no genuine dispute exists must first be identified. Defendants set out their statement of material facts in their brief in support of summary judgment, as required by the local rules. Plaintiff, in addition to responding to defendants' statement of material facts as required by the local rules, also offered their own separate statement of material facts in their brief in opposition to summary judgment. Neither the local rules of the District of Maryland nor the Eastern District of Virginia require plaintiff, as the non-moving party, to set forth a statement of material facts. *See generally* D. Md. Local Rules; E.D. Va. Local Civ. R. 56(B). In the interest of completeness, however, and because each party has responded to the other party's statement of material facts, all facts, and disputes as to those facts, have been considered in deriving from the record the following undisputed material facts.

1. The Internet is a global collection of networks, large and small, interconnected by

a set of routers.[18] Together, these large and small networks function as a single, large virtual network, on which any device connected to the network can communicate with any other connected device.

2. To communicate over the Internet, an individual user connects with the network of a local Internet Service Provider ("ISP"), either directly (typically for a monthly fee) or indirectly through an organization (*e.g.*, a place of business, an Internet café). In turn, the local ISP's network connects to the networks of larger regional and national ISPs, the largest of which are called "Tier 1" telecommunication service providers (*e.g.*, AT&T, CenturyLink, Cogent, Verizon).

3. Tier 1 providers and other large carriers maintain high-capacity terrestrial fiber-optic networks, known generally as Internet "backbone" networks, that use long-haul terrestrial cables to link large metropolitan areas across a nation or region. Data travel across these cables in the form of optical signals, or pulses of light.

4. The Internet backbone also includes transoceanic cables linking North and South America with each other and with Europe, Asia, the Middle East, and Africa. These undersea cables reach shore at points known as cable landing stations, from which they are linked to the terrestrial telecommunications network.

5. Tier 1 providers and other large carriers typically connect separate legs of their own networks using high-capacity switches. To allow users of different providers' networks to communicate with one another, Tier 1 providers and other large carriers typically interconnect their networks using high-capacity routers.[19]

6. Generally speaking, to send a communication on the Internet, the transmitting device (*e.g.*, a personal computer, a cell phone) first converts the communication into one or more small bundles of data called "packets," configured according to globally accepted protocols.[20]

7. When a communication is broken into separate packets, each packet includes (i) a "header," which consists of the routing, addressing, and other technical information required to facilitate the packets' travel from its source to its intended

---

[18] Routers are specialized computers that ensure that Internet communications travel an appropriate path across the Internet. Routers serve a similar role for the Internet as switches (or switchboards) do on the telephone network.

[19] Routers and switches perform similar functions, namely directing the transport of Internet communications across the network. Routers generally connect one communications service provider's network to a different communications service provider's network, whereas switches generally connect a single communication provider's network.

[20] Protocols can be thought of as electronic languages. Each protocol, or language, has its own rules and vocabulary. For example, instead of English and Spanish, there is Transmission Control Protocol ("TCP") and User Datagram Protocol ("UDP").

destination, and (ii) a "payload," which consists of a portion of the contents of the communication being transmitted.

8. A packet's header contains three relevant pieces of address and routing information: (i) the packet's source and destination Internet Protocol ("IP") addresses; (ii) the source and destination ports; and (iii) protocol numbers.

9. IP addresses, which are included in packet headers, are unique numeric identifiers assigned to particular computers, devices, or systems connected to the Internet.[21] IP addresses are used to direct data back and forth between one computer (or other online device) and another online device. IP addresses may be analogized to the destination and return addresses on a mailing envelope.

10. The IP addresses of entities with a large, fixed presence on the Internet do not change and are publicly accessible.[22]

11. Port numbers, which are also included in packet headers, are used to identify communications of different kinds (*e.g.*, webpage requests, or email) so that servers hosting multiple communications services (*e.g.*, a website and an email service) can distinguish packets destined for one service from those meant for another. Port numbers for common applications, like web-browsing and email, are assigned in a common industry registry maintained by the IANA. Whereas IP addresses can be analogized to the street address on a letter, port numbers are roughly analogous to the apartment numbers at a multi-unit dwelling.

12. Protocol numbers, which are also included in packet headers, are used by receiving devices to determine the appropriate method of interpreting data (*e.g.*, HTTP, TCP/IP). A protocol defines the actions taken upon the transmission and/or receipt of a message or other transmission. Protocols are also assigned numbers maintained in a common industry registry maintained by the IANA.

---

[21] There are circumstances, however, in which IP addresses do not uniquely identify individual Internet users. For example, residential Internet customers ordinarily get exactly one "dynamic" IP address at a time, which is assigned on a temporary basis by their ISP. Dynamic IP addresses may be assigned for a day, an hour, or some other period of time depending on the needs, resources, and business practices of a particular ISP, after which the dynamic IP addresses are assigned to other customers. Thus, although the IP addresses of business customers of ISPs almost never change, the IP addresses of individual ISP customers can change fairly often, with the same IP address subsequently being assigned to a different customer of the ISP. *See* Dr. Henning Schulzrinne Decl. ¶¶ 30, 33-34, ECF No. 162-2. As another example, the IP addresses in the packets that make up email messages sent or received by an email server on behalf of its users may have the IP address of the server as the source or destination IP address, not an IP address associated with the individual email user. In other words, the IP address in packets transmitting email messages might be the IP address of the email server (*e.g.*, Gmail, Yahoo), rather than the IP address of the individual user of the email address. Scott Bradner Decl. ¶¶ 244-46, ECF No. 168-2.

[22] Each Internet Service Provider or other large enterprise with a fixed presence on the Internet (*e.g.*, Amazon, Wikimedia) acquires blocks of "static" IP addresses assigned on a permanent basis from the appropriate regional Internet registry affiliated with the global Internet Assigned Numbers Authority ("IANA"). There are public databases that record, with very high accuracy, which address blocks are used by what entities.

13. After a communication has been broken into packets by the transmitting device, specialized computers called routers and switches ensure that the packets travel an appropriate path across the Internet to their destination IP address.

14. Each router or switch through which a packet transits scans the packet's header information, including its destination IP address, and determines which direction (path) the packet should follow next in order to reach its intended destination. The router or switch operates somewhat similarly to Google Maps, updating the fastest route to take between a user's starting point and his or her destination.

15. When packets transmitting a communication arrive at the receiving computer, smartphone, or other online device, the receiving device reassembles the packets into the original communication, such as a webpage or email.

16. Traffic "mirroring" is a technical term for a process by which a router or switch, in addition to determining where on the Internet each packet should be forwarded next, can also identify certain packets to be copied ("mirrored") and divert the designated copies off-network for separate processing. In other words, traffic mirroring can create a copy of all communications, or a subset of all communications, passing through a router or switch without interrupting the flow of those communications.

17. Traffic mirroring is accomplished by programming routers and switches with access control lists ("ACLs") to determine whether packets will be copied and collected at a certain link (the "interface") between the router or switch and another device. The criteria used in the ACL can include a packet's source or destination IP address, the port number, the protocol numbers, or other information contained in a packet header.

18. The router or switch examines the header information of each packet it processes, and compares it to the ACL for each interface, to determine which interfaces the packet may or may not pass through without mirroring (copying).

19. Tier 1 providers and other smaller service providers employ traffic mirroring in the normal course of their operations for such purposes as monitoring traffic load, conducting quality-control processes, and rejecting unwanted traffic.

20. At any link on the Internet where surveillance may be conducted, traffic mirroring with ACLs can be used in several ways to make only certain packets available for inspection by a collecting entity.[23]

---

[23] Plaintiff disputes this fact, as well as facts 22-24, to the extent the "collector" or the "collecting entity" is the NSA conducting Upstream surveillance. These facts, as stated, do not put forth that the "collector" or the "collecting entity" is the NSA. In fact, these facts simply establish that any entity, government or private, trying to collect Internet communications could, *hypothetically*, employ traffic mirroring in this manner. Plaintiff's argument that the

21. To conduct traffic mirroring, an interface (a fiber-optic link) would have to be established between the router or switch directing traffic at the selected location and the separate equipment used by the collecting entity (hereinafter, the "collector interface").

22. After the collector interface is established, communications traffic passing through the carrier's router or switch to the collector's equipment can be filtered by "whitelisting" or "blacklisting" techniques. "Whitelisting" or "blacklisting" involves configuring an ACL to allow only packets meeting the ACL's criteria to be copied and passed through the collector interface to the collector's equipment.

23. For example, the collector could configure an ACL containing a "whitelist" of specific IP addresses of interest. When the router or switch examines the header information of each packet it processes, it would then, (i) as usual, forward a copy of the packet toward its intended destination, (ii) perhaps forward additional copies through other interfaces, per the carrier's routine business practices, and (iii) if, and only if, the packet header contains a source or destination IP address on the whitelist, create an additional copy of the packet, and forward it through the collector interface into the collector's possession and control. In other words, packets containing IP addresses on the whitelist would be copied and sent through to the collector's equipment. Packets not meeting the whitelist criteria would not be copied for, or made available to, the collector's equipment for any purpose.

24. Blacklisting, conversely, involves configuring an ACL to allow all packets to be copied to the collector interface *except* those matching the ACL's criteria. With a blacklist, the router or switch would examine each packet header and (i) as usual, forward a copy of the packet toward its intended destination, (ii) perhaps forward additional copies through other interfaces, per the carrier's routine business practices, and (iii) create an additional copy of every packet and forward it through the collector interface into the collector's possession and control, *except* for those packets with source or destination IP addresses on the blacklist. In other words, if the router or switch finds that a packet header contains a source or destination IP address on the blacklist, an additional copy of that packet is not created or forwarded through the collector interface.

25. Whitelisting and blacklisting techniques can also be used to limit mirroring to particular sources of traffic, such as only cables used by specific carriers, or only cables linked to specific countries or regions.

26. In addition, ACLs can be configured to whitelist or blacklist particular types of communication based on their port or protocol numbers, such as email communications or communications from accessing websites.

NSA does not use traffic mirroring in this way when the NSA conducts Upstream surveillance is discussed at length *infra* Part V.C.

27. Wikimedia operates twelve free-knowledge projects on the Internet, including Wikipedia. Wikipedia, a free-access, free content encyclopedia, is one of the top ten most-visited websites in the world. In 2017, Wikipedia's website received visits from more than 1 billion unique devices each month.

28. Wikimedia engages in more than a trillion international Internet communications each year, with individuals in every country on the planet. This includes communications between foreign users and Wikimedia's U.S.-based servers, and communications between U.S. users and Wikimedia's foreign-based servers.

29. Wikimedia has identified three categories of its international Internet communications that it contends are subjected to Upstream surveillance collection by the NSA: (i) communications with its community members[24] ("Category 1"), (ii) internal "log" communications ("Category 2"), and (iii) the electronic communications of Wikimedia's staff ("Category 3").

30. Category 1 consists of communications with and among Wikimedia's community members, including requests from foreign and domestic users to view or download content from Wikimedia websites, and email communications sent from foreign users to Wikimedia servers.[25] All of these communications were directed to the public IP address ranges assigned to and used by Wikimedia.

31. Category 2 consists of internal log communications transmitted from Wikimedia's servers in the Netherlands to its servers in the United States. These communications are encrypted and received at one of the same public IP address ranges as Wikimedia's communications in Category 1.[26]

32. Category 3 consists of communications by Wikimedia's staff using various protocols, some of which are encrypted, some of which are not. These communications, like those in Categories 1 and 2, are sent and received from the public IP address ranges assigned to and used by Wikimedia.[27]

33. The total volume of Wikimedia's international Internet communications exceeds the number of cables transporting Internet communications between the U.S. and other countries. Moreover, Wikimedia's communications are broadly distributed, with users in every country in the world.

---

[24] Wikimedia community members are people who read or contribute to Wikimedia's twelve free-knowledge projects.

[25] According to Wikimedia, the volume of the email communications in Category 1, and the countries from which those emails are received, are unknown. Defs.' Ex. 4, Pl.'s Am Resps. & Objs. to ODNI Interrog. No. 19, Ex. 1 (hereinafter, "Technical Statistics Chart"), ECF No. 162-5.

[26] Technical Statistics Chart; Schulzrinne Decl. ¶¶ 83-84.

[27] Technical Statistics Chart; Schulzrinne Decl. ¶¶ 85-87.

34. It is "virtually certain" that Wikimedia's communications traverse every cable carrying public Internet traffic that connects the U.S. to other countries.

35. The government has described Upstream surveillance as involving three steps. First, "certain Internet transactions transiting the Internet backbone network(s) of certain electronic communication service provider(s) are filtered for the purpose of excluding wholly domestic communications." Second, these Internet transactions "are then scanned to identify for acquisition those transactions [that contain communications] to or from . . . persons targeted in accordance with the applicable NSA targeting procedures." And third, "those transactions that pass through both the filtering and the scanning are ingested into Government databases."[28]

36. Prior to April 2017, Upstream surveillance involved "about" collection (i.e., a communication containing a reference in the communication's text to a selector tasked for acquisition under § 702). "About" communications were not necessarily sent to or from the user of a § 702 tasked-selector.

37. The statement—the "NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA or is routed through a foreign server"—was accurate as of October 3, 2011.[29]

IV.

Summary judgment is appropriate when there is "no genuine issue as to any material fact" and based on those undisputed facts the moving party "is entitled to judgment as a matter of law." *Celotex Corp. v. Catrett*, 477 U.S. 317, 322 (1986). To serve as a bar to summary judgment, facts must be "material," which means that the disputed fact "might affect the

---

[28] Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2.

[29] R. Richards Dep. at 160:4-17; [Redacted], 2011 WL 10945618, at *15. Defendants' Rule 30(b)(6) witness confirmed the accuracy of this statement as of October 2011. Defendants argue that statements of fact in a judicial opinion, such as this statement from a FISC Opinion, are inadmissible hearsay, and thus, plaintiff cannot rely on such statements at summary judgment. Summary judgement evidence must either be in admissible form or capable of being rendered admissible at trial. *Humphreys & Partners Architects, LP v. Lessard Design, Inc.*, 790 F.3d 532, 538-39 (4th Cir. 2015); Fed. R. civ. P. 56(c)(2). Statements of fact in judicial opinions that are offered for the truth of the matter asserted are hearsay. *Nipper v. Snipes*, 7 F.3d 415, 417-18 (4th Cir. 1993); *see also Zeus Enter., Inc. v. Alphin Aircraft, Inc.*, 190 F.3d 238, 242 (4th Cir. 1999); *Carter v. Burch*, 34 F.3d 257, 265 (4th Cir. 1994). Even though the 2011 FISC Opinion is inadmissible hearsay, defendants' Rule 30(b)(6) witness testimony, confirming the accuracy of this specific statement as of October 3, 2011, is not hearsay. Thus, this statement is admissible, but solely this statement because it is as a statement of a party opponent.

outcome of the suit under the governing law." *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 248 (1986). Where a party "fails to make a showing sufficient to establish the existence of an element essential to that party's case, and on which that party will bear the burden of proof at trial," there can be no genuine issue as to any material fact. *Celotex Corp.* 477 U.S. at 322.

Article III limits the jurisdiction of federal courts to actual "Cases" or "Controversies." *See* U.S. Const. art. III, § 2, cl. 1. As the Supreme Court has made clear, one "essential and unchanging part of the case-or-controversy requirement" is that a plaintiff must establish Article III standing to sue. *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560 (1992). A plaintiff establishes Article III standing by showing that he, she, or it seeks relief from an injury that is "concrete, particularized, and actual or imminent; fairly traceable to the challenged action; and redressable by a favorable ruling." *Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1147 (2013) (quoting *Monsanto Co. v. Geerston Seed Farms*, 561 U.S. 139, 149 (2010)). In other words, a plaintiff must establish (1) an injury-in-fact; (2) a casual connection between the injury and the alleged conduct; and (3) the redressability of the injury by a court.

To establish injury-in-fact, the alleged injury must be "real and immediate," not "conjectural or hypothetical." *City of Lost Angeles v. Lyons*, 461 U.S. 95, 201 (1983). The Supreme Court has "repeatedly reiterated that '[a] threatened injury must be *certainly impending* to constitute injury in fact,' and that '[a]llegations of *possible* future injury' are not sufficient." *Clapper*, 133 S. Ct. at 1147 (quoting *Whitmore v. Arkansas*, 495 U.S. 149, 158 (1990)) (emphases in original). In some cases, injury-in-fact can also be established "based on a 'substantial risk' that the harm will occur, which may prompt plaintiffs to reasonably [sic] incur costs to mitigate or avoid that harm."[30] *Id.* at 1150 n. 5. Importantly, the standing inquiry is

[30] The parties disagree on whether the appropriate standard for determining injury-in-fact sufficient to establish

"especially rigorous when reaching the merits of the dispute would force [a court] to decide whether an action taken by one of the other two branches of the Federal Government was unconstitutional," particularly "in the fields of intelligence gathering and foreign affairs." *Clapper*, 133 S. Ct. at 1147.

Because standing is a threshold jurisdictional requirement, it may be attacked at any time, including at summary judgment. As the Supreme Court has made clear, each element of standing must be supported "in the same way as any other matter on which the plaintiff bears the burden of proof, *i.e.*, with the manner and degree of evidence required at the successive stages of the litigation." *Defenders of Wildlife*, 540 U.S. at 561. Where, as here, standing is challenged at the summary judgment stage, "'the party invoking federal jurisdiction bears the burden of establishing' standing—and…such a party 'can no longer rest on…mere allegations, but must set forth by affidavit or other evidence specific facts'" to establish standing. *Clapper*, 133 S. Ct. at 1148-49 (quoting *Defenders of Wildlife*, 504 U.S. at 561).

Thus, if a plaintiff cannot set forth, by affidavit or other evidence that will be in admissible form at trial, specific facts sufficient to show a genuine issue for trial on standing,

---

standing is a "certainly impending" standard or a "substantial risk" standard in this case. The Supreme Court has not been clear as to whether the "substantial risk" standard applies and whether that standard is distinct from the "certainly impending" requirement in cases such as this that involve government surveillance. *See Clapper*, 133 S. Ct. at 1150 n. 5. But the Supreme Court has "found standing based on a 'substantial risk' that harm will occur" in some cases. *Id.*

The Fourth Circuit has indicated that injury-in-fact may be established under either the "certainly impending" or the "substantial risk" standard, and thus, standing should be analyzed under both standards in some cases. *See Beck v. McDonald*, 848 F.3d 262, 275 (4th Cir. 2017) (after determining that the threatened harm was not "certainly impending," the Fourth Circuit stated "our inquiry on standing is not at an end, for we may also find standing based on a 'substantial risk' that the harm will occur, which in turn may prompt a party to reasonably [sic] incur costs to mitigate of avoid that harm"). Importantly, the "substantial risk" standard does not change "the common-sense notion that a threatened event can be 'reasonabl[y] likel[y]' to occur but still be insufficiently 'imminent' to constitute an injury-in-fact." *Id.* at 276.

In this opinion, both standards are applied. Moreover, the injury-in-fact standard, whether "certainly impending," "substantial risk," or both, does not impact the outcome in this case because under whichever standard applies, litigation of any remaining dispute of material fact as to Wikimedia's Article III standing cannot be further litigated without violating the state secretes doctrine, as further discussed *infra* Part VI.

then Rule 56(c) mandates entry of summary judgment against the plaintiff. *See Celotex Corp.*, 477 U.S. at 322.

V.

At this stage of the litigation, Wikimedia must present specific facts, supported by admissible record evidence, that are sufficient to show a genuine issue for trial on Wikimedia's Article III standing. In other words, Wikimedia must present specific facts which show that defendants, through the Upstream surveillance program, have copied and collected Wikimedia's international Internet communications, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection.[31]

Both parties have focused their discussion of Wikimedia's standing on the three prongs necessary to establish the Wikimedia Allegation,[32] which were enumerated in the Fourth Circuit's remand order in this case. *See Wikimedia Found.*, 857 F.3d at 210–11. The three prongs are: (A) Wikimedia's communications almost certainly traverse every international Internet backbone link connecting the United States with the rest of the world; (B) the NSA conducts Upstream surveillance at one or more points along the Internet backbone; and (C) the NSA, for technical reasons, must be copying and reviewing all the text-based communications that travel across a given Internet backbone link upon which it conducts Upstream surveillance. Together,

---

[31] *See Obama v. Klayman*, 800 F.3d 559, 562 (D.C. Cir. 2015) ("In other words, plaintiffs here must show *their own* metadata was collected by the government.") (emphasis in original); *Halkin v. Helms*, 690 F.2d 977, 999-1000 (D.C. Cir. 1982) ("[T]he absence of proof of actual acquisition of appellants' communications is fatal to their watchlisting claims.").

[32] The Wikimedia Allegation is the allegation that the sheer volume of Wikimedia's communications makes it virtually certain that the NSA has intercepted, copied, and reviewed at least some of Wikimedia's communications through the Upstream surveillance program, even if the NSA conducts Upstream surveillance on only a single Internet backbone link. *See supra* page 7.

these three prongs would establish that the NSA has copied and collected some of Wikimedia's communications in the course of the NSA's Upstream surveillance program, thereby providing Wikimedia standing to sue here.

The sufficiency of the evidence with respect to each of these prongs is discussed in detail below. The summary judgment record contains specific facts which show no genuine dispute as to the veracity of the first two prongs of the Wikimedia Allegation. With respect to the third prong, however, the summary judgment factual record contains specific facts that establish, without a genuine dispute of material fact, that the NSA, in the course of Upstream surveillance, does not need to be copying any of Wikimedia's communications as a technological necessity. Thus, the summary judgment record does not contain the facts necessary for Wikimedia to establish standing at summary judgment via the Wikimedia Allegation.

A.

The first prong of the Wikimedia Allegation is that Wikimedia's communications almost certainly traverse every international Internet backbone link connecting the United States with the rest of the world.

Wikimedia primarily supports this contention through the declarations of Scott Bradner, plaintiff's Internet expert.[33] Mr. Bradner states that "it is virtually certain that Wikimedia's international communications traverse every circuit carrying public Internet traffic on every international cable connecting the U.S. to other countries." Bradner Decl. ¶ 6(d), ECF No. 168-2. Mr. Bradner supports this conclusion with evidence of the volume and global distribution of Wikimedia's communications and the relatively few international circuits connecting the U.S. to

[33] Mr. Bradner worked at Harvard University from 1966 to 2016 in a variety of technical and educational roles, including service as Harvard University's Chief Technology Security Officer for a number of years.

other countries. *Id.* at ¶¶ 346-47, 201-05, 209, 218, 220. Thus, Mr. Bradner concludes, to a virtual certainty, that every international fiber-optic cable that transports Internet communications between the U.S. and the rest of the world transports at least some of Wikimedia's international communications.

Defendants have not disputed this fact. *See* Defs.' Brief in Support of Motion for Summary Judgment, Dkt. 162 at 1 (referring to Wikimedia's standing argument as a "one-legged stool" and taking issue with the other two prongs of Wikimedia's standing argument, but not with the argument that Wikimedia's communications traverse every international Internet backbone link).[34]

Thus, there is no genuine dispute between the parties in the summary judgment record that Wikimedia's communications almost certainly traverse every international Internet backbone link connecting the United States with the rest of the world. Wikimedia has presented specific facts, supported by the conclusion of Mr. Bradner, that establish the first prong of the Wikimedia Allegation.

B.

The second prong of the Wikimedia Allegation is that the NSA conducts Upstream surveillance at one or more international Internet backbone links, all of which, as established in the first prong, some of Wikimedia's communications traverse.

Wikimedia primarily relies upon a sentence in a redacted 2011 FISC Opinion and on language describing the Internet backbone in the PCLOB 702 Report to establish this prong. The

---

[34] The government has not explicitly conceded this prong of the Wikimedia Allegation, that Wikimedia's communications traverse every international Internet backbone link connecting the United States with the rest of the world. But the government has indicated that even assuming *arguendo* that Wikimedia has presented sufficient facts to establish this first prong, Wikimedia still does not have standing in this case. *See also id.* at 21.

sentence in the 2011 FISC Opinion states: the "NSA will acquire a wholly domestic 'about' communication if the transaction containing the communication is routed through an international Internet link being monitored by NSA or is routed through a foreign server." [Redacted], 2011 WL 10945618, at *15. Defendants' Rule 30(b)(6) witness confirmed the accuracy of this statement as of October 2011.[35] *See* R. Richards Dep. at 160:4-17. Thus, as a statement of a party opponent, this statement is admitted as part of the summary judgment record.

Based on this admission, plaintiff contends that Upstream surveillance involves monitoring "international Internet link[s]." Defendants, however, assert that the meaning of the term "international Internet link" is protected by the state secrets privilege and cannot be confirmed or denied by defendants. Defendants' Rule 30(b)(6) witness testified that "unlike the other words you had me go through in terms of definitions…[which were] what a teleco[m] expert would" provide, the "NSA has an understanding of this term [international Internet link] that is specific to how [the FISC Judge] described it, but it's classified to provide any further information." R. Richards Dep. at 160:19-161:22. Thus, the differences between the term "international Internet link" and the term "circuits," which is a colloquial term used in the telecom industry and is used to describe where along the Internet backbone Upstream collection occurs in the PCLOB 702 Report,[36] cannot be known without violation of the state secrets

---

[35] *See supra* note 29 for further detail as to why the statement in the 2011 FISC Opinion is not inadmissible hearsay in the context of this litigation as a result of defendants' Rule 30(b)(6) testimony regarding the statement.

[36] It is worth noting that the PCLOB 702 Report's reference to "circuits" does not suggest that the NSA is conducting surveillance on more than one circuit. To be sure, the PCLOB 702 Report does use the term "circuits," but it does not do so to refer to the number of sites the NSA is monitoring. Instead, the PCLOB 702 Report uses the term "circuits" in the context of defining the "Internet backbone." Specifically, the PCLOB 702 Report explains that the "Internet backbone" consists of "circuits that are used to facilitate Internet communications[.]" PCLOB 702 Report at 36-37.

privilege.[37] *See* PCLOB 702 Report, at 35-37. Moreover, that this statement was accurate on October 3, 2011 says nothing of this statement's accuracy either in 2015, when this suit was filed, or today.[38]

Rather than belabor the squabble between the parties about the meaning of this particular term from a 2011 FISC Opinion, a different, admissible record document sheds significantly more light on this prong of the Wikimedia Allegation. The Public Declaration of Daniel R. Coats, Director of National Intelligence ("DNI"), states that the United States Intelligence Community "has publicly acknowledged that Upstream surveillance is conducted on one or more points on the Internet backbone" and that the United States Intelligence Community "has publicly acknowledged that…NSA is monitoring at least one circuit carrying international Internet communications." Pub. Decl. of Daniel R. Coats, DNI, ¶¶ 30, 37, ECF No. 138-2.[39] In other words, the DNI, who oversees the United States Intelligence Community, has admitted, in the course of this litigation, that the NSA conducts Upstream surveillance on at least one point on the Internet backbone and, to the extent the terms Internet backbone and international Internet circuit are not interchangeable, on at least one circuit carrying international Internet communications.[40]

---

[37] The state secrets privilege's applicability to this case is discussed in significantly greater depth *infra* Part VI.

[38] The statement from the 2011 FISC Opinion pertains to the Upstream surveillance program's collection of "about" communications. As of April 2017, Upstream surveillance no longer involves any "about" collection. Thus, at least the conclusion of this conditional statement is no longer accurate today.

[39] Neither party has cited to these specific paragraphs of the Public Declaration of the DNI in their briefs. Nonetheless, the Public Declaration of the DNI is clearly part of the evidentiary record in this matter, as defendants have cited to other paragraphs of this declaration in their statement of undisputed material facts. Moreover, as the "oversee[r of] the United States Intelligence Community," the DNI is in a position to make such statements from personal knowledge.

[40] In this context, the terms Internet backbone and international Internet circuits both refer on some level to the transoceanic fiber-optic cables that transport Internet communications and connect the U.S. to the rest of the world.

Accordingly, the undisputed summary judgment record adequately establishes that the NSA monitors at least one circuit carrying international Internet communications in the course of Upstream surveillance and that Wikimedia's communications traverse every circuit carrying international Internet communications from the United States to the rest of the world. Thus, Wikimedia has established the first two prongs of the Wikimedia Allegation with the support of admissible record evidence and without a genuine dispute as to any material fact.

C.

With respect to the third prong, however, the summary judgment factual record contains specific facts that establish, without a genuine dispute of material fact, that it is *not* a technological necessity that the NSA has copied or collected some of Wikimedia's communications over the one circuit that the NSA admits monitoring to conduct Upstream surveillance.[41] Accordingly, the summary judgment record does not contain the facts necessary for Wikimedia to establish standing at summary judgment via the Wikimedia Allegation.

To address this prong of the Wikimedia Allegation, both parties have submitted extensive expert reports. The government's expert, Dr. Henning Schulzrinne,[42] has provided expert testimony that details a method of collecting Internet communications, which could, *hypothetically*, avoid collecting any of Wikimedia's communications. Dr. Schulzrinne Decl. ¶

---

[41] Importantly, to establish standing, Wikimedia need only prove that the NSA has copied or scanned some of its communications as part of the Upstream surveillance program, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Wikimedia has chosen to prove that it is a technological necessity that the NSA has copied or scanned some of its communications only because the government's assertion of the state secrets privilege prevents Wikimedia from posing the more direct question of whether the NSA has actually copied or scanned any of Wikimedia's communications as part of the Upstream surveillance program. *See Wikimedia Found. v. Nat'l Sec. Agency*, 335 F. Supp. 3d 772, 788-90 (D. Md. 2018).

[42] Dr. Henning Schulzrinne has been a professor of computer science at Columbia University since 1996 and holds a Ph.D. and a Master's Degree in Electrical Engineering.

77-88. Thus, Dr. Schulzrinne concludes that the NSA, via Upstream surveillance, does not *have* to be collecting any of Wikimedia's communications "as a matter of technological necessity." Dr. Schulzrinne 2d Decl. ¶ 2. Importantly, Dr. Schulzrinne does not provide testimony about the actual operational details of Upstream surveillance because the actual operational details of Upstream surveillance are classified and protected by the state secrets privilege, and thus, Dr. Schulzrinne does not know any of the classified operational details. *Id.* at ¶ 3-4.

On the other side, Wikimedia's expert, Scott Bradner, has provided expert testimony in which he opines, based on a combination of technical and practical factors, that the NSA "most likely" copies all communications transported across an international Internet circuit *before* filtering any of the communications. Bradner Decl. ¶ 282. As a result, Mr. Bradner concludes that "even if the NSA were monitoring only a single circuit under [U]pstream collection, it would be copying and reviewing at least some of Wikimedia's communications." *Id.* at ¶ 353.

Each expert unsurprisingly takes issue with the other's findings. Dr. Schulzrinne claims that Mr. Bradner has provided "no support, and certainly none based in Internet technology and engineering, for concluding that the NSA 'almost certainly' (Bradner Decl. ¶ 6(a)) copies and scans all communications traversing any circuit it monitors, including Wikimedia's." Dr. Schulzrinne 2d Decl. ¶ 5. And Mr. Bradner claims that Dr. Schulzrinne's conclusion that the NSA does not have to be collecting any of Wikimedia's communications as a matter of technological necessity "is simply implausible as a practical matter given everything that is known about [U]pstream collection." Bradner Decl. ¶ 362. For the reasons that follow, this dispute does not present a triable issue of fact.

To begin with, it is necessary to address the practical grounds on which Mr. Bradner reaches his conclusions. Mr. Bradner contends that the NSA could not accomplish its stated goal

of "*comprehensively acquir[ing]* communications that are sent to or from its targets" through Upstream surveillance without first copying all international communications transported over the circuit(s) that the NSA is monitoring. *Id.* at ¶ 333 (quoting PCLOB 702 Report, at 10, 123, 143 (emphasis added)); *Id.* at ¶ 335. To accomplish this goal, Mr. Bradner opines that the NSA is "most likely" copying all of the communications traveling across a circuit before later filtering those communications based on the NSA's targeted selectors. *Id.* at ¶¶ 282, 289. As the basis for this opinion, Mr. Bradner claims (i) that any other method would require the NSA to share sensitive information about its targets and/or filtering criteria with an assisting provider, which the NSA would prefer not to do, (ii) that any other method would require the NSA to place an NSA-operated device into the heart of an ISP's network, which the NSA would prefer not to do, and (iii) that the NSA has no operational incentive to reduce the number of communications it scans for selectors. *Id.* at ¶¶ 283-88.

None of Mr. Bradner's bases for this opinion, however, have a non-speculative foundation in technology. Instead, speculative assumptions about the NSA's surveillance practices and priorities and the NSA's resources and capabilities form the basis for Mr. Bradner's opinion in this regard.[43] *See* Dr. Schulzrinne 2d Decl. ¶ 73. Simply put, Mr. Bradner does not know what the NSA prioritizes in the Upstream surveillance program because that information is classified, and therefore Mr. Bradner has no knowledge or information about it.

---

[43] *See, e.g., Obama v. Klayman*, 800 F.3d 559, 567 (D.C. Cir. 2015) (rejecting a plaintiff's claim that the NSA's collection must be comprehensive to be effective because "there are various competing interests that may constrain the government's pursuit of effective surveillance. Plaintiffs' inference fails to account for the possibility that legal constraints, technical challenges, budget limitations, or other interests prevented NSA from collecting metadata from Verizon Wireless."). Wikimedia has gone significantly further than the plaintiffs in *Klayman* to address the technological issues pertinent to the effectiveness of a less comprehensive surveillance system, but Mr. Bradner still takes significant speculative leaps about the NSA's practical and operational decision-making to reach these particular aspects of his conclusions. These specific conclusions require speculative leaps which are too significant to accept as the foundational basis for an expert's opinion.

As a result, Mr. Bradner's opinions as to these specific propositions are inadmissible pursuant to Rule 702, Fed. R. Evid., and the standards articulated in *Daubert v. Merrell Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).[44]

Moreover, even if Mr. Bradner's opinions on these specific propositions were admissible, any conclusions drawn from those opinions would be barred by the state secrets doctrine, as further discussed *infra* Part VI. No matter how intuitively appealing Mr. Bradner's opinions about the NSA's operational priorities may seem, courts have consistently recognized that "judicial intuition" about such propositions "is no substitute for [the] documented risks and threats posed by the potential disclosure of national security information." *Al-Haramain Islamic Found., Inc. v. Bush*, 507 F.3d 1190, 1203 (9th Cir. 2007). Importantly, defendants cannot effectively defend themselves against Mr. Bradner's speculations without disclosing information about the operational details of the NSA's Upstream surveillance program. But defendants have thoroughly documented the risks of such a disclosure in the classified declaration, explaining that to reveal such facts regarding the operational details of the Upstream surveillance collection process, even considering the public disclosures made to date, would provide insight into the structure and operations of the Upstream surveillance program and in so doing, undermine the effectiveness of this important intelligence method. Thus, even if Mr. Bradner's conclusions,

---

[44] Rule 702 provides that an expert may offer opinion testimony if "the expert's scientific, technical, or other specialized knowledge" will be helpful to understand the evidence or to determine a fact in issue, the proffered opinion is "based on sufficient facts or data," and it is "the product of reliable principles and methods…reliably applied…to the facts of the case." Fed. R. Evid. 702(a)-(d). *Daubert* explained that to meet the test of admissibility under Rule 702, an expert's testimony must rest on a reliable foundation, meaning it "must be based on scientific, technical, or other specialized *knowledge* and not belief or speculation." *Oglesby v. Gen. Motors Corp.*, 190 F.3d 244, 250 (4th Cir. 1999) (emphasis in original); *see also Nease v. Ford Motor Co.*, 848 F.3d 219, 229, 231 (4th Cir. 2017). Here, the critical propositions that form the basis for Mr. Bradner's opinion that the NSA is "most likely" copying all communications before any filtering do not meet this requirement as they are based on Mr. Bradner's speculation as to the NSA's operational priorities and capabilities, not on any technical requirements for the collection of Internet communications. Although the NSA has made some public disclosures about Upstream surveillance, Mr. Bradner's interpretations of single sentences within the public disclosures stretches those disclosures far beyond a natural reading of them, and again, is not based on any knowledge, technical or otherwise.

built off assumptions about the NSA's operational goals from the NSA's limited public disclosures, were admissible as expert opinions, the state secrets doctrine would bar any further litigation of this prong of Wikimedia's standing argument, as further discussed *infra* Part VI.

Analysis of the third prong of the Wikimedia Allegation, however, does not end with dismissal of Mr. Bradner's non-technical assumptions. Each expert has also presented technical arguments for and against the proposition that the NSA must be collecting at least some of Wikimedia's communications at the circuit(s) monitored pursuant to the Upstream surveillance program.

Dr. Schulzrinne explains how the NSA, using the technique of "traffic mirroring" in a specific manner,[45] could conduct Upstream surveillance on an international Internet circuit "without intercepting, copying, reviewing, or otherwise interacting with [the] communications of Wikimedia." Dr. Schulzrinne Decl. ¶ 77. To begin with, Wikimedia has been allocated a number of static IP addresses. *Id.* at ¶ 80. A "static" IP address is an IP address that is assigned on a *permanent* basis from the appropriate regional Internet registry. *See id.* at ¶ 32-33. Static IP addresses are generally assigned to large enterprises on the Internet so that users around the world have consistent access to their websites. Public databases record, with very high accuracy, which IP address blocks are used by what entities. *Id.* Thus, any member of the public can ascertain all of the IP addresses assigned to Wikimedia.

Through a process of "blacklisting" Wikimedia's IP addresses, the NSA could conduct Upstream surveillance without receiving access to any of Wikimedia's communications. *Id.* at ¶ 82. To do so, the NSA could blacklist all of Wikimedia's IP addresses using an access control

---

[45] Traffic mirroring, as defined in the statement of material facts in the summary judgment record, is a technical term for a process by which all communications passing through a router or switch can be copied without interrupting the flow of communications.

list, a list employed in the traffic mirroring process that determines which packets carrying Internet communications will be copied and collected at a certain circuit on the Internet backbone. By blacklisting Wikimedia's IP addresses, all Internet communications *except* those containing Wikimedia's blacklisted IP addresses would be copied and collected by the NSA. Importantly, this hypothetical does not propose that the NSA is copying all Internet communications other than Wikimedia's, but rather states that, as a technical matter, the NSA *could* blacklist certain high-frequency, low-interest IP addresses to minimize the collection of communications of little interest to the NSA and that Wikimedia's IP addresses *could* be high-frequency, low-interest IP addresses to the NSA. Thus, strictly considering the technological limitations of copying Internet communication in transit, it is possible that the NSA has not copied and collected any of Wikimedia's communications despite monitoring an international Internet circuit that transmits some of Wikimedia's communications.[46]

In response, Mr. Bradner finds this hypothetical "simply implausible" as a practical matter given everything that is known about Upstream surveillance, although Mr. Bradner does admit that selective collection is technologically possible. Bradner Decl. ¶ 362, 272(b), 280-81, 299, 325, 366. The foundation for Mr. Bradner's response is that the NSA has disclosed to the public that Upstream surveillance operates by identifying "selectors," the specific means by which the targets communicate, such as email addresses or telephone numbers.[47] Because the

---

[46] In addition to blacklisting Wikimedia's IP addresses, Dr. Schulzrinne proposes several other whitelisting or blacklisting options which would prevent the NSA from collecting Wikimedia's international Internet communications. Dr. Schulzrinne Decl. ¶ 77-88. For example, the NSA could blacklist the ports assigned to HTTP and HTTPS communications so as not to collect any web communications that involve accessing websites. *Id.* at ¶ 79.

[47] NSA Director of Civil Liberties and Privacy Office Report, *NSA's Implementation of FISA Section 702* 4 (2014), *available at* https://www.nsa.gov/Portals/70/documents/news-features/press-room/statements/NSAImplementationofFISA70216Apr2014.pdf.

NSA cannot know in advance which communications contain selectors, Mr. Bradner contends, the NSA must first copy all communications before scanning any of them for selectors. Bradner Decl. ¶ 333, 301.

Despite Mr. Bradner's arguments to the contrary, the traffic mirroring hypothetical proposed by Dr. Schulzrinne does not contradict the government's public disclosures about Upstream surveillance. Importantly, the government has described Upstream surveillance as involving three steps—(1) filtering, (2) scanning, and (3) ingesting.[48] The whitelisting and blacklisting process of traffic mirroring proposed by Dr. Schulzrinne would occur at the first step in the NSA's collection process, the filtering, prior to any copying or scanning. Thus, under Dr. Schulzrinne's hypothetical, the first step, filtering, would involve a combination of whitelisting and blacklisting to exclude wholly domestic communications *and* other low interest communications, and Wikimedia's communications may qualify as low interest communications that the NSA filters out.[49] *Second*, and only after filtering, the NSA would scan the remaining communications for "selectors," which could result in the collection of both communications to or from a targeted selector and about a targeted selector. *See* Dr. Schulzrinne 2d Decl. ¶ 50-52. This second step described in the government's public disclosures is the step on which Mr. Brander focuses. Given the distinction between the first two steps, Dr. Schulzrinne's hypothetical is consistent with government's public disclosures about Upstream surveillance.

---

[48] *See* Material Fact 35; Pub. Decl. of Daniel R. Coats, Director of National Intelligence, ¶ 15, ECF No. 138-2.

[49] It is noted that the government has not disclosed that anything other than wholly domestic communications are filtered out at the first step in the Upstream collection process. Given the government's limited disclosures about the technical details of how Upstream surveillance operates, however, this disclosure does not mean that the government does not, or could not, engage in additional filtering at the first step in the Upstream surveillance collection process. Whether or not the government actually engages in additional filtering at the first step in the Upstream surveillance collection process is a fact protected by the state secrets privilege. *See Wikimedia Found. v. Nat'l Sec. Agency*, 335 F. Supp. 3d 772, 789-90 (D. Md. 2018); Pub. Decl. of Daniel R. Coats, DNI, ¶ 18(B), 18(D), ECF No. 138-2.

Moreover, the hypothetical, regardless of whether it is actually how the NSA conducts Upstream surveillance, does show that there is a technological method by which the NSA could conduct Upstream surveillance on a circuit transporting International internet communications without copying, collecting, or otherwise reviewing any of Wikimedia's communications that traverse that path.

But this does not end the analysis, for there is a technological hurdle that remains. Even if the NSA used the whitelisting and blacklisting techniques proposed by Dr. Schulzrinne to filter the communications it collected via Upstream surveillance, Mr. Bradner maintains that there are three scenarios in which Wikimedia's communications would still be copied and scanned by the NSA. Bradner Decl. ¶¶ 367(b), 370. In these three specific scenarios—namely (i) communications contained within a multi-communication transaction,[50] (ii) emails to or from Wikimedia involving a person located abroad who is using an email service located in the U.S.,[51] or (iii) a person located abroad who accesses Wikimedia's websites through a U.S.-based Virtual

---

[50] A "multi-communication transaction" (MCT) is "an Internet transaction that contain[s] multiple discrete communications." NSA Response to Plaintiff's Interrogatory No. 8 (Dec. 22, 2017). When an email user logs into their email service to check his or her email, the group of all unread email messages is transmitted together as a single communication from the email service to the subscribing user's inbox. This transmission of multiple emails in a single communication might be considered an MCT. Bradner Decl. ¶¶ 67, 132, 317. In transit, an MCT of this type would contain the IP address of the email service as the sender and the IP address of the user as the recipient. If an email to or from Wikimedia were contained within the batch of emails sent as an MCT, the Wikimedia email would be transmitted to the user's inbox without Wikimedia's IP address in the individual packet headers of the MCT. Dr. Schulzrinne 2d Decl. ¶ 78. Thus, this specific type of Wikimedia communication could be transmitted from an email service to a user of the email service without Wikimedia's IP address being the source or destination IP address. And as a result, blacklisting Wikimedia's IP addresses would not prevent the NSA's collection of such an email from an international Internet circuit which the NSA is monitoring.

[51] This scenario is similar to the first MCT scenario. If (i) an email user sent an email to Wikimedia or received an email from Wikimedia, (ii) that email user was abroad, and (iii) that email user utilized a U.S.-based email service, the communication between the email user and the email service would not include Wikimedia's IP address in the packet headers and would need to traverse an international Internet circuit between the U.S.-based email service and the user located abroad. Bradner Decl. ¶ 367(b)(2); Dr. Schulzrinne 2d Decl. ¶ 81. Thus, this specific type of Wikimedia communication could be transmitted from an email service to a user of the email service without Wikimedia's IP address being the source or destination IP address. And as a result, blacklisting Wikimedia's IP addresses would not prevent the NSA's collection of such an email from an international Internet circuit which the NSA is monitoring.

Private Network (VPN),[52] Wikimedia's IP address would not appear as the source or destination IP address on the packet header traversing the international Internet circuit into or out of the U.S. *See* Bradner Decl. ¶ 367(b)(1)-(3); Dr. Schulzrinne 2d Decl. ¶ 77-87. Thus, these communications would not be blocked by the NSA's hypothetical blacklist of Wikimedia's IP addresses because the communications would not contain Wikimedia's IP address in the packet header, despite involving a Wikimedia communication.

Dr. Schulzrinne admits that each of these scenarios is "theoretically possible" but "could come to pass only in the uncertain event that particular conditions are met." Dr. Schulzrinne 2d Decl. ¶ 77. For communications in each of these three scenarios to be collected by the NSA through Upstream surveillance, at least four conditions would have to be met,[53] none of which Wikimedia has established as to any of their communications in this case. Specifically, for Wikimedia communications to exist in either of the first two scenarios, an email user in a foreign location must be downloading emails from a server located in the United States (such that the communication would traverse an international Internet circuit monitored by the NSA) *and* the email user must be sending email to and/or receiving email from Wikimedia. *Id.* at ¶¶ 78, 81. Wikimedia has not presented evidence of any such subset of its communications.[54] For

---

[52] When a user communicates via a Virtual Private Network (VPN), all of the user's communications are encrypted and first routed through the VPN server before being directed to their ultimate destination. Dr. Schulzrinne 2d Decl. ¶ 57. As a result, first, each communication's packet is assigned the VPN server's address as its destination IP address, not the IP address of the ultimate destination. *Id.* Then, once the communication has reached the VPN server (destination one), the communication travels from the VPN server to the ultimate destination (destination two), with the VPN server IP address as the source IP address, rather than the individual user's IP address. Therefore, if a person is located abroad and accesses Wikimedia's website while using a U.S.-based VPN and the first leg communication between the VPN user and the VPN server traverses an international Internet circuit that the NSA is monitoring, the NSA could collect that communication even if the NSA has blacklisted Wikimedia's IP addresses. Bradner Decl. ¶ 367(b)(3).

[53] Dr. Schulzrinne 2d Decl. ¶ 78, 81, 83.

[54] It is worth noting that Wikimedia has acknowledged that is does not know the volume of its international email communications, or the countries from which the emails are received. *See* Technical Statistics Chart. In addition to

Wikimedia's communications to exist in the third scenario, a user of a Virtual Private Network (VPN) that is based in the United States must use that VPN while abroad to visit one of Wikimedia's websites, and the NSA must monitor the international Internet circuit that transmits that communication from the user abroad to the domestic VPN. Again, Wikimedia has not presented evidence of any such subset of its communications. As a result, satisfaction of the chain of conditions necessary to establish that the NSA collected Wikimedia's communications in one of these three circumstances is too speculative to establish standing. *See Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1148, 1150 (2013) (holding that a speculative chain consisting of five contingencies was insufficient to establish standing). Thus, although it is possible that such communications exist,[55] the summary judgment record does not contain any evidence that such communications actually exist, a requirement at this stage of the litigation. *See Clapper*, 133 S. Ct. at 1148-49.

In sum, the undisputed summary judgment record does not establish that the NSA has copied any of Wikimedia's international Internet communications in the course of Upstream surveillance, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Specifically, the summary judgment record establishes that it is not a technological necessity that the NSA must copy all of the text-based Internet communications traversing a circuit that the NSA monitors while conducting Upstream surveillance. The NSA could, *hypothetically*, utilize a process of

---

the total volume and location of all of Wikimedia's international email communications being unknown, this particular subset of Wikimedia's international email communications is also unknown – in volume, in geographic diversity, or even whether such communications exist.

[55] It is worth noting that if such communications exist, they are likely to be far fewer in number than the trillions of international Wikimedia communications every year that traverse every International circuit connecting the U.S. to the rest of the world. Thus, a finding that such communications exist could trigger a re-evaluation of the first prong of Wikimedia's standing argument, *i.e.* that Wikimedia's subject international Internet communications traverse every international Internet backbone link connecting the United States with the rest of the world.

whitelisting and blacklisting to filter out low-interest Internet communications, including Wikimedia's communications, prior to scanning the Internet communications for targeted selectors. At most, there is a genuine dispute of material fact as to whether the NSA can conduct Upstream surveillance without copying Wikimedia's communications, if any, that (i) are contained within a multi-communication transaction, (ii) are emails to or from Wikimedia involving a person located abroad using an email service located in the U.S., or (iii) involve a person located abroad accessing Wikimedia's websites through a U.S.-based Virtual Private Network (VPN) *and* that traverse an NSA-monitored circuit. To the extent there is a genuine issue of material fact with respect to the NSA's collection of this currently unidentified subset of Wikimedia's international communications, that issue cannot be further litigated given the state secrets doctrine, as further discussed *infra* Part VI.

VI.

Even assuming *arguendo* that, there is a genuine dispute of material fact as to the third prong of the Wikimedia Allegation, the question remains as to how the matter should proceed consistent with Supreme Court and Fourth Circuit precedent regarding the state secrets doctrine. Wikimedia's standing cannot be fairly litigated any further without disclosure of state secrets absolutely protected by the United States' privilege. For Wikimedia to litigate the standing issue further, and for defendants to defend adequately in any further litigation, would require the disclosure of protected state secrets, namely details about the Upstream surveillance program's operations. For the reasons that follow, therefore, the standing issue cannot be tried, or otherwise further litigated, without risking or requiring harmful disclosures of privileged state secrets, an outcome prohibited under binding Supreme Court and Fourth Circuit precedent. Thus, the case must be dismissed, and judgment must be entered in favor of defendants.

A.

It is necessary first to review the well-settled Supreme Court and Fourth Circuit precedent concerning the state secrets doctrine. Settled Supreme Court and Fourth Circuit precedent make clear that "[u]nder the state secrets doctrine, the United States may prevent the disclosure of information in a judicial proceeding if 'there is a reasonable danger' that such disclosure 'will expose…matters which, in the interest of national security should not be divulged.'" *Abilt v. CIA*, 848 F.3d 305, 310-11 (4th Cir. 2017) (quoting *El-Masri v. United States*, 479 F.3d 296, 302 (4th Cir. 2007)) (quoting *United States v. Reynolds*, 345 U.S. 1, 10 (1953)). In this regard, the Fourth Circuit has recognized that the state secrets doctrine "performs a function of constitutional significance, because it allows the executive branch to protect information whose secrecy is necessary to its military and foreign-affairs responsibilities." *Id.* at 312 (quoting *El-Masri*, 479 F.3d at 303).

The Fourth Circuit has mandated a three-step analysis for resolution of the state secrets question:

> First, "the court must ascertain that the procedural requirements for invoking the state secrets privilege have been satisfied." Second, "the court must decide whether the information sought to be protected qualifies as privileged under the state secrets doctrine." Third, if the "information is determined to be privileged, the ultimate question to be resolved is how the matter should proceed in light of the successful privilege claim."

Abilt, 848 F.3d at 311 (quoting *El-Masri*, 479 F.3d at 304). Previously, an Order and Memorandum Opinion issued in this case, which concluded that defendants satisfied the procedural requirements necessary to invoke the state secrets privilege, that the information sought to be protected qualified as privileged under the state secrets doctrine, and that therefore, Wikimedia's motion to compel certain information in discovery had to be denied. *Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 790 (D. Md. 2018). The seven categories of

information determined to be privileged under the state secrets doctrine in relation to plaintiff's motion to compel discovery are the same categories of information at issue for plaintiff to establish standing via further litigation of this case.[56] Thus, as already established in the previous Memorandum Opinion and Order, the first two steps of the state secrets analysis have been resolved, and the step that remains is "how the matter should proceed in light of the successful privilege claim." *Abilt*, 848 F.3d at 311.

B.

How the matter should proceed turns on the centrality of the privileged information to the issue at hand. Whether the NSA has copied and collected any of Wikimedia's international Internet communications, or such collection is certainly impending, or there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection, is the threshold issue for Wikimedia to establish standing in this litigation. Where, as here, the privileged information is so central to the subject matter of the litigation, dismissal is the appropriate, and only available, course of action.

As the Fourth Circuit has made quite clear, "both Supreme Court precedent and our own cases provide that when a judge has satisfied himself [or herself] that the dangers asserted by the government are substantial and real, he [or she] need not—indeed, should not—probe further." *Sterling v. Tenet*, 416 F.3d 338, 345 (4th Cir. 2005). Moreover, Fourth Circuit precedent establishes that where "circumstances make clear that sensitive military secrets will be so central to the subject matter of the litigation that any attempt to proceed will threaten disclosure of the

---

[56] The seven categories of information privileged pursuant to the state secrets doctrine are: (i) individuals or entities subject to Upstream surveillance activities, (ii) operational details of the Upstream collection process, (iii) locations at which Upstream surveillance is conducted, (iv) categories of Internet-based communications subject to Upstream surveillance activities, (v) the scope and scale on which Upstream surveillance is or has been conducted, (vi) the NSA's cryptanalytic capabilities, and (vii) additional categories of classified information contained in FISC opinions, orders and submissions.

privileged matters, dismissal is the appropriate remedy." *El-Masri v. Tenet*, 437 F. Supp. 2d 530, 538-39 (E.D. Va. 2006) (quoting *Sterling*, 416 F.3d at 348), *aff'd*, 479 F.3d 296 (4th Cir. 2007).[57]

As such, "[i]f a proceeding involving state secrets can be fairly litigated without resort to the privileged information, it may continue." *El–Masri*, 479 F.3d at 306. On the other hand, "a proceeding in which the state secrets privilege is successfully interposed must be dismissed if the circumstances make clear that privileged information will be so central to the litigation that any attempt to proceed will threaten that information's disclosure." *Id.* at 308 (citations omitted).[58] Such a decision is never taken lightly, as "dismissal is appropriate '[o]nly when no amount of effort and care on the part of the court and the parties will safeguard privileged material.'" *Sterling*, 416 F.3d at 348 (quoting *Fitzgerald v. Penthouse Int'l, Ltd.*, 776 F.2d 1236, 1244 (4th Cir. 1985)) (alteration in original). Nonetheless, "dismissal follows inevitably when the sum and substance of the case involves state secrets." *Id.* at 347. In this regard, the Fourth Circuit has identified three examples of circumstances in which the privileged information is so central to the litigation that dismissal is required. First, "dismissal is required if the plaintiff cannot prove the *prima facie* elements of his or her claim without privileged evidence." *Abilt*, 848 F.3d at 313-14 (citing *Farnsworth Cannon, Inc. v. Grimes*, 635 F.2d 268, 281 (4th Cir. 1980) (en banc) (per curiam)). Second, "even if the plaintiff can prove a *prima facie* case without resort to privileged information, the case should be dismissed if 'the defendants could not properly defend themselves without using privileged evidence.'" *Id.* at 314 (quoting *El-Masri*, 479 F.3d at 309).

---

[57] Importantly, "state secrets and military secrets are equally valid bases for invocation of the evidentiary privilege." *Sterling*, 416 F.3d at 343 (internal quotation marks and alterations omitted).

[58] *See also Sterling*, 416 F.3d at 347–48 ("We have long recognized that when 'the very subject of [the] litigation is itself a state secret,' which provides 'no way [that] case could be tried without compromising sensitive military secrets,' a district court may properly dismiss the plaintiff's case." (quoting *Fitzgerald*, 776 F.2d at 1243) (alterations in original)); *Bowles v. United States*, 950 F.2d 154, 156 (4th Cir. 1991) (per curiam) ("If the case cannot be tried without compromising sensitive foreign policy secrets, the case must be dismissed.").

Third, "dismissal is appropriate where further litigation would present an unjustifiable risk of disclosure" of state secrets. *Id.* (citing *El-Masri*, 479 F.3d at 308).

C.

Given these principles and given "the delicate balance to be struck in applying the state secrets doctrine," it is appropriate to analyze the litigation at hand, namely the centrality of state secrets to Wikimedia's standing. *El-Masri*, 479 F.3d at 308. To establish standing, Wikimedia must prove (1) injury-in-fact, (2) causation, and (3) redressability. Through an extensive jurisdictional discovery process, Wikimedia has established that the NSA monitors at least one circuit carrying international Internet communications in the course of Upstream surveillance and that Wikimedia's communications traverse every circuit carrying international Internet communications from the United States to the rest of the world. Importantly, this extensive jurisdictional discovery process has resulted in the compilation of a voluminous record, including hundreds of pages of expert reports, government disclosures and declassified documents regarding Upstream surveillance, Rule 30(b)(6) testimony from an NSA representative, and extensive interrogatory responses from the parties. Thus, Wikimedia has been granted the opportunity to establish its standing without resort to privileged information, and Wikimedia has made significant progress on that front.

Nonetheless, the summary judgment record does not establish that the NSA has copied or collected any of Wikimedia's communications via Upstream surveillance conducted on an NSA-monitored circuit, that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Wikimedia has been unable to make this showing because it is not true, as a technological necessity, that the NSA must be copying every text-based communication that traverses a circuit that the NSA

monitors. Indeed, Dr. Schulzrinne has convincingly demonstrated that there are technologically feasible methods by which the NSA could hypothetically operate Upstream surveillance that would result in the NSA not copying or collecting any of Wikimedia's communications. Thus, the undisputed summary judgment record establishes that Wikimedia does not have Article III standing sufficient to survive summary judgment.

Even if Wikimedia could establish a *prima facie* case of its standing based solely on the public, unclassified record, which it has not been able to do thus far in this case, the state secrets doctrine still requires dismissal because the defendants cannot properly defend themselves without using privileged evidence. The Fourth Circuit "ha[s] consistently upheld dismissal when the defendants could not properly defend themselves without using privileged information." *Abilt v. CIA*, 848 F.3d 305, 316 (4th Cir. 2017). As in *El-Masri*, "virtually any conceivable response to [Wikimedia's] allegations [that the NSA has copied and collected some of Wikimedia's international Internet communications] would disclose privileged information." *El-Masri*, 479 F.3d at 310. Defendants have provided a detailed and persuasive explanation, in more than 60 pages of classified declarations, that disclosure of the entities subject to Upstream surveillance activity and the operational details of the Upstream collection process would (i) undermine ongoing intelligence operations, (ii) deprive the NSA of existing intelligence operations, and significantly, (iii) provide foreign adversaries with the tools necessary both to evade U.S. intelligence operations and to conduct their own operations against the United States and its allies. *Wikimedia Found. v. Nat'l Sec. Agency*, 335 F. Supp. 3d 772, 789 (D. Md. 2018). Accordingly, defendants could not properly defend themselves in any further litigation of Wikimedia's standing, and thus, the case must be dismissed.

Moreover, if the issue of Wikimedia's standing were further adjudicated, "the whole

object of the [adjudication]…[would be] to establish a fact that is a state secret," presenting an unjustifiable risk of disclosing privileged information. *Sterling*, 416 F.3d at 348. Courts have concluded that where, as here, the information sought to be disclosed involves the identity of parties whose communications have been acquired, this information is properly privileged. *See Al-Haramain Islamic Found., Inc. v. Bush*, 507 F.3d 1190, 1203-04 (9th Cir. 2007) (finding that the fact of a plaintiff's surveillance by the NSA was covered by the state secrets privilege); *Halkin v. Helms*, 598 F.2d 1, 9 (D.C. Cir. 1978) (upholding assertion of state secrets privilege with respect to "the identity of particular individuals whose communications have been acquired"). Accordingly, because the privileged information, namely the operational details of the Upstream collection process and whether any of Wikimedia's international Internet communications have been copied or collected by the NSA, is so central to the litigation of Wikimedia's standing, the case must be dismissed, and judgment must be entered in favor of defendants.

VII.

To avoid the conclusion that the case must be dismissed, Wikimedia revives its argument that 50 U.S.C. § 1806(f) displaces the state secrets doctrine in cases challenging electronic surveillance pursuant to FISA and provides for *in camera* review of the materials related to the NSA's Upstream surveillance program. This argument, however, has already been considered and rejected in this litigation. *See Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 786 (D. Md. 2018). Specifically, the "§ 1806(f) procedures do not apply where, as here, a plaintiff has not yet established that it has been the subject of electronic surveillance" as required by the statute. *Id.* at 780. Nonetheless, plaintiff raises two additional arguments as to why *in camera* review pursuant to § 1806(f) is appropriate in this case: (i) plaintiff has now established a

genuine dispute of material fact concerning its status as an "aggrieved person"[59] before invoking FISA's procedures and (ii) the Ninth Circuit recently held that § 1806(f) displaces the state secrets privilege in an affirmative legal challenge to electronic surveillance pursuant to FISA. *See Fazaga v. Fed. Bureau of Investigation*, 916 F.3d 1202 (9th Cir. 2019).

First, plaintiff has not established a genuine dispute of material fact concerning its status as an aggrieved person, *i.e.*, that plaintiff's communications have been the subject of electronic surveillance, as discussed *supra* Part V.C. As previously explained, "the text of § 1806(f) points persuasively to the conclusion that Congress intended § 1806(f) procedures to apply only after it became clear from the factual record that the movant was the subject of electronic surveillance." *Wikimedia Found.*, 335 F. Supp. 3d at 781. To be sure, "affirmative government acknowledgement of surveillance of a specific target is not the only means by which a plaintiff can establish evidence of his or her 'aggrieved person' status." *Id.* at 784. But here, despite the extensive jurisdictional discovery undertaken in this case, plaintiff has been unable to make a factual showing that Wikimedia was the subject of electronic surveillance using admissible record evidence. Thus, the §1806(f) *in camera* review procedures remain inapplicable to this case.

In addition, no binding authority establishes that § 1806(f)'s review procedures displace the state secrets doctrine even if a plaintiff survived summary judgment on the issue of whether plaintiff has been the target of electronic surveillance, which again is not the case here. Specifically, in *ACLU Foundation of Southern California v. Barr*, 952 F.2d 457 (D.C. Cir. 1991), the D.C. Circuit reasoned that "legitimate concerns about compromising ongoing foreign

---

[59] For the purposes of FISA, an "aggrieved person" is "a person who is the target of an electronic surveillance or any other person whose communications or activities were subject to electronic surveillance." 50 U.S.C. § 1801(k).

intelligence investigations" are more properly considered at the summary judgment stage, not upon the pleadings. *Id.* at 469. In doing so, the D.C. Circuit only considered what a party must show to establish his or her "aggrieved person" status and therefore invoke § 1806(f) review. Simply put, the D.C. Circuit did not consider whether or when § 1806(f) *in camera* review is inappropriate or unnecessary because of the state secrets doctrine.

Moreover, the Ninth Circuit's opinion in *Fazaga* does not hold that § 1806(f) displaces the state secrets doctrine in this case, despite plaintiff's arguments to the contrary. The Ninth Circuit reasoned in *Fazaga* that § 1806(f)'s procedures displace a dismissal remedy for the *Reynolds* state secrets doctrine *only where § 1806(f)'s procedures apply.*[60] *Fazaga*, 916 F.3d at 1234. Specifically, the Ninth Circuit held that for FISA's § 1806(f) procedures to apply, "[p]laintiffs must satisfy the definition of an 'aggrieved person.'" *Id.* at 1238. In this case, as previously discussed at length, Wikimedia has not established it is an "aggrieved person" as defined in § 1801(k). *See Wikimedia Found. v. Nat'l Sec. Agency,* 335 F. Supp. 3d 772, 780, 786 (D. Md. 2018). Thus, § 1806(f) does not apply to this case, and dismissal on state secrets grounds is appropriate, as discussed *supra* Part VI.

Notably, the only court to address this issue post-*Fazaga* held that "where the very issue

---

[60] *Fazaga* addressed a challenge to an allegedly unlawful FBI counter-terrorism investigation involving electronic surveillance. *Id.* at 1210-11. Specifically, in that case, "several sources" confirmed the identity of a confidential FBI informant and disclosed that that specific confidential informant "created audio and visual recordings" for the FBI. *Id.* at 1214. The district court dismissed all but one of plaintiff's claims at the pleading stage without further discovery based on the government's assertion of the state secrets privilege. *Id.* at 1211. The Ninth Circuit reversed, concluding that § 1806(f)'s procedures are to be used when "aggrieved persons" challenge the legality of electronic surveillance and that the district court erred by dismissing the case without reviewing the evidence. *Id.* at 1238, 1252. In remanding for further proceedings, the *Fazaga* court held that "[t]he complaint's allegations are sufficient *if proven* to establish that Plaintiffs are 'aggrieved persons.'" *Id.* at 1216 (emphasis added). Thus, the Ninth Circuit's decision reasoned that at the pleading stage of the litigation, where plaintiffs have alleged sufficient facts, assumed to be true at that stage of the litigation, to establish they are "aggrieved persons" as required for application of Section 1806(f), dismissal on the basis of the state secrets doctrine was inappropriate. This holding says nothing, however, about the relationship between § 1806(f) and the state secrets doctrine dismissal remedy where, as here, a plaintiff has not established that he, she, or it is an "aggrieved person" using admissible record evidence, after a lengthy jurisdictional review process, at the summary judgment stage of the litigation.

of standing implicates state secrets," the holding in *Fazaga* and § 1806(f) do not foreclose "dismissing [the case] on state secrets grounds" at the summary judgment stage of the litigation.[61] *Jewel v. Nat'l Sec. Agency*, No. C 08–04373, at *24 (N.D. Cal. April 25, 2019), *appeal docketed*, No. 19–16066 (9th Cir. May 21, 2019). Accordingly, because plaintiff has not established it is an "aggrieved person" as defined in the statute, and hence § 1806(f) does not apply, and because the issue of standing in this case necessarily implicates state secrets, dismissal of the case is appropriate.

VIII.

To avoid dismissal of the litigation on state secrets grounds, Wikimedia has raised several additional standing arguments separate and apart from the Wikimedia Allegation—namely (i) Upstream surveillance has impaired Wikimedia's communications with its community members, (ii) Upstream surveillance has required Wikimedia to take costly protective measures, and (iii) Wikimedia has third-party standing to assert the rights of its users. Wikimedia's arguments fail as to each of these theories of standing for the reasons discussed below.

First, Wikimedia argues it has standing because Upstream surveillance has impaired Wikimedia's communications with its community members, as evidenced by a drop in the readership of certain Wikipedia pages. In *Clapper* and *Laird*, however, the Supreme Court unequivocally held that "[a]llegations of a subjective 'chill' are not an adequate substitute for a claim of specific present objective harm or a threat of specific future harm." *Clapper v. Amnesty*

---

[61] To be sure, the district court in California did review "classified evidence submitted by Defendants in response to Plaintiffs' discovery requests" pursuant to the procedures of § 1806(f) of FISA prior to its summary judgment ruling dismissing the case. *Id.* at *24-25. That court did not, however, consider the question of whether plaintiffs were "aggrieved persons" prior to undertaking § 1806(f)'s procedures for *in camera* review. Nevertheless, that court still found that where, as here, "the answer to the question of whether a particular plaintiff was subjected to surveillance – *i.e.*, is an 'aggrieved person' under Section 1806(f) – is the very information over which the Government seeks to assert the state secrets privilege," dismissal of the case and entry of judgment in favor of the government is the appropriate action at summary judgment. *Id.* at *23, *25.

Int'l USA, 133 S. Ct. 1138, 1152 (2013) (quoting *Laird v. Tatum*, 408 U.S. 1, 13-14 (1972)). To avoid the conclusion that any drop in readership is the result of a "subjective chill," Wikimedia relies upon a statistical analysis performed by Dr. Jonathon Penny, which concludes it is "highly likely" that "public awareness of NSA surveillance programs, including Upstream surveillance,…ha[s] had a large-scale chilling effect on Wikipedia users" since June 2013. Dr. Jonathon Penney Decl. ¶ 10-11. But Dr. Penney's conclusion that Wikipedia's readership has suffered an actual chill as the result of Upstream surveillance is undermined for two principal reasons. First, Dr. Penney's data only covers a 32-month period which ends in August 2014, before this lawsuit was even filed. Thus, Dr. Penney's evidence, even if reliable, does not say anything about any ongoing harm suffered by Wikimedia that is traceable to Upstream surveillance. Second, these alleged readership effects were from public awareness of "media coverage of NSA surveillance" generally, not Upstream surveillance specifically. *Id.* at ¶ 26. Thus, Dr. Penney's findings do not demonstrate an ongoing and sustained drop in Wikimedia's readership stemming from the NSA's Upstream surveillance program.

Moreover, "a 'chilling effect aris[ing] merely from the individual's knowledge that a governmental agency was engaged in certain activities or from the individual's concomitant fear that, armed with the fruits of those activities, the agency might in the future take some other and additional action detrimental to that individual'" is insufficient to establish standing.[62] *Clapper*, 133 S. Ct. at 1152 (quoting *Laird*, 408 U.S. at 11). This is exactly the situation here—Wikimedia claims that this decreased readership is a result of individual's fear that the government might be

---

[62] It is worth noting that the Fourth Circuit and the Supreme Court have explained that "standing requirements are somewhat relaxed in First Amendment cases." *Cooksey v. Futrell*, 721 F.3d 226, 235 (4th Cir. 2013) (citing *Secretary of State of Md. v. Joseph H. Munson Co., Inc.*, 467 U.S. 947, 956 (1984)). Even though the standing requirements are somewhat relaxed in the First Amendment context, subjective and speculative fears of government surveillance, such as in this case, do not establish Article III standing at summary judgment, as the Supreme Court specifically held in *Clapper* and *Laird*. *See Clapper*, 133 S. Ct. at 1151-52; *Laird*, 408 U.S. at 10-15.

monitoring their Internet activity and might use that information at some later date. Moreover, the Supreme Court has specifically found that a claimed reluctance by third parties to communicate with a plaintiff, due to their subjective fears of surveillance, is not fairly traceable to the alleged surveillance, and is thus foreclosed as a basis for standing. *Clapper*, 133 S. Ct. at 1152 n.7. Accordingly, Wikimedia cannot establish standing under this theory given the Supreme Court's holdings in *Clapper* and *Laird*.

Second, Wikimedia argues it has standing because Upstream surveillance has required Wikimedia to take costly protective measures—namely, transitioning its Internet communications into encrypted formats such as HTTPS and IPSec, acquiring new technical infrastructure, and hiring a full-time engineer to manage the protective measures. The Supreme Court has already foreclosed this alternative theory of standing where, as here, a plaintiff has failed to establish that their communications have been collected by the government, or that such collection is certainly impending. *Clapper*, 133 S. Ct. at 1151. Applicable here is the Supreme Court's statement in *Clapper* that a plaintiff "cannot manufacture standing merely by inflicting harm on themselves based on their fears of hypothetical future harm that is not certainly impending." *Id.*

Wikimedia attempts to distinguish this case from *Clapper* by arguing that the harm Wikimedia faces from Upstream surveillance is well-established, not some "hypothetical future harm." As discussed at length *supra* in Part V, however, the summary judgment record does not establish that Wikimedia's communications have been collected by the NSA during Upstream surveillance, or that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. Thus, any harm to Wikimedia from the Upstream surveillance program remains a purely hypothetical harm

insufficient to establish standing. As the Supreme Court has sensibly observed, to find otherwise "would be tantamount to accepting a repackaged version of [plaintiff's] first failed theory of standing," namely the Wikimedia Allegation. *Id.* (citing *Am. Civil Liberties Union v. Nat'l Sec. Agency*, 493 F.3d 644, 655–56 (6th Cir. 2007)). Accordingly, Wikimedia's alleged expenditures to protect its communications from Upstream surveillance collection do not establish its standing.[63]

Third, Wikimedia argues it has third party standing to assert the rights of its users. In the Fourth Circuit, a plaintiff must demonstrate "(1) an injury-in-fact; (2) a close relationship between [itself] and the person whose right [it] seeks to assert; and (3) a hindrance to the third party's ability to protect his or her own interests" to "overcome the prudential limitation on third-party standing."[64] *Freilich v. Upper Chesapeake Health Inc.*, 313 F.3d 205, 215 (4th Cir. 2002) (citing *Powers v. Ohio*, 499 U.S. 400, 410–11 (1991)). Wikimedia has met none of these requirements. As discussed at length *supra* in Part V, Wikimedia has been unable to establish injury-in-fact in this case. In addition, Wikimedia has not presented admissible evidence that establishes a "close relationship" between Wikimedia and its largely unidentified contributors.[65]

---

[63] Moreover, without evidence that the alleged injuries from implementing these protective measures would be redressed by the injunctive relief plaintiff seeks, these alleged injuries cannot confer standing to sue. *See Clapper*, 568 U.S. at 409; *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560-61 (1992). Given the number of other reasons that plaintiff has admitted influenced its decision to implement these protective measures, including protecting against individual computer hackers and keeping their company policies up-to-date and transparent, injunctive relief enjoining the NSA from conducting the Upstream surveillance program would not redress any alleged injury from these protective expenditures. In fact, Wikimedia began the process of switching to HTTPS as early as 2011, years before any disclosures about the NSA's Upstream surveillance program. *See* ECF No. 178-8.

[64] As the Supreme Court has appropriately warned, "[f]ederal courts must hesitate before resolving a controversy, even one within their constitutional power to resolve, on the basis of the rights of third persons not parties to the litigation." *Singleton v. Wulff*, 428 U.S. 106, 113 (1976).

[65] Close relationships that have established third-party standing in the past include lawyer-client and doctor-patient. *See Department of Labor v. Triplett*, 494 U.S. 715 (1990) (lawyer-client); *Singleton v. Wulff*, 428 U.S. 106 (1976) (doctor-patient). Wikimedia's relationship with its unidentified contributors clearly does not rise to the level of those protected, close relationships.

In fact, Wikimedia has only presented declarations from one single contributor who has edited Wikimedia's web projects while abroad, and this single contributor has stated that her "workload as a medical student" makes it "impossible" for her to bring a lawsuit as a plaintiff.[66] Such "normal burdens of litigation," however, are insufficient to satisfy the third requirement that an obstacle exists that prevents the third party from bringing the lawsuit herself or himself.[67] *See Lawyers Ass'n v. Reno*, 199 F.3d 1352, 1364 (D.C. Cir. 2000). Thus, Wikimedia has also failed to satisfy the third requirement to establish third-party standing. Accordingly, Wikimedia's third-party standing argument clearly fails.

For the reasons stated above, Wikimedia's three additional standing arguments clearly fail because Wikimedia has not established an injury-in-fact using admissible record evidence and Wikimedia has not satisfied the strict requirements to proceed on the basis of third-party standing.

IX.

In sum, Wikimedia has failed to present specific facts which show that defendants, through the Upstream surveillance program, have copied and collected Wikimedia's international Internet communications, that such collection is certainly impending, or that there is a substantial risk that collection will occur such that Wikimedia must incur costs to avoid collection. More specifically, the summary judgment record establishes that it is not a technological necessity that the NSA must copy all of the text-based Internet communications

---

[66] Temple-Wood Decl. ¶ 26, ECF No. 168-10.

[67] Thus, Ms. Temple-Wood, a contributor to Wikimedia's free-knowledge projects, also states that "serving as a plaintiff in a lawsuit would threaten the anonymity [upon which Wikimedia] users depend." Temple-Wood Decl. ¶ 27, ECF No. 168-10. Although privacy and anonymity are valid concerns, in this case a putative plaintiff would not need to reveal the contents of their communications with Wikimedia in order to serve as a plaintiff; they would only need to disclose the form in which the communications were sent (*i.e.*, sending an email or accessing or editing a web project), and the location from which the communications were sent.

traversing a circuit that the NSA monitors while conducting Upstream surveillance. Thus, there is no genuine dispute of material fact that the NSA could conduct Upstream surveillance without collecting any of Wikimedia's communications, and Wikimedia has been unable to present specific facts that establish otherwise, largely because the necessary facts are protected by the state secrets privilege.

Moreover, even if Wikimedia had established a genuine issue of material fact as to whether the NSA has copied or collected any of its international Internet communications, which Wikimedia has not done on this record, further litigation of this matter is precluded by the state secrets doctrine, which has been properly invoked by defendants. The extensive jurisdictional discovery process in this case has made clear that the very issue of standing implicates state secrets and that despite plaintiff's valiant efforts, establishing standing solely on the basis of the public, unclassified record is not possible in this case. Pursuant to Supreme Court and Fourth Circuit precedent, at this stage of the litigation, namely summary judgment post-jurisdictional discovery, dismissal and entry of judgment in favor of defendants is the appropriate, and only available, remedy because the issue of standing in this case necessarily implicates state secrets.

It is important to acknowledge the unfortunate burden that this decision places on Wikimedia. *See Abilt*, 848 F.3d at 317; *Sterling*, 416 F.3d at 348; *El-Masri*, 479 F.3d at 313 ("As we have observed in the past, the successful interposition of the state secrets privilege imposes a heavy burden on the party against whom the privilege is asserted."). Wikimedia suffers dismissal of its claim "not through any fault of [its] own, but because [its] personal interest in pursuing [its] civil claim is subordinated to the collective interest in national security." *El-Masri*, 479 F.3d at 313; *see also Abilt*, 848 F.3d at 318; *Fitzgerald*, 776 F.2d at 1238 n.3 ("When the state secrets privilege is validly asserted, the result is unfairness to individual litigants—through the loss of

important evidence or dismissal of a case—in order to protect a greater public value."). It is appropriate, however, "in limited circumstances like these, [that] the fundamental principle of access to court must bow to the fact that a nation without sound intelligence is a nation at risk." *Sterling*, 416 F.3d at 348.

Plaintiff contends that a holding which finds plaintiff does not have standing and precludes further litigation of this matter because of defendants' invocation of the state secrets doctrine leads to the result that "the Executive Branch alone controls who can and cannot challenge unlawful surveillance."[68] This contention is incorrect; the Supreme Court addressed and rejected a similar argument in *Clapper*. There, the Supreme Court explained that Section 702 surveillance orders are not insulated from judicial review because (i) the FISC reviews the government's certifications, targeting procedures, and minimization procedures for Section 702 surveillance, including whether the targeting and minimization procedures comport with the Fourth Amendment, (ii) criminal defendants prosecuted on the basis of information derived from Section 702 surveillance are given notice of that surveillance and can challenge its validity, and (iii) electronic communications service providers directed to assist the government in surveillance may challenge the directive before the FISC. *Clapper v. Amnesty Int'l USA*, 133 S. Ct. 1138, 1154 (2013). Even if those other avenues for judicial review were not available, the Supreme Court has made clear that "'[t]he assumption that if [plaintiff has] no standing to sue, no one would have standing, is not a reason to find standing.'" *Id.* (quoting *Valley Forge Christian College v. Americans United for Separation of Church and State, Inc.*, 454 U.S. 464, 489 (1982)).

Moreover, since this litigation began in 2015, FISA Section 702, pursuant to which the

---

[68] Plaintiff's Br. in Op. to Defs.' Motion for Summary Judgment, ECF No. 168, at 2.

NSA Upstream surveillance program operates, was reauthorized by Congress. FISA Section 702 was set to expire on December 31, 2017, but Congress voted in January 2018 to extend FISA Section 702 for an additional six years (the "FISA Amendment Reauthorization Act of 2017").[69] This reauthorization process sparked significant public debate, and the FISA Amendment Reauthorization Act of 2017 enacted a number of reforms to address the public's civil liberties concerns.[70]

Thus, rather than the executive branch alone controlling who can and cannot challenge unlawful surveillance, the judicial branch provides for review and oversight via the limited avenues outlined by the Supreme Court in *Clapper*, including the significant role of the FISC, and the legislative branch provides for review and oversight via the FISA reauthorization process and the executive branch's ongoing reporting requirements to Congress. These avenues are sufficient to meet Constitutional requirements while at the same time precluding the unnecessary disclosure of state secrets.

* * *

For the reasons set forth above, this case must be dismissed, and judgment must be entered for defendants.

An appropriate order will issue separately.

---

[69] FISA Amendments Reauthorization Act of 2017, PL 115-118, January 19, 2018, 132 Stat 3.

[70] For example, the FISA Amendment Reauthorization Act of 2017 added a requirement that the DNI adopt procedures consistent with the requirements of the Fourth Amendment for querying information collected pursuant to Section 702 authority and made these querying procedures subject to FISC review. *See id.* at Sec. 101 Querying Procedures Required. The FISA Amendment Reauthorization Act of 2017 also restricted the use of U.S. person information obtained under Section 702 as evidence in a criminal proceeding and amended the mandatory reporting requirements to require the release of information on the breakdown of U.S. and non-U.S. person targets of electronic surveillance. *See id.* at Sec. 102. These represent only a few of a number of reforms enacted by the FISA Amendment Reauthorization Act of 2017. These reforms, combined with the short period of reauthorization, demonstrate the legislative branch's focused oversight of the executive branch's Section 702 authority.

The Clerk is directed to provide a copy of this Opinion to all counsel of record.

Alexandria, Virginia
December 13, 2019

/s/
T. S. Ellis, III
United States District Judge

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION,	)	
Plaintiff,	)	
	)	
v.	)	**Case No. 1:15-cv-662**
	)	
NATIONAL SECURITY AGENCY/ CENTRAL SECURITY SERVICE, *et al.*,	)	
	)	
Defendants.	)	

ORDER

For the reasons stated in the accompanying Memorandum Opinion,

It is hereby **ORDERED** that defendants' motion for summary judgment (Dkt. 161) is **GRANTED.**

Accordingly, it is further **ORDERED** that this matter is **DISMISSED without prejudice.**

The Clerk is directed to enter Rule 58 judgment on behalf of defendants and against plaintiff and place this matter among the ended causes.

The Clerk is further directed to send a copy of this Order to all counsel of record.

Alexandria, Virginia
December 13, 2019

/s/
T. S. Ellis, III
United States District Judge

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF MARYLAND

WIKIMEDIA FOUNDATION, *Plaintiff*, v. NATIONAL SECURITY AGENCY, *et al.*, *Defendants*.	Hon. T.S. Ellis, III Civil Action No. 15-cv-662-TSE

NOTICE OF APPEAL

Wikimedia Foundation, Plaintiff in the above-captioned case, hereby appeals to the United States Court of Appeals for the Fourth Circuit from the final judgment entered in this action on the 17th day of December, 2019, granting Defendants' motion for summary judgment.

Dated: February 14, 2020

Respectfully submitted,

/s/ Deborah A. Jeon
Deborah A. Jeon (Bar No. 06905)
David R. Rocah (Bar No. 27315)
AMERICAN CIVIL LIBERTIES UNION FOUNDATION OF MARYLAND
3600 Clipper Mill Rd., #350
Baltimore, MD 21211
Phone: (410) 889-8555
Fax: (410) 366-7838
jeon@aclu-md.org

Benjamin H. Kleine (pro hac vice)
COOLEY LLP
101 California Street, 5th Floor
San Francisco, CA 94111
Phone: (415) 693-2000
Fax: (415) 693-2222
bkleine@cooley.com

Counsel for Plaintiff

/s/ Patrick Toomey
Patrick Toomey (pro hac vice)
(*signed by Patrick Toomey with permission of Deborah A. Jeon*)
Ashley Gorski (pro hac vice)
AMERICAN CIVIL LIBERTIES UNION FOUNDATION
125 Broad Street, 18th Floor
New York, NY 10004
Phone: (212) 549-2500
Fax: (212) 549-2654
ptoomey@aclu.org

Alex Abdo (pro hac vice)
Jameel Jaffer (pro hac vice)
KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY
475 Riverside Drive, Suite 302
New York, NY 10115
Phone: (646) 745-8500
alex.abdo@knightcolumbia.org