Before the
Federal Communications Commission
Washington, DC 20554

In the Matter of	)	
Wireline Competition Bureau Seeks Comment	)	
on Petition of Public Knowledge for	)	
Declaratory Ruling that Section 222 of the	)	WC Docket No. 13-306
Communications Act Prohibits	)	
Telecommunications Providers from Selling	)	
Non-Aggregate Call Records Without	)	
Customers' Consent	)	

COMMENTS OF THE FUTURE OF PRIVACY FORUM

Jules Polonetsky
Co-Chair and Director

Christopher Wolf
Founder and Co-Chair
Future of Privacy Forum
919 18th Street, NW Suite 901
Washington, DC 20006
(202) 713-9466

Mark W. Brennan
Partner
Hogan Lovells US LLP
555 13th Street, NW
Washington, DC 20004
(202) 637-6409
mark.brennan@hoganlovells.com

January 17, 2014

TABLE OF CONTENTS

Before the
Federal Communications Commission
Washington, DC 20554

In the Matter of	)	
Wireline Competition Bureau Seeks Comment	)	
on Petition of Public Knowledge for	)	
Declaratory Ruling that Section 222 of the	)	WC Docket No. 13-306
Communications Act Prohibits	)	
Telecommunications Providers from Selling	)	
Non-Aggregate Call Records Without	)	
Customers' Consent	)	

COMMENTS OF THE FUTURE OF PRIVACY FORUM

I. INTRODUCTION AND SUMMARY

The Future of Privacy Forum ("FPF") submits these comments in response to the Federal Communications Commission ("Commission") Wireline Competition Bureau's December 18, 2013 Public Notice in the above-captioned proceeding,[1] which seeks comment on a Petition for Declaratory Ruling filed by Public Knowledge, et al. ("Public Knowledge").[2] In the Petition, Public Knowledge requests that the Commission clarify that "anonymized" or "deidentified" but non-aggregate call records constitute individually identifiable "customer proprietary network information" ("CPNI") under Section 222 of the Communications Act.[3]

---

[1] *Wireline Competition Bureau Seeks Comment on Petition of Public Knowledge for Declaratory Ruling that Section 222 of the Communications Act Prohibits Telecommunications Providers from Selling Non-Aggregate Call Records Without Customers' Consent*, WC Docket No. 13-306, Public Notice, DA 13-2415 (rel. Dec. 18, 2013) ("*Notice*").

[2] *See Public Knowledge, et al.*, Petition for Declaratory Ruling, WC Docket No. 13-306, (filed Dec. 11, 2013) ("Petition").

[3] *See* 47 U.S.C. § 222.

FPF submits these comments to address the Petition's argument that all anonymized records[4] must be considered "personally identifiable" records because there have been instances in which some publicly available, anonymized records have been reidentified.[5] Although reidentification may be possible in some specific circumstances, when proper anonymization practices are used, anonymization is a valuable and effective way to advance the goal of protecting individual privacy while allowing for beneficial uses of data. As the Federal Trade Commission ("FTC") and others have recognized, the mere mathematical possibility that an anonymized data set may be reidentified should not suffice to make the data set "individually identifiable" – that extreme approach would be contrary to expert findings and statements from regulators in in the United States and abroad. Instead, whether a data set is individually identifiable should require a determination of whether, in the context in which the specific data will be used, there is a reasonable basis to believe that the information can be used to identify an individual.

II. ABOUT THE FUTURE OF PRIVACY FORUM

FPF is a Washington, DC-based think tank founded in 2008 and focused on advancing responsible data practices. FPF is led by privacy leaders Jules Polonetsky and Christopher Wolf and includes an advisory board comprised of leading figures from industry, academia, law, and advocacy groups.[6]

---

[4] For purposes of these comments and to reflect Public Knowledge's usage, FPF defines "anonymized records" as records that have been stripped of personal identifiers but may still retain individual characteristics (*e.g.*, an individual's call records that have been stripped of the individual's name and number but retain information about incoming and outgoing calls, call durations, and call times).

[5] Petition at 6-8.

[6] The positions taken by FPF are entirely its own and do not necessarily reflect those of its supporters and advisory board members.

FPF seeks to identify and develop leading practices for the promotion of consumer privacy and has significant experience working with stakeholders on anonymization practices. For example:

- FPF has launched a De-ID Project to focus on the de-identification landscape;[7]
- On December 5, 2011, FPF held a workshop that brought together academics, advocates, chief privacy officers, and policy makers to identify leading practices;[8]
- In January 2013, FPF issued a white paper explaining that anonymization can promote privacy and that effectively anonymized data should not be considered "personal data" under the European Union's proposed General Data Protection Regulation;[9] and
- On September 13, 2013, Stanford Law Review Online published an article co-authored by FPF Co-Chair Jules Polonetsky and FPF Legal and Policy Fellow Yianni Lagos, which demonstrated that effective anonymization practices can greatly reduce privacy risks.[10]

III. PUBLIC KNOWLEDGE OVERSTATES THE RISKS OF REIDENTIFICATION

Public Knowledge argues that because researchers have been able to reidentify some publicly disclosed data sets that were purged of personally identifying information, all datasets that have been purged of personally identifying information must necessarily be considered individually identifiable. Logically, this argument is flawed. It is analogous to the argument that because some locks have been broken, there is no such thing as a reasonably secure door.

---

[7] *De-identification*, Future of Privacy Forum, http://www.futureofprivacy.org/de-identification/ (last visited Jan. 16, 2014).

[8] *De-Identification Workshop*, Future of Privacy Forum, http://www.futureofprivacy.org/de-identification-workshop/ (last visited Jan. 16, 2014).

[9] Omer Tene & Christopher Wolf, *Future of Privacy Forum, The Definition of Personal Data: Seeing the Complete Spectrum* (2013), *available at* http://www.futureofprivacy.org/wp-content/uploads/FINAL-Future-of-Privacy-Forum-White-Paper-on-De-Id-January-201311.pdf.

[10] Yianni Lagos & Jules Polonetsky, *Public vs. Nonpublic Data: The Benefits of Administrative Control*, 66 Stan L. Rev. Online 103 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/public-vs-nonpublic-data.

Public Knowledge's argument that all anonymized data sets should be considered personally identifiable relies on three studies in which researchers were able to reidentify anonymized public data sets[11] and a recent study that illustrated the uniqueness of mobile location histories but did not involve the reidentification of individuals.[12] This research illustrates that some anonymization practices may allow for the reidentification of data in certain circumstances, but Public Knowledge overstates the degree to which data sets may be reidentified.

As an example, all of the reidentification studies cited by Public Knowledge involved anonymized data sets that were publicly disclosed. Publicly disclosed data sets are at the greatest risk for reidentification,[13] and focusing on the reidentification risks associated solely with publicly disclosed, anonymized data sets leads to an overestimation of the risks associated with reidentification.[14] As discussed in Section III below, effective, contemporary anonymization

[11] Latanya Sweeney, was able to reidentify publicly disclosed, anonymized data sets containing hospital discharge data. Latanya Sweeney, *Simple Demographics Often Identify People Uniquely*, Carnegie Mellon University, Data Privacy Working Paper 3, *available at* http://dataprivacylab.org/projects/identifiability/paper1.pdf (last visited Jan. 14, 2014). Researchers at the University of Texas at Austin identified Netflix subscribers from a publicly disclosed, anonymized data set. Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets,* Proceedings of the 2008 IEEE Symposium on Security and Privacy, 111 (2008). And researchers reidentified some publicly disclosed, anonymized search records relating to AOL subscribers in 2006. Michael Barbaro & Tom Zeller Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. Times, Aug. 9, 2006, *available at* http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted= 1&_r=0.

[12] Yves-Alexandre de Montjoye et al., *Unique in the Crowd: The Privacy Bounds of Human Mobility,* 3 Sci. Rep. (Article 1376) 1 (2013), *available at* http://www.nature.com/srep/2013/130325/srep01376/full/srep01376.html.

[13] Lagos & Polonetsky, *supra* note 10.

[14] *See id.*

practices "drastically reduce[] the risk that personal information will be used or disclosed for unauthorized or malicious purposes."[15]

The Petition also includes the erroneous claim that "[anonymized] records often contain sufficient information to discover the true identity of the person whose records they are."[16] This is not the case. Reidentification of anonymized records requires the use of additional information. For example, Public Knowledge offers a hypothetical in which an individual is identified by spotting frequent calls made to the individual's mother.[17] But in the hypothetical, the anonymized records are not alone sufficient to identify the individual. To identify Doe in the Petition's hypothetical, one needs to review the anonymized phone records and combine that information with the knowledge that Doe calls his mother a lot and that the frequently called number in the anonymized records is associated with Doe's mother. The reidentification requires outside information. On their own, the anonymized records show only that some individual calls a particular number quite often – there is no inherent reason why the number called must belong to Doe's mother, or even to a family member.

The studies cited by Public Knowledge also illustrate that reidentifying anonymized records requires information beyond the anonymized records themselves. Contrary to what is stated in the Petition, the recent study involving anonymized location data did not show that "95 percent of individual users could be uniquely identified using just four location data points."[18] The study actually concludes that 95% of location histories could be uniquely specified using

---

[15] Ann Cavoukian & Khaled El Emam, *Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy* 4 (2011), *available at* http://www.ipc.on.ca/images/Resources/anonymization.pdf.

[16] Petition at 7

[17] *Id.*

[18] *Id.* at 7-8.

just four data points.[19] As the study notes, identifying the individuals associated with those location histories would still require "outside information."[20] In addition, Latanya Sweeney demonstrated that 87% of the U.S. population can be uniquely specified by 5-digit ZIP, gender, and date of birth.[21] However, to identify who those specified individuals were, Sweeney had to use additional information (*e.g.*, voter registration records).[22] When researchers at the University of Texas at Austin reidentified Netflix subscribers from a publicly released dataset, they used additional information obtained from the Internet Movie Database.[23] The reidentification of AOL users also required additional information.[24]

To be sure, when anonymized data sets are disclosed publicly, in some cases it may require little outside information to reidentify anonymized data sets.[25] However, as discussed in the next section, today's sophisticated anonymization tools can make reidentification unlikely and difficult even with publicly data sources at hand. For example, only certain information will provide the "key" to reidentification, and those who attempt to reidentify anonymized data sets must determine what specific additional information is required and then obtain access to that information. This can be a very difficult task. When appropriate safeguards are used along with sophisticated deidentification tools, it can be extremely difficult to discover the "key" – and the risks of reidentification can be drastically reduced.[26]

---

[19] de Montjoye et al., *supra* note 12, at 2.

[20] *Id.*

[21] Sweeney, *supra* note 11, at 2.

[22] *Id.*

[23] Narayanan & Shmatikov, *supra* note 11.

[24] Barbaro & Zeller Jr., *supra* note 11.

[25] *Id.*

[26] Cavoukian & El Emam, *supra* note 15, at 13-14; Lagos & Polonetsky, *supra* note 10.

IV. APPROPRIATE ANONYMIZATION PRACTICES CAN PREVENT THE REIDENTIFICATION OF INDIVIDUALS

The value and effectiveness of appropriate anonymization procedures have been confirmed and espoused by various experts and regulatory bodies. The FTC, for example, recognized the value of anonymization in its 2012 privacy report.[27] Specifically, the FTC's proposed privacy framework does not apply to data that has been reasonably anonymized.[28] In determining whether data has been reasonably anonymized, the FTC proposes looking to whether the technical measures reasonably anonymize the data set and whether appropriate administrative safeguards are implemented.[29]

In 2012, the United Kingdom Information Commissioner's Office ("the ICO") published a guide on anonymization.[30] The ICO wrote that "the effective anonymization of personal data is possible, desirable, and can help society make rich data resources available whilst protecting individuals' privacy."[31] Although the ICO acknowledges that there may be some circumstances in which purportedly anonymized data can be reidentified,[32] the ICO concluded "that adopting

---

[27] FTC, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers* 31 (2012), *available at* http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

[28] *Id.* at 20.

[29] *Id.* at 22.

[30] Information Commissioner's Office, *Anonymisation: Managing Data Protection Risk, Code of Practice* (2012), *available at* http://ico.org.uk/for_organisations/data_protection/topic_guides/~/media/documents/library/Data_Protection/Practical_application/anonymisation-codev2.pdf.

[31] *Id.* at 7.

[32] *See, e.g., id.* at 19.

the techniques and procedures recommended in [the ICO's] code will guard against re-identification."[33]

The value of anonymization has also been recognized by Ontario, Canada's Privacy & Information Commissioner, Ann Cavoukian. As Cavoukian stated in a recent report co-authored with Professor Khaled El Emam, anonymization is "a valuable and important mechanism in protecting personal data, and must not be abandoned."[34] In that report, Cavoukian and El Emam addressed criticisms of the utility of anonymization as a tool to protect privacy (including the research of Latanya Sweeney and the Netflix reidentification discussed above). Cavoukian and El Emam noted that "contrary to what has been suggested in recent articles, re-identification of properly de-identified information is not in fact an 'easy' or 'trivial' task."[35] They cite a 2011 study concluding that "the evidence indicates that there are few cases in which properly de-identified data have been successfully reidentified."[36]

FPF agrees with the FTC, the ICO, and Ontario's Privacy & Information Commissioner that anonymization should be assessed in context, and that when appropriate technical measures are combined with reasonable safeguards, anonymization protects the privacy of individuals. The mere mathematical possibility that an anonymized data set could be reidentified should not suffice to make that data set "individually identifiable." Instead, whether a data set is individually identifiable should require a determination of whether,

[33] *Id.* at 27.

[34] Cavoukian & El Amam, *supra* note 15, at 4

[35] *Id.* at 1.

[36] *Id.* at 6 (citing Khalid El Emam et al., *A Systematic Review of Re-Identification Attacks on Health Data*, subsequently published in PLoS One 6:12 (2011), *available at* http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0028071.

in the context in which the data will be used, there is a reasonable basis to believe that the information can be used to identify an individual.

The Commission has adopted a similar "reasonableness" approach in its rule regarding the safeguarding of CPNI. Section 47 C.F.R. § 64.2010(a) of the Commission's rules requires telecommunications carriers (including interconnected VoIP providers) to take "reasonable measures to discover and protect against attempts to gain unauthorized access to CPNI."[37] Safeguards can be reasonable even if there is a mathematical possibility that they can be circumvented. Similarly, anonymization practices can be reasonable and effective even if there is a mathematical possibility that the anonymized data can be reidentified.

Even when only technical measures are used, anonymization can significantly reduce the risks of reidentification. For example, in a 2009 study performed for the U.S. Department of Health and Human Services Office of the National Coordinator for Health Information Technology, statistical experts assessed the likelihood of reidentifying records that were anonymized under Health Insurance Portability and Accountability Act standards. [38] The experts used commercially available data sources and were able to reidentify only 2 out of 15,000 individuals (or 0.013%).[39] Other research involving sophisticated anonymization tools have led to similar results.[40]

When these technical measures are used along with administrative safeguards, the privacy protections are multiplied. Administrative safeguards include such protections as

[37] 47 C.F.R. § 64.2010(a) (emphasis added).

[38] Deborah Lafkey, *The Safe Harbor Method of De-Identification: An Empirical Test*, ONC Presentation, Oct. 8, 2009, *available at* http://www.ehcca.com/presentations/HIPAAWest4/lafky_2.pdf.

[39] *Id.*

[40] Cavoukian & El Amam, *supra* note 15, at 6-7.

internal controls restricting access to data sets (*e.g.*, security policies and access restrictions) and contractual terms that restrict how the recipients of data may use and share information. Administrative safeguards reduce the likelihood that data will be accessed or used without authorization. For example, "[w]here data are maintained by a controller or shared with a restricted group of service providers or business associates, additional safeguards–administrative and legal–beyond technical de-identification can prevent re-identification."[41]

A statistical example serves to illustrate the effectiveness of using both administrative safeguards and technical anonymization measures: suppose that the implementation of certain administrative safeguards reduces the risk of reidentification to 0.05% (*i.e.*, there is a 1 out of 2,000 chance that an individual will be reidentified). Suppose further that the technical deidentification measures that are used (*e.g.*, the manner in which personal information is deleted, scrubbed, or obscured) also reduce the risk of reidentification to 0.05%. If the probability of reidentification based on technical and administrative safeguards are independent of each other, the likelihood of reidentification in these circumstances is 1 in 4 million.[42] In those circumstances, it would not be reasonable to believe that the information could be used to identify an individual. Combining effective technical measures with effective administrative safeguards, therefore, effectively promotes privacy.

As mentioned above, the determination of whether data are anonymized should depend on context, not the mere mathematical possibility of reidentification. In the cases where individuals were identified, the data sets involved were publicly disclosed. "Non-publicly disclosed datasets have a lessened risk of re-identification than publicly disclosed datasets due to

[41] Tene & Wolf, *supra* note 9, at 7.

[42] See Lagos & Polonetsky, *supra* note 10 for further discussion of the interconnection between technical de-identification procedures and administrative safeguards.

the added protection of administrative controls."[43] When appropriate administrative safeguards (*e.g.*, access controls and restrictions, contractual data use restrictions, and data deletion protocols) are used in conjunction with sophisticated anonymization techniques, reidentification "remains a relatively difficult task," and "in the vast majority of cases, de-identification will protect the privacy of individuals." [44]

V. EFFECTIVE ANONYMIZATION ALLOWS FOR USES OF DATA THAT SUBSTANTIALLY BENEFIT OUR SOCIETY

As the ICO has stated, "the effective anonymization of personal data is possible, desirable, and can help society to make rich data resources available whilst protecting individuals' privacy."[45] The societal value of anonymized data is immense, and anonymization is integral to a wide range of business models.[46] For example, anonymized data drives innovation in "healthcare, energy conservation, fraud prevention, and data security."[47] The availability of anonymized data for health research has led to important discoveries regarding "disease trends, risk factors, outcomes of treatment, and patterns of care."[48] In addition, technology companies are using anonymization to increase the security of cloud infrastructure. [49] As an example, when United Nations Global Pulse made an anonymized mobile telephony data

---

[43] *Id.*

[44] Cavoukian & El Emam, *supra* note 15, at 15.

[45] ICO, *supra* note 30, at 7.

[46] Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 Stan. L. Rev. Online 63 (2012), http://www.stanfordlawreview.org/online/privacy-paradox/big-data.

[47] Tene & Wolf, *supra* note 9, at 4.

[48] Cavoukian & El Emam, *supra* note 15, at 4-5.

[49] *See* Jeff Sedayao, Enterprise Architect, Intel IT, *Enhancing Cloud Security Using Data Anonymization* (June 2012), *available at* http://www.intel.com/content/dam/www/public/us/en/documents/best-practices/enhancing-cloud-security-using-data-anonymization.pdf.

set available for a research challenge, participants showed that the data could be used to improve transport infrastructure, analyze social divisions, and contain the spread of disease.[50]

VI. CONCLUSION

The Commission should recognize the value of effective anonymization practices and the privacy protections that such practices offer. Contrary to the arguments in the Petition, reidentifying anonymized data sets is not an easy task and requires the use of additional information. The FTC and privacy regulators from other countries recognize that contemporary anonymization techniques can be effective tools that promote privacy by preventing the reidentification of individuals. When effective anonymization technical tools are used in combination with appropriate administrative safeguards, the risks of reidentification can be drastically reduced. Moreover, the use of anonymized data redounds to the benefit of individuals and society as a whole, and stakeholders can use deidentified data to serve a wide range of business models and research programs while promoting privacy.

Respectfully submitted,

/s/ Jules Polonetsky
Jules Polonetsky
Co-Chair and Director

Christopher Wolf
Founder and Co-Chair
Future of Privacy Forum
919 18th Street, NW Suite 901
Washington, DC 20006
(202) 713-9466

Mark W. Brennan
Partner
Hogan Lovells US LLP
555 13th Street, NW
Washington, DC 20004
(202) 637-6409
mark.brennan@hoganlovells.com

January 17, 2014

[50] Jennifer Poole, *Winning Research from the Data 4 Development Challenge*, U.N. Global Pulse (May 6, 2013), http://www.unglobalpulse.org/D4D-Winning-Research (last visited Jan. 17, 2014).