Before the
Federal Communications Commission
Washington, D.C. 20554

In the Matter of	)	
	)	
Public Knowledge et al. Petition for Declaratory Ruling	)	WC 13-306
	)	

COMMENTS OF THOMAS M. LENARD, PH.D
PRESIDENT AND SENIOR FELLOW, TECHNOLOGY POLICY INSTITUTE

These comments are in response to a petition by selected public interest groups asking the Commission to "clarify that under Section 222 of the Communications Act, 'anonymized' or 'de-identified' but non-aggregate call records constitute individually identifiable customer proprietary network information ('CPNI'), and must not be sold to or otherwise shared with third parties without customers' consent."[1]

According to the petitioners, major carriers believe it is consistent with Section 222 to share data in either anonymous or aggregate form without customers' consent, although the petitioners do not know whether the carriers currently do so.[2] The petitioners argue that carriers should only be able to share aggregated data.

The petitioners make two arguments. They first argue that in Section 222 of the Communications Act, the "structure and definition of 'aggregate customer information' indicate under this section, 'individually identifiable' means 'not aggregate'."[3] Not being a lawyer, I do not address this legal argument.

Second, the petitioners argue that "even when carriers have 'anonymized' or 'de-identified' call records by removing personal identifiers from them they still constitute identifiable CPNI."[4] They add, "Even if 'individually identifiable' were interpreted to mean personally identifiable, 'anonymized' call records must still fall into this category because in many cases sufficient information remains in anonymized records to link them back to individual people."[5]

Thus, the petitioners seem to be arguing that it is not possible to anonymize data. Some recent privacy literature is also consistent with this assertion. As Jane Yakowitz observes, "Privacy advocates, the media, and the Federal Trade Commission ('FTC') have accepted uncritically the notion that anonymization is impossible, and they advocate for the wholesale dismantling of the

---

[1] Public Knowledge et al., *Petition for Declaratory Ruling*, WC 13-306 (filed December 11, 2013) at 1.
[2] *Id.* at 9.
[3] *Id. at* 3.
[4] *Id.* at 3.
[5] *Id.* at 6.

concept of anonymization."[6] However, the notion that data cannot be anonymized does not stand up to scrutiny.

As part of their argument, the petitioners point to a well-known 2000 study by Professor Latanya Sweeney, currently Chief Technologist at the FTC, that re-identified Massachusetts Governor Weld's medical records using hospital data that contained a 5-digit zip code, full birth date and gender.[7] Dr. Sweeney found that 87 percent of the U.S. population could be uniquely identified with these three pieces of data. However, this study only shows that pre-HIPAA data are an example of ineffective anonymization.

Indeed, Dr. Sweeney's subsequent research shows something quite different.[8] "Data properly de-identified under the requirements of HIPAA are quite robust against re-identification attacks. In contrast to the high resolution health data that allowed Dr. Sweeney to re-identify Governor Weld fifteen years ago, Dr. Sweeney estimates that HIPAA-compliant data reporting a patient's gender, year of birth (rather than full birth date), and 3-digit zip code (rather than 5-digit) produces a re-identification risk of only 0.04 percent. That is, only four people in 10,000 have a unique combination of gender, age in years, and 3-digit zip code."[9]

The petitioners also refer to a widely cited study in which researchers identified Netflix subscribers from an anonymized Netflix database. However, as Yakowitz and Barth-Jones note, "the applicability of this study is much more limited than the privacy advocates suggest."[10] This "proof of concept" study was able to re-identify only two individuals (out of a few dozen), and the authors did not report verifying those matches.

Importantly, the petitioners do not analyze the carriers' methods of anonymizing CPNI data or in any way show that their methods are deficient. The petition contains no examples of CPNI having been successfully re-identified or misused. In fact, "as of today, there is no evidence that

---

[6] Jane Yakowitz, *Tragedy of the Data Commons*, Harvard Journal of Law & Technology, v. 25, no. 1, Fall 2011, at 3, available at: http://jolt.law.harvard.edu/articles/pdf/v25/25HarvJLTech1.pdf.

[7] *See* Latanya Sweeny, *Uniqueness of Simple Demographics in the U.S. Population,* Carnegie Mellon University, School of Computer Science, Data Privacy Laboratory, Technical Report LIDAP-WP4, 2000

[8] *See footnote 16* of National Committee on Vital and Health Statistics Report to the Secretary of the U.S. Department of Health and Human Services, *Enhanced Protections for Uses of Health Data: A Stewardship Framework for "Secondary Uses" of Electronically Collected and Transmitted Health Data"*, December 19, 2007, at 36, available at: www.ncvhs.hhs.gov/071221lt.pdf.

[9] Jane Yakowitz and Daniel Barth-Jones, *The Illusory Privacy Problem in Sorrell v. IMS Health*, Technology Policy Institute, May 2011, available at: http://www.techpolicyinstitute.org/files/the%20illusory%20privacy%20problem%20in%20sorrell1.pdf.

[10] *Id.* at 6.

re-identification [of any data set] by a true adversary (somebody other than a researcher or journalist interested in the efficacy of privacy protections) has actually happened. This is not particularly surprising when one considers the skill and effort required to launch a deanonymization attack on a properly anonymized dataset."[11]

The proper method of anonymizing data is a legitimate topic for discussion, but the notion that anonymization is not possible or that any reasonably competent computer science graduate student can easily re-identify anonymized data is not correct.

Moreover, acceptance of the notion that anonymization is impossible would adversely affect research across a broad array of subjects that rely on anonymous data as well as commercial activity.[12] Thus, the issue presented by this petition has implications that are broader than the sharing of CPNI. Whatever the Commission decides, it should not base its decision on the notion that data cannot be de-identified.

Respectfully Submitted,

Thomas M. Lenard, President
Technology Policy Institute
1099 New York Avenue, NW, Suite 520
Washington, DC 20001
(202) 828-4405
tlenard@techpolicyinstitute.org

[11] *Id.* at 7.
[12] *See* Yakowitz, Section II, at 5-17.