

**BEFORE THE PUBLIC UTILITIES COMMISSION
OF THE STATE OF CALIFORNIA**



FILED

12-17-12
04:59 PM

Order Instituting Rulemaking to Consider Smart Grid Technologies Pursuant to Federal Legislation and on the Commission's own Motion to Actively Guide Policy in California's Development of a Smart Grid System.

Rulemaking 08-12-009
(Filed December 18, 2008)

**OPENING COMMENTS OF THE ELECTRONIC
FRONTIER FOUNDATION ON ENERGY DATA CENTER**

Lee Tien
ELECTRONIC FRONTIER FOUNDATION
454 Shotwell Street
San Francisco, CA 94110
Telephone: (415) 436-9333 x 102
Facsimile: (415) 436-9993
E-Mail: tien@eff.org

Counsel for
ELECTRONIC FRONTIER FOUNDATION

Jennifer M. Urban
Assistant Clinical Professor of Law
SAMUELSON LAW, TECHNOLOGY & PUBLIC
POLICY CLINIC
UC-Berkeley School of Law
585 Simon Hall
Berkeley, CA 94720-7200
Telephone: (510) 642-7338
E-mail: jurban@law.berkeley.edu

Counsel for
SAMUELSON LAW, TECHNOLOGY & PUBLIC
POLICY CLINIC

Dated: December 17, 2012

**BEFORE THE PUBLIC UTILITIES COMMISSION
OF THE STATE OF CALIFORNIA**

Order Instituting Rulemaking to Consider Smart Grid Technologies Pursuant to Federal Legislation and on the Commission’s own Motion to Actively Guide Policy in California’s Development of a Smart Grid System.

Rulemaking 08-12-009
(Filed December 18, 2008)

**OPENING COMMENTS OF THE ELECTRONIC
FRONTIER FOUNDATION ON ENERGY DATA CENTER**

I. INTRODUCTION

The Electronic Frontier Foundation (“EFF”)¹ files these comments pursuant to the “Assigned Commissioner’s Scoping Memo and Ruling Amending Scope of Proceeding to Seek Comments and to Schedule Workshops on Energy Data Center,” (“Ruling”) filed on November 13, 2012. Consistent with the directions in section 12 of the Ruling, EFF understands these comments to establish party status in the proceeding without the need to request party status in a separate motion.

Energy data is surely a rich resource for researchers who wish to explore important questions about energy efficiency, sustainability, and technical solutions to energy problems, among others. As the Commission found in Decision (D.) 11-07-056, however, laudable and important societal energy goals can and must be met while protecting the privacy of Californians. Because the same richness that makes detailed energy usage data promising for research also makes it highly revealing of activities within a premises, the Commission should examine both the risks and benefits of making energy usage data available to researchers, through the proposed energy data center (“EDC”) or otherwise, and use the information gained through that examination to establish privacy and security safeguards that are sufficient to protect Californians’ data security and privacy interests before releasing the data.

¹ EFF is a non-profit member-supported organization based in San Francisco, California, that works to protect free speech and privacy rights in an age of increasingly sophisticated technology.

Specifically, EFF believes that the Commission should closely examine whether researchers' interests can be met without compelling the transfer of energy usage data from utilities that have clear legal responsibilities for protecting such data's privacy and security to a separate, centralized data repository. As we note below, state law imposes significant restrictions on research use of state-held personal data precisely because of previous failures to safeguard such data. EFF is concerned that it may not be possible for the proposed EDC to function as envisioned without harm to customer privacy and data security unless significant technical, legal and administrative safeguards are employed.

Put another way, the key question raised by the EDC proposal is not the merits of an EDC *per se*, but the proper framework for researcher access to energy usage data given the recognized need for meaningful privacy and security safeguards as well as respect for consumers' rights over their personal data. As discussed in more detail below, established state law and policy already speaks to many aspects of this question. But even the best law and policy cannot be properly applied without a solid understanding of the facts. For instance, the current baseline of researcher access to energy data should be clarified: what are current utility practices for researcher access, what problems do these practices pose for researchers, how satisfactory are these practices from the perspective of privacy and security, and how might they be improved? In EFF's view, the proposed EDC must be evaluated in comparison to existing utility practices (and how those practices could be standardized or improved).

Accordingly, EFF urges the Commission to focus on what technical, legal, and administrative protocols should be required for research access to and use of energy usage data, regardless of the method used to make the data available.

EFF's comments are necessarily preliminary and will be updated and revised as appropriate as the Commission moves forward with workshops to address the concept of an EDC in more detail.

II. BACKGROUND

EFF was an active participant in the earlier phase of this proceeding,² in which the Commission undertook a careful review of the privacy issues presented by energy usage data collected by smart meters, and adopted a significant and pioneering privacy and security framework for energy usage data based on the “Fair Information Practice” principles (FIPs)³— and then promulgated balanced, privacy-protective rules for disseminating and using energy usage data. Among other things, the Commission’s rules require: customer consent for disclosure of customer-specific energy usage data for a purpose that is unrelated to the electric or natural gas services provided by the utility; minimization of customer-specific energy usage data that is collected, disclosed or shared for any purpose; notification of utility customers prior to the disclosure or use of their private customer-specific energy usage data for purposes unrelated to the utility services that are directly provided to them; and compliance with information security standards.

EFF commends the Commission for recognizing that the proposed EDC poses privacy and security issues surrounding the dissemination and use of energy usage data given that:

The information generated by smart meters creates individual privacy concerns because household energy consumption, particularly when measured in near-real time and traced back to its sources, tells a startling amount about life and behavior within the home. While a more traditional meter records monthly energy consumption as a single lump figure, smart meters may collect 750 to 3,000 distinct and time-stamped data points per month. Some smart meters record energy usage every fifteen minutes, and advanced versions may shrink this window to as few as six seconds or permit measurement in real time. This information can be analyzed to reveal medical conditions, criminal activity, and other information about life within the home.

² See, e.g., comments of EFF and the Center for Democracy and Technology (CDT), filed March 9, 2010, available at <http://docs.cpuc.ca.gov/efile/CM/114696.pdf>, and October 15, 2010, available at <http://docs.cpuc.ca.gov/efile/CM/125121.pdf>.

³ “In conclusion, this decision adopts the FIP principles as the framework for developing specific regulations to protect consumer privacy because these principles are consistent with California law, consistent with emerging national privacy and security policies, and supported by the record in this proceeding. A statement of the FIP principles brings clarity to the goals of California privacy and security regulations.” D.11-07-056, at 21, available at http://docs.cpuc.ca.gov/published/FINAL_DECISION/140369.htm.

Individual appliances and other sources of energy use have unique “load signatures,” which are the distinct energy consumption patterns specific to each source. A refrigerator, for example, draws power in a different way than a television, a respirator, or high-wattage indoor marijuana “grow lights.” When aggregated over time, this data can be used to infer the number of people occupying a home, their mundane or illicit habits, and the rhythm of their movements, both in general and on a particular day. Anyone with access to smart meter data can deduce the “avocations, finances, occupation, general reputation, credit, health, or any other personal characteristics of the customer or the customer’s household.”⁴

As Justice Scalia recognized in *Kyllo v. United States*, “at what hour each night the lady of the house takes her daily sauna and bath” is “a detail that many would consider ‘intimate.’”⁵

The state Article I, § 1 constitutional right to privacy, added to the state constitution by ballot amendment in 1972, also applies generally to the collection, use and dissemination of energy usage data whether by government or private entities. “The proliferation of government snooping and data collecting is threatening to destroy our traditional freedoms. Government agencies seem to be competing to compile the most extensive sets of dossiers of American citizens. Computerization of records makes it possible to create ‘cradle-to-grave’ profiles of every American.” *White v. Davis*, 13 Cal.3d 757, 774 (1975); *id.* at 775 (listing “the principal ‘mischiefs’ at which the amendment is directed” as “(1) ‘government snooping’ and the secret gathering of personal information; (2) the overbroad collection and retention of unnecessary personal information by government and business interests; (3) the improper use of information properly obtained for a specific purpose, for example, the use of it for another purpose or the disclosure of it to some third party; and (4) the lack of a reasonable check on the accuracy of existing records.”).

Thus, EFF agrees that “[t]o make an energy data center possible, the Commission would need to decide what constitutes appropriately aggregated and anonymized data,” Ruling, at 3, in order to fully protect customer privacy.

⁴ Sonia R. McNeil, Note, *Privacy and the Modern Grid*, 25 Harv. J.L. & Tech. 199, 204-205 (2011) (footnotes omitted).

⁵ *Kyllo v. United States*, 533 U.S. 27, 38 (2001).

III. SHOULD THE COMMISSION ESTABLISH AN ENERGY DATA CENTER?

Without additional information, EFF is skeptical about the need for an EDC specifically to address researcher access to energy data. The Briefing Paper suggests that researchers have experienced difficulties in obtaining access to energy data from the utilities. As a threshold matter, the proceeding record must establish exactly the problem to be solved, and then review whether it would best be solved by an EDC. We suggest a number of questions to establish a clear record of the problem:

- How often do researchers request energy usage data from the Commission or the investor-owned utilities?
- What entities make these requests— for example, are requesters academic researchers, marketers, local governments, energy start-ups, or others?
- For what purposes do researchers request data, and what kinds of data do they request?
- When the utilities permit access to energy usage data for research purposes, what procedures do they follow and what privacy and security safeguards are implemented?
- What have been the results of such data-access arrangements?

The answers to such questions will help the Commission establish more clearly the scope of the problems.

Moreover, it is presently unclear that any difficulties faced by researchers in gaining access to energy usage data are attributable to the absence of, or would be eliminated by, the proposed EDC. The Commission and the utilities are operating in a world of heightened privacy concerns about energy usage data, and working with a new privacy-protecting regulatory framework. In this transitional period, the rules of the road simply are not yet well settled (including, as discussed below, the definition of crucial terms); it may be that this uncertainty, and not the lack of a central data repository, is the current problem for research access.

In short, EFF believes that the threshold question is not whether to establish an EDC, but to identify the precise problems to be solved and then to decide what rules and practices should govern researchers' access to customer energy usage data regardless of the method used to make it available. As such, the Commission should carefully consider the best approach to safeguarding energy usage data and making it available to researchers in light of its findings.

IV. THRESHOLD QUESTION: THE DEFINITION OF “ANONYMIZED AND AGGREGATED” ENERGY USAGE DATA

The Briefing Paper appropriately asks: “...[H]ow does one determine what is aggregated enough or anonymized enough?” (Briefing Paper, p. 2.) It is necessary for the Commission to consider carefully exactly what data could fall outside the Privacy Rules or the requirement of a non-disclosure agreement (NDA). The Commission began to address this question in D.11-07-056 in at least two places.

Covered Information. “Covered information” is any usage information obtained through the use of the capabilities of Advanced Metering Infrastructure when associated with any information that can reasonably be used to identify an individual, family, household, residence, or non-residential customer, except that covered information does not include usage information from which identifying information has been removed such that an individual, family, household or residence, or non-residential customer cannot reasonably be identified or re-identified.⁶

Availability of Aggregated Usage Data. Covered entities shall permit the use of aggregated usage data that is removed of all personally-identifiable information to be used for analysis, reporting or program management provided that the release of that data does not disclose or reveal specific customer information because of the size of the group, rate classification, or nature of the information.⁷

Unfortunately, while these provisions provide some guidance, they need further elaboration. An initial confusion stems from the word “aggregated.” Often, aggregated or aggregate data refers to collective data that has already been processed, with identifiers removed. The federal Telecommunications Act defines “aggregate information” as “collective data that

⁶ D.11-07-056, at 40.

⁷ D.11-07-056, at 87.

relates to a group or category of services or customers, from which individual customer identities and characteristics have been removed.” 47 U.S.C. § 222(h)(2).

In the financial arena, companies often prepare “aggregate or average data on consumers. Trans Union's SUM-it product, for example, creates aggregate financial information, such as the average mortgage and bank card balance, for consumers who live within a particular zip code, zip code-plus-two digits, or zip code-plus-four digits. The information in the SUM-it database is then used to create models that predict consumers’ financial characteristics or their propensity to purchase certain goods or services. This aggregate information is then made available to marketing firms. Individual information reported by financial institutions about particular customers is not disclosed in this material.” *Individual Reference Services Group, Inc. v. FTC*, 145 F. Supp. 2d 6, 15 (D.D.C. 2001) (internal citations omitted).

On the other hand, “aggregated data” can also simply mean a collection of individual records aggregated together, such as a data set of all customer energy usage records for a given ZIP code. Notably, it is eminently possible to identify individuals from such data, even without standard identifiers. The World War II internment of Japanese-Americans was partly enabled by disclosure of census data. The Census Bureau never released individual names and addresses, only aggregated data for certain localities. “But while the Bureau achieved technical compliance with legal restrictions on releasing information relating to individuals, the practical effect of its actions was tantamount to individual disclosure given that the released population figures were sufficiently detailed to ‘provide[] the parameters for finding and interning the [Japanese-American] population.’” Douglas A. Kysar, Book Review, *Kids & Cul-De-Sacs: Census 2000 and the Reproduction of Consumer Culture*, 87 Cornell L. Rev. 853, 873-874 (2002) (footnotes omitted); *id.* at n. 124 (noting that “the Census Bureau [today] engages in complex data-blurring techniques known as ‘data-swapping,’ ‘random noise,’ and ‘coarsening,’ all designed to protect the integrity of the aggregated data while heightening the security of individual-level information” and that “even with such statistical counter-maneuvers at its disposal, the Bureau remains concerned about re-identification.”) (citations omitted).

EFF believes that the proceeding must address which of these two very different meanings is applicable here. This is especially important because the 15/15 Rule for aggregated data mentioned in the Briefing Paper is clearly inadequate to protect privacy,⁸ and may be unreasonable in light of the definition of “covered information.”

Similarly, it is currently unclear what “anonymized” means. One important technical problem today is re-identification of supposedly de-identified data. (Current practice in the privacy arena tends to prefer the term “de-identified,” because it is so unclear whether data that has had identifiers or other personal details removed can truly be anonymized.) Only in the last few years have researchers come to realize how difficult it is to truly de-identify data.

In particular, researchers Arvind Narayanan and Vitaly Shmatikov have revolutionized the field of re-identification. Based on their statistical research and techniques for re-identifying purportedly anonymous datasets, they conclude that “[t]he emergence of powerful re-identification algorithms demonstrates not just a flaw in a specific anonymization technique, but the fundamental inadequacy of the entire privacy protection paradigm based on ‘de-identifying’ the data.”⁹

An important aspect of this problem is that re-identifiability is a function of both the putatively de-identified data set and other available data. Thus, it may be difficult to know if data is effectively de-identified, as a dataset that is de-identifiable today may become identifiable tomorrow, as more information about persons or households becomes available. Statisticians and computer scientists have been exploring ways to accommodate interests in using large datasets while preserving privacy. As noted earlier, the Census Bureau researches techniques

⁸ The 15/15 Rule states that an aggregation sample must have more than 15 customers and no single customer’s data may comprise more than 15 percent of the total aggregated data. EFF criticized this rule at the October 2012 workshop, based on conversations with a smart grid privacy researcher who gave the following example: a researcher could send two queries to the database—the first for the sum of households 1 through 16, and the second for the sum of households 1 through 17. Simply by subtracting the first from the second, the researcher would be able to derive individual data for household 17. This admittedly naïve technique would work even if the number of customers were increased.

⁹ Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of “Personally Identifiable Information,”* 53 *Comms. of the ACM* 24, 26 (2010); see Arvind Narayanan & Vitaly Shmatikov, *Robust De-Anonymization of Large Sparse Datasets*, 29 *Procs. of the 2008 IEEE Symp. on Security & Privacy* 111 (2008); see also Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 *UCLA L. Rev.* 1701, 1704 (2010).

such as “synthetic data.”¹⁰ Other scientists have done much work in the area of “differential privacy,” which “works by inserting an intermediary piece of software between the analyst and the database,” which then “acts as a privacy-protecting screen or filter, effectively serving as a privacy guard.”¹¹ Other researchers study these privacy-technology issues in the smart grid context.¹²

EFF believes that the Commission should ensure that this state-of-the-art technical research into privacy protection is considered within this proceeding, and that the Commission and the parties obtain expert advice and recommendations from privacy experts and technical experts.

V. UNDER EXISTING LAW AND HUMAN SUBJECTS PROTOCOLS, RESEARCH ACCESS TO AND USE OF SENSITIVE DATA LIKE CUSTOMER DATA MUST BE CAREFULLY CONSIDERED AND CONTROLLED.

Although EFF has not yet deeply researched the issue, we believe that the state Information Practices Act may significantly limit Commission-initiated provision of energy usage data to researchers. First, Civil Code § 1798.24 generally prohibits a state agency from disclosing “any personal information in a manner that would link the information disclosed to the individual to whom it pertains” subject to enumerated exceptions. One such exception permits disclosure “[t]o a person who has provided the agency with advance, adequate written assurance that the information will be used solely for statistical research or reporting purposes, but only if the information to be disclosed is in a form that will not identify any individual.” Civil Code § 1798.24(h). If this section applies here, it is unclear whether the proposed EDC could receive individual-level energy usage data in the first place.

Second, Civil Code § 1798.24(t) specifically addresses state agency disclosure of personal information to “the University of California or a nonprofit entity conducting scientific research” and establishes a significant process for independent review of the need for and details

¹⁰ <http://www.census.gov/icf/docs/synthetic.pdf>.

¹¹ Available at <http://www.microsoft.com/en-us/download/details.aspx?id=35409>.

¹² See, e.g., http://research.microsoft.com/en-us/projects/privacy_in_metering/;
<http://research.microsoft.com/apps/pubs/default.aspx?id=178055>.

of disclosure by “the Committee for the Protection of Human Subjects (CPHS) for the California Health and Human Services Agency (CHHSA) or an institutional review board.”

Importantly, this provision was added by Senate Bill 13 to the Information Practices Act (effective January 1, 2006) in response to a high-profile computer hacking incident in which the state Department of Social Services disclosed the names and Social Security Numbers of In-Home Supportive Services (IHSS) providers and recipients to a researcher at UC Berkeley who was conducting research of IHSS provider wages and benefits into how to better deliver care to people who are homebound. Only a random sample of IHSS data from four counties was needed for the project. The entire IHSS database was downloaded to the researcher in lieu of a partial county sample of the data. A computer hacker took advantage of a known system vulnerability to crack the system that housed the database.¹³ According to a state official, the university had not been in compliance with state security rules for research access to sensitive data.¹⁴

As noted above, EFF does not yet know whether these state-law provisions apply to the disclosure of energy usage data contemplated by the proposed EDC. At a minimum, however, EFF believes the Commission must examine this legal issue. And whether or not these provisions apply, they strongly suggest that existing state policy specifically establishes a need for stringent controls for any disclosure of energy usage data for research purposes.

VI. CONCLUSION

EFF sees at least three broad issues for the Commission going forward. First, because the propriety of an EDC is partly an empirical question of incremental costs and benefits, the Commission should gather facts about the utilities’ current or contemplated practices regarding researchers’ access to energy usage data in order to enable informed comparison to an EDC approach. Second, as a technical matter, the Commission should clarify the meaning of “anonymized and aggregated data” in light of evolving re-identification techniques and the FIPs-

¹³ Available at <http://www.dhcs.ca.gov/formsandpubs/laws/priv/Pages/PrivacyLegislation.aspx> (SB 13 (Bowen; Chapter 241, Statutes of 2005)).

¹⁴ Kevin Poulsen, *California reports massive data breach*, SecurityFocus (2004-10-19), available at <http://www.securityfocus.com/news/9758>.

based privacy and security framework of D.11-07-56. Third, as a legal matter, the Commission should examine the extent to which state law and policy already constrains researchers' access to energy usage data.

EFF therefore suggests that the Commission conduct some form of proceeding to consider: 1) the problem to be solved for researchers; 2) the privacy and security issues to be solved; and 3) the measures that must be put into place to protect privacy and comply with the various legal frameworks identified above, and that the proceeding include both briefings from technical privacy experts and workshops.

EFF appreciates the opportunity to provide these preliminary comments on the energy data center concept discussed in the Ruling and Briefing Paper and looks forward to further discussion next year.

Dated: December 17, 2012

Respectfully submitted,

By: /s/ Lee Tien

Lee Tien

ELECTRONIC FRONTIER FOUNDATION

454 Shotwell Street

San Francisco, CA 94110

Telephone: (415) 436-9333 x 102

Facsimile: (415) 436-9993

E-Mail: tien@eff.org

Counsel for

ELECTRONIC FRONTIER FOUNDATION

Jennifer M. Urban

Assistant Clinical Professor of Law

SAMUELSON LAW, TECHNOLOGY & PUBLIC
POLICY CLINIC

UC-Berkeley School of Law

585 Simon Hall

Berkeley, CA 94720-7200

Telephone (510) 642-7338

E-mail: jurban@law.berkeley.edu

Counsel for

SAMUELSON LAW, TECHNOLOGY & PUBLIC
POLICY CLINIC