



FILED

07-10-13

04:59 PM

**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.

(U39E)

Rulemaking 08-12-009
(Filed December 18, 2008)
Phase III Energy Data Center

**WORKING GROUP REPORT PURSUANT TO
FEBRUARY 27, 2013 ADMINISTRATIVE LAW JUDGE'S
RULING**

CHRISTOPHER J. WARNER

Pacific Gas and Electric Company
77 Beale Street
San Francisco, CA 94105
Telephone: (415) 973-6695
Facsimile: (415) 972-5220
E-Mail: CJW5@pge.com

Dated: July 10, 2013

Attorney for
PACIFIC GAS AND ELECTRIC COMPANY

Pursuant to Ordering Paragraph 4 of the February 27, 2013, “Administrative Law Judge’s Ruling Setting Schedule to Establish ‘Data Use Cases,’ Timelines for Provision of Data, and Model Non-Disclosure Agreements,” as subsequently modified (February 27 ALJ Ruling), Pacific Gas and Electric Company (PG&E), on behalf of itself, Southern California Edison Company (SCE), San Diego Gas & Electric Company (SDG&E), and Southern California Gas Company (SoCalGas), provides the Working Group Report that summarizes the results of the collaborative working group in the areas of use cases, definitions, and non-disclosure agreements in this phase of the proceeding.^{1/} The Working Group Report is attached as Appendix A to this pleading.

A draft of the Working Group Report was provided for comment to the working group participants and other parties on July 3, 2013, and the Working Group Report reflects comments received by the close of business July 9, 2013, including alternate views incorporated directly into the Report. The Working Group Report does not and should not be interpreted as reflecting a consensus of the working group participants or formal concurrence by the working group participants in the Report. The Working Group Report is a factual summary of the working group discussions and views of working group participants who contributed to the Report or provided comments on the Report. An opportunity for all working group participants and interested parties including the utilities, to provide views and comments is provided pursuant to the formal comment period subsequent to the filing of the Report.

^{1/} Counsel for SCE, SDG&E and SoCalGas have authorized PG&E to file and serve this Working Group Report on their behalf.

The working group participants express appreciation for the facilitation, advice and assistance provided by Commission staff during the working group sessions, including ALJ Jessica Hecht who served as informal facilitator for the working group sessions.

Respectfully Submitted,
CHRISTOPHER J. WARNER

By: /s/ Christopher J. Warner
CHRISTOPHER J. WARNER

Pacific Gas and Electric Company
77 Beale Street
San Francisco, CA 94105
Telephone: (415) 973-6695
Facsimile: (415) 972-5220
E-Mail: CJW5@pge.com
Attorney for
PACIFIC GAS AND ELECTRIC COMPANY

Dated: July 10, 2013

**Working Group Report
R.08-12-009
Phase III Energy Data Center
February 27, 2013 ALJ Ruling
July 10, 2013**

I. INTRODUCTION AND EXECUTIVE SUMMARY

This Working Group Report for the Energy Data Center phase of CPUC Rulemaking 08-12-009 is organized as follows:^{1/}

Section I includes an executive summary of the Working Group Report, including a summary of the scope of the Report and potential consensus and alternative proposals and recommendations in the Report.

Section II includes a summary of the Working Group meetings and sessions conducted by the investor-owned utilities (utilities or IOUs)^{2/} and interested parties and facilitated by ALJ Jessica Hecht and Commission staff. The IOUs and interested parties express their appreciation for the support and facilitation provided by Judge Hecht; Judge Sullivan; Audrey Lee, advisor to President Peevey; Alope Gupta of the Commission's Energy Division; and Chris Villareal of the Commission's Policy & Planning Division.

1/ A draft of the Working Group Report was provided for comment to the working group participants and other parties on July 3, 2013, and the Working Group Report reflects comments received by the close of business July 9, 2013, including alternate views incorporated directly into the Report. The Working Group Report does not and should not be interpreted as reflecting a consensus of the working group participants or formal concurrence by the working group participants in the Report. The Working Group Report is a factual summary of the working group discussions and views of working group participants who contributed to the Report or provided comments on the Report. An opportunity for all working group participants and interested parties, including the investor-owned utilities, to provide views and comments is provided pursuant to the formal comment period subsequent to the filing of the Report.

2/ Pacific Gas and Electric Company (PG&E), San Diego Gas & Electric Company (SDG&E), Southern California Edison Company (SCE), and Southern California Gas Company (SoCalGas).

Section III provides a summary of the legal and public policy framework on customer privacy and energy usage data that informed the Working Group discussions. This summary is supported largely by a legal memorandum provided by the Samuelson Law, Technology & Public Policy Clinic at the University of California, Berkeley, School of Law, representing the Electronic Frontier Foundation (EFF), an interested party in the proceeding.

Section IV includes a summary of technical advice and recommendations on privacy protections provided by data scientists assisting EFF in the proceeding. The IOUs and interested parties appreciate this extensive and helpful technical advice and “tutorials” on technical privacy issues provided on behalf of EFF by Moritz Hardt from IBM Research and Cynthia Dwork from Microsoft Research.

Section V includes a discussion of the definitions and classification of energy usage data used by the interested parties to evaluate the “use cases” and proposed protocols for energy usage data access in the proceeding.

Section VI discusses and provides potential recommendations regarding each of the 8 energy usage data “use cases” in the February 27 ALJ Ruling and 4 additional use cases suggested by interested parties.

Section VII evaluates and provides potential recommendations regarding energy usage data access protocols for third parties to access energy usage data, including comments on a “strawperson” streamlined data access process proposed by the IOUs and a draft standardized non-disclosure master data access agreement attached to the February 27 ALJ Ruling.

Section VIII provides a general conclusion to the Report and potential recommendations on “next steps” in the proceeding.

A. Scope of Working Group Report

On February 27, 2013, the Administrative Law Judge issued a ruling (ALJ Ruling) in the Energy Data Center phase of Rulemaking 08-12-009, establishing “the next steps for receiving proposals to ensure the timely provision of energy usage data, particularly when personally identifiable information (PII) has been removed, to requestors of data interested in topics of policy interest to California ratepayers, utilities, and policy makers.”^{3/} The Ruling required Pacific Gas and Electric Company (PG&E), Southern California Edison Company (SCE), Southern California Gas Company (SoCal Gas) and San Diego Gas & Electric Company (SDG&E) to form a working group, including representatives of interested parties, to make proposals on certain energy usage data access issues for third parties and tasks identified in the ALJ Ruling.^{4/} The ALJ Ruling as subsequently revised also required the utilities to file a working group report on the issues and proposals by July 10, 2013.^{5/}

The energy usage data access issues and tasks identified by the ALJ Ruling include:

1. Propose refinements to eight use cases listed in the ALJ Ruling, and develop additional use cases as needed, using the template included in Attachment B to the ALJ Ruling. According to the ALJ Ruling, the purpose of the use cases is to assess whether a particular use case raises issues that require resolution, including assessing

3/ ALJ Ruling, p. 1.

4/ ALJ Ruling, pp. 17- 19. CPUC D.11-07-056, D.12-08-045 and other statutes and CPUC decisions and orders establish the privacy rules applicable to third-party access to customer energy usage data and other customer-specific information. See, also, Attachment B to D.11-07-056.

5/ ALJ E-mail, June 20, 2013.

the risk to privacy that providing access to data under the use case entails, and to forecast the value to ratepayers that access to the data can produce. In addition, the use cases should include estimates of the costs of preparing or maintaining the data and the access to the data described in the use case.^{6/}

2. Propose definitions for eight energy data access terms listed in the ALJ Ruling. According to the ALJ Ruling, the purpose of the definitions is to ensure that there is common understanding of the key terms used to describe data, and thus facilitate the development of policies that provide easier access to data while protecting privacy.^{7/}

3. Propose refinements to a draft energy usage data access non-disclosure agreement, including data security protocols, included as Attachment A to the ALJ Ruling. According to the ALJ Ruling, the purpose of this task is to provide a starting point for discussion on the elements of a standard non-disclosure agreement that could be used by all California energy utilities or other agencies that provide data to eligible recipients.^{8/}

On May 8, 2013, the ALJ issued a subsequent ruling, adding two memoranda by the Electronic Frontier Foundation to the record, and asking for comments by the interested parties on two memoranda provided for the record by EFF.^{9/} The first memorandum is titled “Legal Considerations for Smart Grid Energy Data Sharing” and is attached as Appendix C to the Working Group Report. This memorandum “covers legal background relevant to this proceeding, providing a brief explanation of important laws

6/ ALJ Ruling, pp. 16- 17.

7/ *Id.*, p. 11.

8/ *Id.*, p. 18.

9/ ALJ email, May 8, 2013.

that apply to energy usage data sharing, as well as a brief background of the legal landscape covered in the proceeding to date.”^{10/} The second memorandum is titled “Technical Issues with Anonymization & Aggregation of Detailed Energy Usage Data as Methods for Protecting Customer Privacy.” The memorandum states that it “addresses the technical issues surrounding aggregation and anonymization of customer data.”^{11/} The memorandum appends a paper titled *Privacy Technology Options for Protecting and Processing Utility Readings*. A copy of the second memorandum is included as Appendix D to the Working Group Report.

In particular regarding the second memorandum, the ALJ Ruling indicated that there appears to be an inherent tension between the memo, which focuses on the failings of techniques for protecting the privacy of data, and Appendix A of the memo, which proposes “Robust Privacy Technology Options.” Accordingly, the ALJ Ruling invited comments on the “*Laplacian* mechanism” and “the *Subsample and Aggregate* mechanism” for incorporating “noise” from a specific noise distribution. How robust are these techniques for protecting the privacy of a particular statistic? Does the addition of “noise” dilute the power of subsequent statistical analyses of the data, or does the fact that the noise is generated by a known distribution enable adjustments that eliminate bias? What effect, if any, does the addition of noise have on the variance of statistical estimators? In what settings should these mechanisms be used and where are they not needed? In addition, the ALJ Ruling invited comments on other techniques for protecting the privacy of data.^{12/}

10/ *Id.*, ALJ email, draft ruling, p. 2.

11/ *Id.*

12/ *Id.*, pp. 2- 3.

Sections II – IV of the Working Group Report provide a summary of the working group meetings and the legal, policy and technical framework for the Working Group report. In more detail in Sections V- VII, below, the Working Group Report discusses the proposals, recommendations and comments of the interested parties on the issues and tasks identified in the above-referenced ALJ Rulings and within the scope of this phase of R.08-12-009.

B. Summary of Potential Recommendations and Comments of Parties

The interested parties participating in the Working Group generally agreed that energy usage data is useful for various energy and environmental policy analyses and research, as well as for evaluating the cost effectiveness of various energy efficiency, renewable energy and climate planning programs. The interested parties discussed but did not agree formally on the definitions and legal and policy criteria that should be applied to energy usage data access. The parties also discussed a broad outline of a streamlined process by which the utilities should expedite the granting of third-party access to certain, pre-approved categories of energy usage data that do not include customer-specific or personally identifiable information and that do not present significant privacy, commercial, grid security or competitive concerns.

Additionally, the interested parties have been unable to agree on the overall legal and policy criteria that should apply to energy usage data access, including how certain technical, practical, and legal constraints should be addressed for certain categories of energy usage data access. These issues include whether and how to distinguish between research proposals that provide concrete benefits to utility customers on energy and environmental programs and policies, and those that may be worthy of research but do not provide potential benefits to customers. These issues also include

how the Commission, the utilities and interested parties should evaluate the trade-offs between the risk of violation of customers' privacy and the benefits of making available customer-specific information for public interest research and program planning. Because of these disagreements, the unresolved technical and implementation issues regarding how to screen energy usage data access research proposals and how to protect and "anonymize" customer-specific information from "re-identification" need to be the subject of further discussion and informal resolution within the broad policy parameters that the Commission may adopt on energy usage data access in this proceeding.

1. Potential Recommendations and Proposals

The following is a summary of the proposals which the interested parties discussed and on which agreement may be reached:

1. The utilities should adopt streamlined processes for granting third-party researchers, local governments and accredited public institutions with access to monthly electric and natural gas consumption data that does not contain personally identifiable information or customer-specific data and that has been adequately "anonymized" and aggregated such that the personally identifiable information of utility customers cannot reasonably be identified or "re-identified." Customer-specific energy usage data that has not been "anonymized" and aggregated would only be disclosed to third parties if authorized by the customer or if used for a specific utility operational purpose as provided by the Public Utilities Code or ordered by the Commission. The streamlined process should be consistent across the utilities and should provide for access to pre-approved energy usage data "templates" for pre-approved "use cases"

where the requests can be processed in less than 30 days from the time of receipt of a complete data access request from a qualified researcher, local government or other accredited public institution. The process should also incorporate common information security controls agreed to by the utilities, such as secure access and data transmission methods, where required by the utility, as well as a standard non-disclosure agreement where appropriate.

2. Examples of energy usage data access cases that should be subject to pre-approved, streamlined access include aggregated data (data from which personally identifiable information cannot reasonably be identified or re-identified) such as:

(a) aggregated monthly energy consumption data made available to local governments for climate planning and for local government energy efficiency programs;

(b) aggregated monthly energy consumption data made available to accredited academic and public research institutions in California for California energy or environmental policy research purposes that clearly demonstrate the potential value of their research to ratepayers; and

(c) aggregated monthly energy consumption data made available or reported to California state government energy and environmental agencies for energy or environmental policy analysis or planning, including the CPUC, Energy Commission and California Air Resources Board.

3. Customer-specific energy usage data, particularly interval AMI-generated customer specific energy usage data, should not be provided to third-parties without customer authorization or consent, except for utility operational or program purposes consistent with Public Utilities Code Section 8380(e)(2) or pursuant to a specific state

law or regulation authorizing disclosure for specific purposes under specific privacy protections enforced directly against the third-parties, such as Community Choice Aggregation programs under Public Utilities Code Section 366.2(c)(9) and CPUC Decision 12-08-045.

4. For building benchmarking programs and regulations, such as the City and County of San Francisco building benchmarking ordinance and the AB 1103 Energy Commission building benchmarking program, building tenant energy usage that is sufficiently “anonymized” to avoid “re-identification” can be provided to building owners or landlords without customer consent. The “building benchmark” data should be sufficiently “blurred” through anonymization techniques that minimize the risk of “re-identification” of customers’ identities and personal information.

5. In compliance with the prohibition in Public Utilities Code Section 8380(b)(2), the utilities should not sell, license or transfer customer-specific energy usage data to third parties for commercial gain or profit-making purposes. This prohibition applies to “use cases” involving disclosure of customer-specific energy usage data for purposes of promotion and marketing of rooftop solar systems by retail solar vendors and installers (Use Case 6 in the February 27, 2013, ALJ Ruling); for purposes of marketing and promotion of energy efficiency retrofits by energy efficiency contractors (Use Cases 8 and 11 in the ALJ Ruling); and for purposes of promotion and marketing of loans by financial institutions for energy retrofits under utility “on-bill” financing programs (Use Case 5 in the ALJ Ruling.) In addition, the utilities, as well as third-parties that receive the data, may not sell, license or transfer aggregated or “anonymized” energy usage data to third parties for such commercial gain or profit-

making purposes, unless the Commission has approved the sale, licensing or transfer as providing adequate benefits to utility ratepayers in accordance with the “gain on sale” rules applied by the Commission under Public Utilities Code Section 851, or unless the energy usage data is already in the public domain for other purposes.

6. As a condition precedent to the utilities implementing a streamlined energy usage data access program as discussed above, the Commission should authorize the utilities to recover, either in their revenue requirements or through user fees, the full reasonable incremental costs the utilities incur to implement the data access program, including start-up and ongoing costs as well as costs associated with any special requests for information or analyses not addressed by the energy usage data access program.

7. The utilities should use consistent and standardized non-disclosure agreements and information security review protocols to ensure the protection of customer-specific information, intellectual property and competitive-sensitive information provided for the research or government planning purposes identified in the use cases in this phase of the proceeding. Information security reviews, to the extent necessary, should be streamlined and simplified so that third-party users of energy usage data can “pre-qualify” and be certified for receipt of data to relevant national standards in advance to the extent practicable. The parties recommend a separate working group process to finalize the standardized NDA forms and information security protocols applicable to energy usage data access.

Further discussion of these potential recommendations and proposals is provided in Sections V – VII of the Working Group Report, below, in connection with discussion of the use cases, definitions, and data access protocols.

2. Issues Not Addressed in Scope of Working Group Discussions

It is important to note what issues have *not* been extensively addressed by the Working Group participants within the scope of this phase of the proceeding. This phase of the proceeding has (properly) *not* included extensive discussion of access to customer-specific information and data that is not energy usage data generated by the utilities' advanced metering infrastructure (AMI). Privacy and access rules for customer-specific data outside the scope of this proceeding are established by previous Commission decisions and utility tariffs. See, e.g., Attachment B, D.11-07-056 and D.12-08-045.

The Working Group was not tasked with addressing AMI-generated energy usage data that is used or disclosed by a utility to its vendors, contractors or agents for utility operational purposes under Public Utilities Code Section 8380(e)(2), e.g. “for system, grid, or operational needs, or the implementation of demand response, energy management, or energy efficiency programs....” (Public Utilities Code Section 8380(e)(2); D.11-07-056 and D.12-08-045, Attachment D, Rules 1(c) and (e), distinguishing between “primary purposes” related to utility operations, for which customer consent to disclose energy usage data is unnecessary, and all other purposes defined as “secondary purposes” for which customer consent and authorization is required.) Utility use of customer energy usage data in order to support utility operations and programs, e.g. billing, metering, and energy efficiency and demand

response programs, is already governed by the detailed tariffs filed by the utilities to implement the privacy rules adopted for such purposes in Attachment D of D.11-07-056 and D.12-08-045. Likewise, this phase of the proceeding does *not* include a utility's use or disclosure of aggregated energy usage data for purposes of "analysis, reporting or program management" pursuant to Public Utilities Code Section 8380(e)(1), as long as "all information has been removed regarding the individual identity of a customer." (Public Utilities Code Section 8380(e)(1).)

Nor was the Working Group tasked with addressing the implementation of customer-authorized access to bulk energy usage data. The electric utilities' respective Customer Data Access applications, currently pending for decision before the CPUC, fully address the requirement of Ordering Paragraph 8 of D.11-07-056 that they provide third parties with access to a customer's usage data via the utility's AMI backhaul when authorized by the customer, including using a common data format among the utilities and conforming to ongoing national standards for such data access. See Applications (A.) 12-03-002 (PG&E), A.12-03-003 (SDG&E), and A.12-03-004 (SCE).

Finally, per the direction of the ALJ Ruling, the Working Group has focused directly on access to energy usage data, and not on customer-specific information that is not energy usage data, such as (a) whether a customer has participated in a particular energy efficiency or weatherization program; (b) the income level of a utility customer for purposes of determining eligibility for governmental assistance programs offered by non-utility public agencies; (c) information regarding the type and specifications of energy efficiency or energy management equipment or devices installed under a utility program but subject to separate regulation by other

governmental agencies, such as Title 24 contractor compliance administered by the Energy Commission; or (d) customer-specific financial or billing information, such as credit and collection history, for purposes of promoting energy efficiency lending to customers, such as under “on-bill financing” programs. Issues and policies regarding access to these types of customer-specific or aggregated data are addressed in other Commission proceedings.

3. Alternative Proposals and Recommendations

Solar City

Re-identification of Customer Level Usage Data and Technical Feasibility

In general, we are concerned that the staff report goes too far in accepting and, in effect, endorsing the arguments presented by the Electronic Frontier Foundation (EFF) that because it may be technically feasible to re-identify customer level energy usage data, even when stripped of personally identifying information (PII), this data should be considered covered information under the Commission’s privacy rules and therefore should be subject to prior customer consent before it can be shared with third parties.

This appears to effectively eliminate SolarCity’s use case, which we believe strikes a reasonable and appropriate balance between the use of AMI data to advance state policy objectives and to empower customers with the knowledge required to make rational energy management decisions and ensuring customers are not unwittingly compromising their privacy.

We believe it is premature to draw the conclusion that customer-level usage data with or without PII is covered information and, in so doing, treating this as the starting point for establishing protocols around data access. In particular, we are concerned that in endorsing EFF's suggested solution, namely by requiring data to be "blurred" or otherwise anonymized to eliminate any possibility of re-identification, the usefulness of the data to facilitate certain use cases will be unnecessarily compromised.

Even assuming that customer-level usage data, stripped of PII, could be used to re-identify a customer, the question of whether this data should be considered covered information under the Commission's privacy rules is highly contestable.

The language adopted by the Commission regarding covered information specifically exempts *"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"*. In adopting this language, we believe the Commission sought to balance effective use of energy data with the privacy concerns associated with the potential threat of re-identification. Whether the sharing of energy usage data, stripped of PII, should be considered covered information hinges on a finding that the energy usage data can be reasonably used to re-identify a household.

EFF argues, and the Draft Report appears to accept, that the technical possibility of re-identification meets this reasonableness standard. If third-parties were given unfettered access to customer-level usage data, stripped of PII, this might technically be true.

However, the concern about re-identification can be addressed in a number of ways.

EFF presents one set of solutions. Unfortunately, the Draft Report appears to take for granted that data blurring or other technical approaches to confound re-identification are the only set of solutions. Another solution would be to simply require third-parties seeking customer level energy usage data to execute a contract, under penalty or appropriate recourse, that forecloses any efforts to re-identify customers. By making re-identification a violation of the contractual terms, such a requirement could mean that usage data cannot be reasonably used to identify or re-identify a customer, given the consequences violators would face.

Additionally, we believe a useful distinction should be made between "*customer-level* usage data," access to which SolarCity believes should be given to third-parties through its use case, and "*customer specific* energy usage data". Customer-specific usage information would include AMI-generated customer energy usage data that could be reasonably used to identify or re-identify a given customer. For example, AMI data paired with a name, address, and/or zip code would clearly be considered customer-specific. However, AMI-generated usage information, stripped of PII and conveyed under contractual terms discussed above, should not be considered customer-specific information. Rather, it should be considered customer-level usage data. With this clarification, we believe that the proposed Consensus Recommendations and Proposals" are reasonable. Absent such a clarification, SolarCity does not believe it is accurate to characterize all of the proposed items listed in this section as reflecting a consensus view, in particular items 3 and 5.

II. SUMMARY OF WORKING GROUP MEETINGS

On January 15 and 16, 2013, the Commission held two days of workshops on Energy Data Access. The workshops sought to work towards “a consistent, uniform, transparent process for access to energy data from the investor-owned utilities and to explore security, legal, economic, and policy issues associated with an energy data center.”^{13/} The topics covered during the workshops are described in detail in the February 27, 2013 Administrative Law Judge (ALJ) ruling (ALJ Ruling) setting the schedule. At the end of the workshops, it was decided to utilize a collaborative workshop process as the next step in this proceeding. The collaborative workshop process would be led by the utilities and would include other interested parties, as well as, subject matter experts in the fields of data privacy and data collection. The utilities were encouraged to work with a Commission-trained professional facilitator to help run the collaborative workshops. The work of the collaborative workshops is summarized in this report.

The following interested parties participated directly and provided their views in the various working group sessions held during April and May, 2013: PG&E, SCE, SDG&E, SoCal Gas, TURN, Division of Ratepayer Advocates, UCLA California Center for Sustainable Communities, Energy Institute at Haas, Local Government Sustainable Energy Coalition, City and County of San Francisco, Electronic Frontier Foundation; Distributed Energy Consumers Advocates, California Energy Commission, County of Los Angeles, Lawrence Livermore National Laboratory, and Brighter Planet Technology Services, (aka Faraday Company). Other parties who submitted information or views to

13/ *Administrative Law Judge’s Ruling Setting Schedule to Establish “Data Use Cases”, Timelines for Provision of Data, and Model Non-Disclosure Agreements*, Phase 3 R.08-12-009, p.3, February 27, 2013.

the Working Group included the California Department of Community Services, Natural Resources Defense Council, and Solar City. The Working Group Sessions were also facilitated by ALJ Jessica Hecht, and informal views and information were provided to the participants by Commission staff.

The first Working Group Session was held on April 3, 2013. ALJ Jessica Hecht was the facilitator and Chris Vera from SDG&E and Jennifer Urban from the University of California, Berkeley, School of Law lead the discussions. The group was unsuccessful in coming to a consensus on definitions of aggregated data vs. anonymous data, personal information vs. personally identifiable information (PII), and validity vs. usefulness of data access. Legal references and legal framework for these definitions were also a topic of discussion. Jennifer Urban, Lee Tien and University of California-Berkeley School of Law students Brady Blasco and Julie Byren, representing the Electronic Frontier Foundation (EFF), presented two memoranda they wrote in preparation for the working group session. One memo addressed *Legal Considerations for Smart Grid Energy Data Sharing* and the other addressed *Technical Issues with Anonymization & Aggregation of Detailed Energy Usage Data as Methods for Protecting Customer Privacy*. Both memos have been submitted into the record for this proceeding as part of ALJ Ruling^{14/} dated May 13, 2013. The Appendix to the second memo on technical issues with anonymization and aggregation is a research paper by George Danezis of Microsoft Research.^{15/} Mr. Danezis was unable to attend the

14/ *Administrative Law Judge's Ruling Adding Technical Memos to the Record, and Inviting Comments and Replies; Revising Schedule for Filing Use Cases, Comments and Replies*, Phase 3 R.08-12-009, May 13, 2013.

15/ *Administrative Law Judge's Ruling Adding Technical memos to the Record, and Inviting Comments and Replies; Revising Schedule for Filing Use Cases, Comments and Replies*, Phase

workshop, however, Moritz Hardt from IBM Research was able to discuss the topics in the research paper and answer questions.

The second Working Group Session was held on April 15, 2013. Once again ALJ Jessica Hecht was the facilitator. This session focused specifically on use cases 5, 6 and 8, as identified in the February 27, 2013 ALJ Ruling, a new use case submitted by Brighter Planet Technology Services (aka Faraday), and a new use case submitted by Distributed Energy Consumer Advocates (DECA). Topics of discussion for each use case included specific type of data needed, granularity of the data (e.g. customer specific PII, non-PII anonymized, non-PII aggregated) or level of aggregation, purpose for which the data is requested, how the data will be used and distributed, and format of the data requested. It was agreed that each party which finds a particular use case relevant, will fill out Attachment B of the February 27, 2013 ALJ Ruling.

The third Working Group Session was held on April 17, 2013 and followed the same format as Working Group Session 2. The use cases discussed at the session were 1, 2, 3, 4, and 7 from the February 27, 2013 ALJ Ruling. It was determined that smaller, more focused discussions need to occur between the privacy experts (EFF and UC Berkeley, School of Law), the data scientists (Moritz Hardt from IBM Research and Cynthia Dwork from Microsoft Research) and the parties that would be requesting and using the data in the use cases.

The smaller working group sessions were held on May 13 and May 15, 2013. The two sessions covered use cases including 1) distributed generation grid use and planning data; 2) retail commercial solar PV and energy efficiency contractors; 3) local

3 R.08-12-009, Appendix A – “Privacy Technology Options for Protecting and Processing Utility Readings”, May 13, 2013.

government access to data for climate planning and energy efficiency programs; 4) local and state government access to data for building benchmarking, such as under AB 1103 and the County and City of San Francisco (CCSF) building benchmarking ordinance; and 5) general “public interest” research for energy policy and program purposes. Key attendees included the CPUC staff (ALJ Hecht, Audrey Lee, and Alope Gupta); representatives of EFF; Solar City; UCLA; Haas Energy Center; the Local Government Sustainability Coalition; DECA; CCSF; County of LA; Energy Commission; Lawrence Livermore National Lab; and Brighter Planet Technology Services (aka Faraday). PG&E and the other utilities monitored the sessions, but did not actively participate, because the focus of the discussions was the technical issue relating to how to sufficiently anonymize customer-specific data in order to protect the customer specific information from being “re-identified.”

The last all-party Working Group Session (#4) was held on May 22, 2013. Alope Gupta, Audrey Lee and ALJ Jessica Hecht of the CPUC facilitated and organized the meeting. Topics discussed included overview of and feedback on the utilities “strawperson” for the “streamlined data access” proposal, feedback on the utilities proposed Non-Disclosure Agreement, data privacy framework as seen by Alope Gupta of the CPUC, summary of the use case proposals, and two technical solutions (data cubing and interactive query system) for public access data presented by Moritz Hardt of IBM Research and Cynthia Dwork of Microsoft Research. Not all questions were covered during this meeting and it was decided to have additional meetings/calls scheduled to discuss costs (development and ongoing), list of data attributes currently

available at the utilities, and possibility and cost of implementing a data cube or interactive query system solution.

A follow up call was held between the utilities and Moritz Hardt of IBM Research on June 5, 2013. The utilities provided ahead of time a list of specific questions related to data cubing and interactive query system solutions. Moritz was unable to provide specific responses without having more information from the utilities about the data attributes currently available, number of years of data available, structure and size of the databases, etc. It was agreed that a representative from SDG&E and SoCalGas would work directly with Moritz to come up with a sample database that Moritz could use to run a data cubing algorithm on to demonstrate how the solution would work and what the output would look like. During this June 5, 2013 call, the technical experts confirmed that no “off-the-shelf” data cube software was currently available, and that utilities wishing to use a data cube would need to retain software development engineers to create the appropriate user interface and to customize the data cube to the databases desired to be used. The technical experts also stated that California utilities would be among the vanguard nationally to use a data cube should the solution be explored.

The utilities then had a follow up call with the CPUC staff on June 10, 2013. The discussion focused on consistent terminology for energy data requests, energy data request template, streamlined process across IOUs to handle energy data requests on a timely basis, guidelines for “scrubbing” energy data to reduce risk of customer re-identification, and disposition of specific energy data request use cases.

III. LEGAL AND POLICY FRAMEWORK

As a threshold matter, the parties agree that energy usage data access and privacy are not a “zero sum,” i.e. personally-identifiable energy usage information, particularly granular, customer-specific interval energy usage data, can be protected from use and disclosure to third parties unless the customer consents and authorizes third-party access to the data. On the other hand, energy usage data that excludes personally identifiable information from direct or indirect disclosure through aggregation or “anonymization” of the data, can be made available, provided that the aggregation or anonymization technique reasonably prevents re-identification.

However, how to strike this balance between the utility of energy usage data access and the protection of customer privacy requires a detailed understanding of the legal and technical aspects of California’s privacy laws and the rapidly changing technologies which enable analysis of personal data.

This section and the following section of the Working Group report summarize the legal, policy and technical issues that have framed the working group discussions on these topics. The summary of legal and policy issues is based on the memorandum dated April 1, 2013 submitted for the record by EFF and the Samuelson Law, Technology & Public Policy Clinic at the University of California, Berkeley, School of Law.

A. The California Constitution

Article I, Section 1 of the California Constitution recognizes each individual’s right to privacy. There is general agreement among the judicial, scholarly, legislative, and regulatory communities that the data collected by smart meters reveals intimate details

about the lives of California citizens. As such, the California Constitution establishes a baseline obligation to protect energy usage data from harmful disclosure or use.

The same interests that motivated California citizens to enact Section 1 by ballot amendment in 1972 still apply today: (1) the overbroad collection and retention of unnecessary personal information by government and business interests; and (2) 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.

Representative of the high value the California public places on privacy, the California Constitution imposes an obligation to protect consumer privacy on all parties—including private parties—engaging in energy usage data sharing. As such, addressing privacy issues are necessarily central to this proceeding.

B. California Information Practices Act

The California Information Practices Act (IPA; California Civil Code section 1798 *et seq.*) governs the manner in which state agencies, as defined in the IPA, collect and disclose personally identifiable information and data (hereafter, “PII”). The statute applies to state-wide agencies, including the CPUC and the California Energy Commission. Should the CPUC designate a state agency as a custodian of customer energy usage data, the IPA will apply to that agency’s collection and disclosure of the data. In addition, the IPA applies to collection and disclosure of personally identifiable information by a state agency such as the Commission, even where the data is indirectly collected from the utilities.

The IPA protects energy usage data that “identifies or describes an individual”—in this context, an individual utility customer.^{16/} The IPA offers a non-exhaustive list of example types of “personal information” that might be used to identify or describe an individual, including an individual’s “name, social security number, physical description, home address, home telephone number, education, financial matters, and medical or employment history.”^{17/} Additional types of information, such as sex, birthdate, and zip code, may operate as “quasi-identifiers,” capable of re-identifying an individual when linked to other available data. The IPA’s list of identifiers would include that information as well.

As a general rule, state agencies are not permitted to disclose any personal information “in a manner that would link the information disclosed to the individual to whom it pertains.”^{18/} However, a number of exceptions apply, subject to varying protocols and approval procedures depending on the data recipient. For example, Section 1798.24 authorizes disclosure of an individual’s personal data in the following pertinent scenarios, among others:

- With the prior written voluntary consent of the individual, Cal. Civ. Code § 1798.24(b);
- To persons, or another state agency, such as the CEC, for whom the information is necessary to fulfill statutory duties, Cal. Civ. Code § 1798.24(e);
- Where the CPUC is required by law to disclose the information to a local government (or federal government) entity,¹³ Cal. Civ. Code § 1798.24(f);

16/ Cal. Civ. Code Section 1798.3(a).

17/ The IPA also includes “statements made by, or attributed to, the individual” within its list of identifiers. Cal. Civ. Code Section 1798.3(a).

18/ Cal. Civ. Code Section 1798.24.

- Disclosure to a researcher, if (1) he or she provides assurance that the information will be used solely for statistical research or reporting purposes, and (2) he or she does not receive the information in a form that will identify the individual, Cal. Civ. Code § 1798.24(h); and
- Disclosure to a researcher within the University of California system, provided that the request is approved by the Committee for the Protection of Human Subjects, Cal. Civ. Code § 1798.24(t).

Of particular relevance to Working Group discussion is Section 1798.24(h), which specifically addresses disclosure for research purposes. This provision underscores the California legislature's commitment to protecting the privacy of the individual(s) to whom the data pertains by explicitly limiting disclosure of personally identifiable information to researchers, while allowing research. Note that Section 1798.24(e) also practically limits the scope of agency disclosures to only those specifically and directly authorized by statute, lest the exception swallow the rule.

One of the fundamental privacy concerns motivating the enactment of the IPA was the risk of data breach, a problem that is prevalent and well-documented among all institutions, including California institutions. An important obligation the IPA imposes on third party data recipients working within the University of California system is that requests for disclosure of personal information must first be approved by the Committee for the Protection of Human Subjects (CPHS), or another institutional review board that has written authorization from the CPHS.

As such, the IPA provides both legal requirements binding on relevant agencies and overall guidance as to how California has thus far approached data risks for

California citizens. Accordingly, although the IPA is not binding directly on utilities or academic or local government researchers, or other parties who cannot be characterized as state agencies, it nevertheless provides useful guidance in this proceeding because it guides how California law might treat the disclosure of energy usage data more generally.

C. SB 1476 and the CPUC Privacy Rules

In the smart grid context, statewide concern in California with consumer privacy has culminated in the Legislature's 2010 adoption of SB 1476(Padilla) and the CPUC's implementation of SB 1476 through Privacy Rules which specifically address the sharing of energy usage data by IOUs. The Privacy Rules most directly address the type of data sharing at issue in this phase of the proceeding: (1) they specifically regulate energy usage data collected by smart meters, and (2) they govern disclosure by the IOUs to third party data requesters.

In addition to implementing the requirements of SB 1476, the CPUC in D.11-07-056 established that the sharing of energy usage data should follow **Fair Information Practice Principles** (FIPPs), a widely accepted international framework for handling electronic information in a privacy-protective manner.^{19/} In the 2011 Decision, the Commission explicitly adopted the FIPPs as California's policy for smart grid privacy. Thus, the foundational principles set forth in the FIPPs provide guidance to the Working Group for determining how to most effectively apply the Privacy Rules to the use cases in this proceeding.

The eight principles embodied in the FIPPs include:

19/ D.11-07-056, pp. 19- 21, July 28, 2011.

1. *Transparency*: Any new collection or disclosure of customer-specific energy usage data that is separate from the IOUs would make it more difficult to provide notice to individual utility customers about the use or dissemination of their PII.

2. *Individual Participation*: The Working Group should continue to consider informed customer consent as the “default” preferred process for data collection, use, dissemination and maintenance. Unlike typical consumers, many utility customers have no choice when buying energy. As a result, foregoing consent for disclosure of their private data is not a matter that they have agreed to as a condition of receiving utility service.

3. *Purpose Specification*: Requesting parties must be required to specify the purpose underlying a request for energy usage data access prior to authorization for disclosure.

4. *Data Minimization*: Only the data actually necessary for the particular purpose identified should be disclosed. The FIPPs’ minimization principle helps in developing data handling practices that limit data breach and other risks before they happen, and helps data handlers decide on data needs in an efficient manner.

5. *Use Limitation*: There must be mechanisms to ensure that the disclosure of information is used solely for the specified purpose(s).

6. *Data Quality and Integrity*: When multiple parties are permitted to collect and store energy usage data, it would be harder to ensure that the data is accurate, relevant, timely, and complete. The problems associated with one data set may be multiplied across parallel data sets.

7. *Security*: Any data collected from the IOUs and stored pursuant to security protocols that are less rigorous than those utilized by the IOUs under the CPUC privacy rules and “best industry practices” may be susceptible to loss, unauthorized access, destruction, modification, or unintended disclosure.

8. *Accountability and Auditing*: Mechanisms are already in place to enforce IOUs’ compliance with the FIPPs directly. However, it is important to ensure equivalent enforceability and accountability against any non-utility, third-party entity collecting and disclosing customer-specific energy usage data.

D. Privacy Rules, adopted in D. 11-07-056 (Attachment D)

Under the CPUC Privacy Rules, “Covered information” is defined similarly to the definition of “personal information” in the IPA. “Covered information” is information that does not include usage information from which identifying information has been removed such that an individual, family, household or residence, or nonresidential customer cannot reasonably be identified or re-identified. Covered information, however, does not include information provided to the Commission pursuant to its regulatory responsibilities.

The Privacy Rules categorize various potential uses into two categories. “Primary purposes” are uses of the data that directly serve utility operations, are specifically authorized by the utility company or the Commission in connection with an energy-related program, or are for services required by state or federal law. “Secondary purposes” cover all other uses. Each category comes with its own list of obligations and security protocols relating to data transfer.

The Rules impose these obligations on both the IOU disclosing the data and the third party recipients of the data.

a. Primary Purpose

Under the Privacy Rules, a covered entity may only disclose covered information without customer consent if the data will be used for a “primary purpose.” The Privacy Rules identify four limited purposes that fit within this category:

- (1) [to] provide or bill for electrical power or gas,
- (2) [to] provide for system, grid, or operational needs,
- (3) [to] provide services as required by state or federal law, or as specifically authorized by an order of the Commission, or
- (4) [to] plan, implement, or evaluate demand response, energy management, or energy efficiency programs under contract with an electrical corporation, gas corporation, community choice aggregator, or electric service provider (when providing services to residential or small commercial customers), under contract with the Commission, or as part of a Commission authorized program conducted by a governmental entity under the supervision of the Commission. Privacy Rules § 1(b).^{20/}

Further, for the purposes of “analysis, reporting or program management,” disclosure of “aggregated usage data that is removed of all personally-identifiable information” is permissible, “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.” Privacy Rules § 6(g).^{21/}

According to D.11-07-056, “[t]o the extent other governmental organizations, such as the California Energy Commission or local governments, may seek Covered

20/ D.11-07-056, Attachment D, Rule 1(b).

21/ *Id.*, Rule 6(g).

Information in a manner not provided in these rules, the Commission will determine such access in the context of the program for which information is being sought absent specific Legislative direction.”^{22/} Accordingly, where the Privacy Rules do not explicitly provide for a certain form of disclosure, the Commission will determine on a case-by-case basis whether the disclosure is appropriate, and whether it is permissible under relevant legislation, such as the IPA.

Section 6 of the Privacy Rules provides additional guidance as to what qualifies as a “primary purpose,” and how disclosures must be carried out.^{23/} Under these provisions, an IOU may share covered information with a third party without customer consent (a) if “explicitly ordered to do so by the Commission” or (b) if the disclosure serves “a primary purpose being carried out under contract with and on behalf of the electrical corporation/gas corporation disclosing the data.”^{24/} These provisions indicate that the Commission intended for the “primary purpose” category to cover a fairly narrow selection of disclosure scenarios, largely directed to IOU operations (such as billing, maintenance, and the like by contractors), along with the noted services, when under direct Commission oversight.

“Primary purpose” disclosures create a chain of obligations that carry down to subsequent custodians of “covered information.” When disclosure occurs for a “primary purpose,” the covered entity disclosing the data “shall, by contract, require the third party to agree to access, collect, store, use, and disclose the covered information under policies, practices and notification requirements no less protective than those under which the covered entity itself operates as required under this Rule, unless otherwise

22/ D.11-07-056, pp. 47- 48.

23/ *Id.*, Attachment D, Rule 6.

24/ *Id.*, Rule 6(c)(1)a. and b.

directed by the Commission.”^{25/} Thus, a “primary purpose” recipient of covered information must employ at least the same privacy and security measures as those implemented within the IOU from which it collected the data. The Privacy Rules attach to all data that originates with the IOUs, regardless as to whom ultimately takes possession of it.

b. Secondary Purpose

Any purpose that does not fall within one of the above categories is considered a “secondary purpose” under the Privacy Rules.^{26/} IOUs are prohibited from disclosing covered information for any secondary purpose without the “prior, express, written authorization” of each utility customer represented in the data.^{27/}

Three limited exceptions to this requirement exist. A covered entity may only disclose smart grid data without customer consent in the following situations: (1) disclosure pursuant to certain types of legal process (such as a warrant or court order); (2) disclosure in “situations of imminent threat to life or property; and (3) disclosure “authorized by the Commission pursuant to its jurisdiction and control.”^{28/} Again, without an authorization order from the Commission, third parties not working on behalf of the utility may not obtain covered information without the prior, express, written authorization from utility customers.

c. Data Minimization Requirements

Under Section 5(c) of the Privacy Rules, covered entities must limit the disclosure of customer-specific energy usage data to only that which is “reasonably

25/ *Id.*, Rule 6(c)(1)b.

26/ *Id.*, Rule 1(e).

27/ *Id.*, Rule 6(d).

28/ *Id.*, Rule 6(d)1-3.

necessary or as authorized by the Commission” to carry out the specific purpose permitted under the Privacy Rules.^{29/} For data uses constituting “secondary purposes,” this means that the covered entity may not disclose more information than is reasonably necessary to carry out the specific purpose authorized by the customer in writing. As noted above, data minimization requires entities to consider, in advance of disclosure, what data is reasonably necessary for the agreed-upon purpose before disclosing the data.

d. Data Security and Breaches

Section 8 of the Privacy Rules establishes the minimum security requirements that covered entities must employ when in possession of covered information.

“Covered entities shall implement reasonable administrative, technical, and physical safeguards to protect covered information from unauthorized access, destruction, use, modification, or disclosure.”^{30/}

Furthermore, when a breach has been detected, a covered third party must notify the disclosing IOU within one week, and the utility must notify the Commission of all breaches affecting one thousand or more customers. Utilities are additionally obligated to file an annual report at the end of the each calendar year, chronicling all security breaches affecting covered information that year.^{31/}

e. Enforcement and Recourse for Privacy Rule Violations

If a recipient party fails to comply with its contractual obligations to handle the covered information in a manner “no less protective” than those under which the

29/ *Id.*, Rule 5(c).

30/ *Id.*, Rule 8(a).

31/ *Id.*, Rule 9(e).

originating entity operates—a “material breach” under the Privacy Rule—“the disclosing entity shall promptly cease disclosing covered information to such third party.”^{32/}

This legal and policy framework for privacy protection provided by the Electronic Frontier Foundation is a useful guide to the development of protocols and processes for sharing access to energy usage data for public purposes in a manner that protects personal privacy as envisioned in this proceeding.

E. Protection of Intellectual Property and Competitively Sensitive Information

In addition to the privacy framework, the parties identified other legal issues that need to be addressed as part of any energy usage data access policy. Among these are the protection of trade secrets, the protection of competitively-sensitive data that could be used to unfairly manipulate energy markets, and the protection of grid security information. These three issues are discussed briefly below.

The data collected by utilities to serve their customers, whether including or excluding personally identifiable information, may include intellectual property owned by the utilities and paid for by the utility’s customers generally, provided that the data is protected by the utility from disclosure to the public or third parties as such. For example, customer-related data and other intellectual property and trade secrets, including “customer lists” and other data about an entity’s customers, are all protected under California law as property of the entity. (California Civil Code Sections 654, 655; 3426- 3426.10 (Uniform Trade Secrets Act); *American Paper & Packaging Products, Inc. v. Kirgan* (2d Dist. 1986) 183 Ca. App. 3d 1318, 228 Cal. Rptr. 713.) Under Public Utilities Code Section 851, intellectual property constitutes an intangible asset of the

^{32/} *Id.*, Rule 6(c)(3).

utility that may not be sold, transferred or disposed of without CPUC authorization, and the “gain” in sale or disposition must be shared directly with utility customers under the CPUC’s “gain on sale” rules. Moreover, SB 1476, enacting the updated privacy rules for energy usage data in Public Utilities Code Section 8380, also makes clear that energy usage data *and any other personally identifiable information* collected by a utility for utility purposes may not be sold “for any purpose.” (Public Utilities Code Section 8380(b)(2).

Similarly, the CPUC and interested parties over the past several years since the 2000- 2001 energy crisis have promulgated rules restricting the access of “market participants” to customer-specific as well as aggregated energy usage data that could potentially be used by the market participants to manipulate prices or supplies in electricity procurement markets. See General Order 66-C, Public Utilities Code Section 583 and D.06-06-066 “Confidentiality Matrix” Rules. Thus, even if customer privacy is protected under energy usage data access protocols, the data itself still will need to comply with other legal and regulatory restrictions on disclosure, including the protection and use of intellectual property, and the prohibition on the disclosure of market-sensitive energy usage data that could be used unfairly to manipulate energy markets.

Finally, to the extent that the utility possesses data that, if disclosed, would compromise grid security, the utility must protect the data from disclosure regardless of whether it includes PII.

F. Alternative Views of Parties

None received as comments on draft Report.

IV. TECHNICAL FRAMEWORK RE. “RE-IDENTIFICATION”

In addition to discussing the legal and policy framework for energy usage data access, the parties addressed technical issues regarding the evolving risks that energy usage data, even with PII removed, could be “reverse-engineered” with the help of sophisticated computer programs available to anyone with access to the Internet that can “re-identify” the PII and potentially hack into or steal the identity and other personally identifiable information from utility customers. Experts retained by EFF provided technical advice assessing these risks for the Working Group. The following is a summary of the EFF experts’ advice and assessment, as contained in the memorandum attached to the Working Group Report as Appendix C.^{33/}

A. Risks to Customer Privacy Related to Disclosure of Customer Energy Consumption Data Collected from Smart Meters

Since the late 1980s, scientists have reported the ability to derive detailed behavioral information about a household or other premise from electrical meter readings. For example, Non-intrusive Appliance Load Monitoring (NALM) “use[d] temporally granular energy consumption data to reveal usage patterns for individual appliances in the house.” These usage patterns revealed, for example, time away from one’s home, cooking and sleeping habits, or the number of inhabitants in a particular household. Not long after its development in 1989, scientists described this technology as capable of remotely identifying patterns based on externally available meter information. In a 1989 paper, NALM creator George Hart simultaneously noted that identifying these patterns created the potential for invasions of private information.

^{33/} The footnotes to the technical assessment and advice by the EFF experts are omitted from the following summary for convenience, but are included in Appendix D.

By tracking the daily or hourly energy usage of a household, it is possible to create a consumption profile and deduce behavior for that household. It exposes not only energy consumption patterns overall, but also intimate behavioral information that most customers would not suspect is being shared, including travel, sleeping, and eating patterns, occupational trends, and even detailed information such as when children are home alone. This type of profiling is attractive for a number of purposes, from behavioral research to marketing. For an example of such consumption profiling used in the retail industry, Target Corporation used data on women's shopping habits to develop a pregnancy detection method so reliable that it often allowed for targeted advertisements before a woman had even revealed her pregnancy to others. Similar predictive algorithms can be used to extend noticeable trends in energy consumption data, such as using real-time data to determine when an occupant is at home for solicitation by the utility or some third party. To continue with family formation as an example, an occupant's consumption profile might indicate a new baby in the house. This would violate the home occupants' privacy and create risks of leaking personal information that the customer had not even considered exposed in the first place.

It is important to consider both existing profiling capabilities and those that are likely to arise in the near future. More recent scientific research on techniques for ascertaining information from energy data describes the developing ability to discern what video content is being viewed on a television or computer monitor. Known as "use-mode detection," this method relies on collecting energy data in real time. Lab scientists tested multiple television sets to determine that the content viewed on those devices left uniquely identifying energy signatures, known as electro-magnetic

interference (EMI). The same video content would produce the same repeatable EMI traces, even across different television sets. Under laboratory conditions, researchers were able to identify 1200 movies at a 92% accuracy rate by reviewing these trace EMI patterns.

Given the present and developing abilities to use energy data to detect appliance usage, discern regular household habits, and review the in-home consumption of video content or online information, the Working Group should assess and consider protections that guard such personal information from unauthorized collection and disclosure, in alignment with the requirements of the IPA and the CPUC's Privacy Rules.

B. Known Limits to Anonymization and Aggregation as Methods for Preventing Re-identification and Protecting Privacy.

Scientists now recognize that aggregating or anonymizing data to sufficiently prevent re-identification of an individual is almost impossible. As such, instead of relying directly on these techniques, instances of re-identification have prompted new efforts among computer science and privacy experts to “balance the risks and value of data sharing in a de-identification regime.” Existing and developing re-identification capabilities should be considered by the Working Group in making recommendations on the dynamic definitions of aggregated/anonymized data to give privacy-protecting protocols any value.

1. Anonymization

Anonymization techniques attempt to protect anonymity of data subjects by removing personal identifiers, such as names and addresses, from the data. Although anonymized data do not, on their own, point to specific individuals, numerous examples

demonstrate that reidentification can be achieved by comparing anonymized data with external information that contains corresponding data points.

As evident in the Netflix, AOL, Amazon and Massachusetts case studies described by the EFF experts in Appendix D to this Report, the removal of key identifiers, such as the data subject's name, address and birthdate, is insufficient to protect customer privacy.

Similar to the data sets in these case studies, energy data changes over time, allowing for noticeable patterns to appear. Unique energy signatures become personally identifying characteristics when compared to external information with shared data points. In addition, many of the same characteristics, such as name, address, birthdate, etc., are collected by utilities, as were in the Massachusetts government health data breach or by online service providers like Amazon, Netflix, and AOL. Further, many of these characteristics are available to the public on other databases, making it possible to identify an individual through linking other data.

These examples, among others, explain why anonymizing data by removing a few key attributes unfortunately may not prevent re-identification. In some cases, it was only a matter of hours before data considered "anonymized" was cross-referenced with external data and re-identified, compromising the data subject's privacy. As such, data that has been "anonymized" is often easily re-identifiable. Accordingly, data that has been processed with these types of anonymization techniques, without additional protective steps, would still be PII or "covered information" requiring protection under the IPA and the CPUC Privacy Rules.

2. Aggregation

The use of the term “aggregated data” has not been consistently used in this proceeding. Based on the scientific literature in this area, aggregated data does not include micro-data—i.e., the underlying, discrete records about individuals from which the aggregation is derived. Unlike attempts to anonymize data, for example by removing certain identifiers from individual records, aggregating data requires processing it such that there are no individual-level records, for example by computing the sum or the average of a group of individual households’ energy usage information. For purposes of the Working Group, “aggregated data” would not include the total annual or average annual energy usage for an individual household, precisely because the data pertains to a specific household.

Despite excluding micro-data, aggregated data can still leak private information. Traditional privacy protections for aggregation, such as the CPUC’s “15/15” rule originally adopted to protect the customer privacy of direct access customers, are sometimes referred to by computer scientists as “naïve aggregation rules” because of the ways in which PII can be reverse-engineered, for example when a requester makes more than one query into related data sets over a short period of time. To use an historical example, this one from as far back as World War II, it is now well known that re-identification of naively aggregated Census Bureau data helped the U.S. military locate and transfer Japanese-Americans to internment camps during World War II. Although naïve aggregation was considered an acceptable privacy policy in the 1940s, today’s Census Bureau employs a series of complex data-blurring techniques to

promote data integrity but maintain heightened security in response to such re-identification risks.

The 15/15 Rule is the most prominent “aggregation” model under review in this proceeding and has been previously approved by the Commission for aggregation of PII under the Commission’s Community Choice Aggregation tariffs. Data privacy experts have noted that a carefully crafted series of queries can generate aggregate results that, when looked at together, reveal customer-specific information. A brief explanation by EFF of how queries can work around the limits imposed by the 15/15 Guideline is given below, followed by an example of the risks of cross-referencing aggregated data with external sources. Please see Appendix A for further discussion of the data security issues EFF identified regarding the 15/15 Guideline.

a. Likely Smart Grid Data Leaks from Naïve Aggregation Rules

According to EFF, the 15/15 Rule and similar well-intentioned standards unfortunately exhibit fundamental flaws that render them unable to effectively defend customer privacy. Numerous researchers have addressed how a combination of queries can enable the re-identification of individuals represented in aggregate data, even though neither query on its own infringes the individual’s privacy.

To illustrate, imagine a quantitative query system under a standard like the 15/15 Rule, which ignores requests when the number of results is less than a particular threshold. In such a case, one need only ask two questions that meet that threshold to obtain an answer otherwise forbidden by the rule:

The first question:

“How many people in this database exhibit power usage patterns consistent with

using a television and video games in the afternoon, but patterns consistent with additional appliances, electric vehicles, and lights in the evening?”

([In an] interactive system designed to answer queries about the health care expenses of the Harvard faculty, which allows queries of the form “how many Harvard faculty satisfy X” where X is a search criterion that can involve attributes like age, health care expenses, and department.)

While “how many” questions may seem relatively safe when computed over a population of 2000+ individuals, they are not. By asking the question “How many Harvard faculty are in the computer science department, were born in the U.S. in 1973, and had a hospital visit during the past year?,” it is possible to find out whether one of the authors of these comments (S.V.) had a hospital visit during the past year (according to whether the answer is 0 or 1), which is clearly a privacy violation. A common “solution” to this sort of problem is to only answer queries whose answers are sufficiently large, say at least 10. But then, by asking two questions --- “how many Harvard faculty had hospital visits during the past year?” and “how many Harvard faculty, other than those in the computer science department and those born in the U.S. in 1973, had hospital visits during the past year?” --- and taking the difference of the results, we can obtain an answer to the original, privacy-compromising question.

The second question:

“How many people in this database who exhibit power usage patterns consistent with using a television and video games in the afternoon, but patterns consistent with additional appliances, electric vehicles, and lights in the evening, do not live at 100 Main Street?”

Although both questions provide aggregated results, the combination of these two questions has effectively "leaked" information about 100 Main Street. The first question essentially asked for the total number of homes where children are likely to be home alone in the afternoon. The second question sought the same information but excluded 100 Main Street. If the answers to these two questions are the same, then one can reasonably infer that there are no latchkey children at 100 Main Street; if the answers differ by 1, then one can reasonably infer that there are.

Unfortunately, it is very difficult for computer programs to detect the query combinations that breach customer privacy in advance. Professor Machanavajjhala pointed out at the January workshop in this proceeding that energy data is dynamic, not static. If aggregated data changes, then individuals can be uniquely identified in ways that computers were not programmed to protect against. For example, if data shows a new house on the block, then an attacker can look at changes in the neighborhood's energy consumption and subtract the new information to attribute change to the new home.

According to EFF, because this simple, two-query process for overcoming the 15/15 Guideline defeats its protective purpose, data masked in this manner is likely to remain re-identifiable. As such, like data that has been subjected to basic anonymization techniques, EFF would consider data aggregated according to these techniques to be "covered information" under the Privacy Rules, and would expose customers to re-identification risks if released without additional protective protocols in place.

b. Re-identification Using Pre-existing Information about an Individual

If an attacker or researcher has background information about an individual represented in an aggregated data set, re-identification becomes even easier. For example, in 2008, a research team, led by Nils Homer, then a graduate student at the University of California at Los Angeles, showed that in many cases, knowing a person's genome can help determine, beyond a reasonable doubt, whether that person had participated in a particular genome-wide test group.

Energy data is susceptible to the same sorts of attacks on other types of personal data. If an attacker knows the unique combination of appliances that a utility customer has in their kitchen, he can examine aggregate energy usage patterns to determine if the data signature corresponding to that combination of appliances fits the aggregate profile, which would lead to an inference that the customer was or was not included in the data.

Accordingly, with certain background information and data manipulation, data aggregated according to these techniques, as well, can easily be re-identified—especially as researchers, marketers, or others combine datasets—and EFF concludes that this would still be considered “covered information” under the Privacy Rules.

3. Expert Technical Solutions Are Required to Develop More Robust Privacy Solutions Because Current Anonymization and Aggregation Techniques May Fail to Protect Private Customer Data

As made clear in the analysis and examples above, when devising protocols for the disclosure of customer data, the Working Group participants should be aware that neither aggregation nor anonymization can be defined or evaluated in static terms if privacy is to be protected. Re-identification is a dynamic concept. Each time there is an influx of publicly available data, an advance in computer technology, or additional

collection of personally identifying characteristics, re-identification strategies will evolve. This means that the techniques required for the “safe” release of customer energy usage data or PII will likely also change. Any definitions adopted by the Working Group will need to accommodate this reality. In order to do this, the Working Groups need to consult experts in the fields of computer science, consumer privacy, and data security at each stage of developing data disclosure procedures, in order to understand the unfortunate, but genuine challenges in securely sharing data and to develop feasible solutions that overcome the shortfalls of anonymization and aggregation.

The EFF experts’ technical assessment of the risks of “re-identification” of PII even when anonymized or aggregated was a common issue that was extensively discussed throughout the Working Group Sessions. As a result, finding an acceptable, reasonable “solution” to the “re-identification problem” was identified by the Working Group as a primary task and condition precedent to development of a practical, standardized process for energy usage data access in this proceeding. Recommendations on a follow-up process for “solving” this problem, including the use of “data cubes,” are included in the evaluation of the various “use cases” and overall recommendations in Sections VI, VII, and VIII, below.

4. Current Anonymization and Aggregation Techniques are Needed if IOUs are to be Able to Continue to Release Customer Information Until More Robust Privacy Solutions may be Developed and Implemented

As a practical matter, some standard to evaluate the risk to customer privacy is essential if the IOUs are to be able to continue to release aggregated or anonymized data. Whatever standard is used must be able to evaluate millions of records of customer data and readily indicate whether information is reasonably aggregated or

anonymized. Notwithstanding the comments above regarding susceptibility to identifying customer-specific information due to the use of multiple datasets, external sources of information or other more sophisticated analytical techniques, the fundamental concept of 15/15 rule (adopted in D.97-10-031) represents the current method utilized by the IOUs for this purpose.

C. Alternative Views of Parties

No comments received on draft Report.

V. DEFINITIONS

A. Potential Recommendations

The February 27 ALJ Ruling proposed the following definitions for use in the interested parties' discussions of data access issues.

Aggregated data means a group or set of data points containing a sufficient number of points removed of personally-identifiable information where one cannot reasonably re-identify an individual customer based on, for example, usage, rate class, or location.

Anonymized data means a data set containing individual sets of information where all identifiable characteristics and information, such as, but not limited to, name, address, account number, or social security number, are removed (or scrubbed) so that one cannot reasonably re-identify an individual customer based on, for example, usage, rate class, or location.

(ALJ Ruling, February 27, 2013, pp. 12- 13.)

Over the course of the Working Group sessions, most of the interested parties as well as Commission staff converged on a somewhat different set of definitions to be used to discuss energy usage data access issues. These different definitions begin with the legal and factual definitions of "covered" customer-specific information in the Commission's privacy rules and the similar definition of "personal information" in the

California Information Practices Act. These definitions collectively define the term “personally identifiable information,” or “PII” for short, that is used routinely by privacy experts and advocates:

“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 nonresidential customer cannot reasonably be identified or re-identified. Covered information, however, does not include information provided to the Commission pursuant to its oversight responsibilities.” (CPUC Privacy Rules, Section 1(b).)

“Personal information’ means any information that identifies, relates to, describes, or is capable of being associated with, a particular individual, including, but not limited to, his or her name, signature, social security number, physical characteristics or description, address, telephone number, passport number, driver’s license or state identification card number, insurance policy number, education, employment, employment history, bank account number, credit card number, debit card number, or any other financial information, medical information, or health insurance information. ‘Personal information’ does not include publicly available information that is lawfully made available to the general public from federal, state, or local government records.” (California Civil Code Section 1798.80(e).)

These definitions do not exclude or supplant the “anonymized” and “aggregated” data definitions in the ALJ Ruling. Instead, “anonymized” and “aggregated” data are types of data that may or may not include PII.

For example, raw energy usage data from which names of customers and other personal information have been removed can be considered “anonymized” data. However, the technology available for “re-identification” of PII from previously “anonymized” raw data may mean that the “anonymized” data may or may not include PII, depending on what techniques may be used to “blur” or insert “noise” into the data.

Likewise, energy usage data that has been aggregated to a higher level than individual, “anonymized” customer records can be considered “aggregated” data. However, as with “anonymized” data, “aggregated” data may or may not include PII, because certain techniques may be used to “re-identify” or “disaggregate” the data to PII via sophisticated computer programs that compare aggregated data sets to other data sets in order to directly derive the identity of persons and other personal information.

For these reasons, based on an updated data classification method proposed by Commission staff in the Working Group sessions, the following factual definitions can be used to discuss energy data access. These definitions are taken directly from a May 22, 2013, presentation by Commission staff to the Working Group participants.

Types of Information Elements

Usage Info (UI) kWh over time unit	Billing Info (BI) \$ over time unit	Personal (PII)	Quasi (QII)
--	---	---------------------------	------------------------

- **PII (Personally Identifiable Information)**
 - Attributes associated with UI/BI directly identifying a customer (essentially, a unique identifier)
 - Examples: Name, address, SSN, [meter ID? Service account #?]
- **QII (Quasi-Identifiable Information)**
 - Attributes associated with UI/BI describing class or group; requires cross-referencing to identify a specific customer
 - Examples: Geography, customer class, rate, programs, ...
 - Some attributes are confidential if associated with PII



Information from Privacy Perspective

CI	AI
-----------	-----------

- **CI (Covered Information)**
 - UI/BI with PII. Requires customer consent (except if primary purpose)
- **AI (Anonymized Information) ~ (aka “Aggregate” Info)**
 - UI/BI without PII; No ability to identify customers (directly or indirectly)
 - Requires no customer consent
 - Two usage data types could be available as anonymized:
 - “Micro” data = individual meter-level readings
 - “Processed” (aka “aggregate”) data = micro data massaged via algorithm
 - Sum or Average, Random Sampling, “Laplacian”, “Subsample + Aggregate”
 - “Blurred” data, Other algorithms ?
 - To what extent can some QII be associated with Micro data while keeping risk of re-identification to an “acceptable” level?



Under the Commission staff definitions, the initial classification is whether the raw data directly includes PII, i.e. the name or other personal information about the customer, coupled with their quantitative energy usage or bill. The next part of the classification is to determine whether the data set includes “quasi-identifiable information” or “QII” even where no PII is directly included. This classification includes “re-identifiable” data that, with computer algorithms, can identify PII indirectly as discussed by the privacy experts sponsored by the Electronic Frontier Foundation in the Workshop sessions.

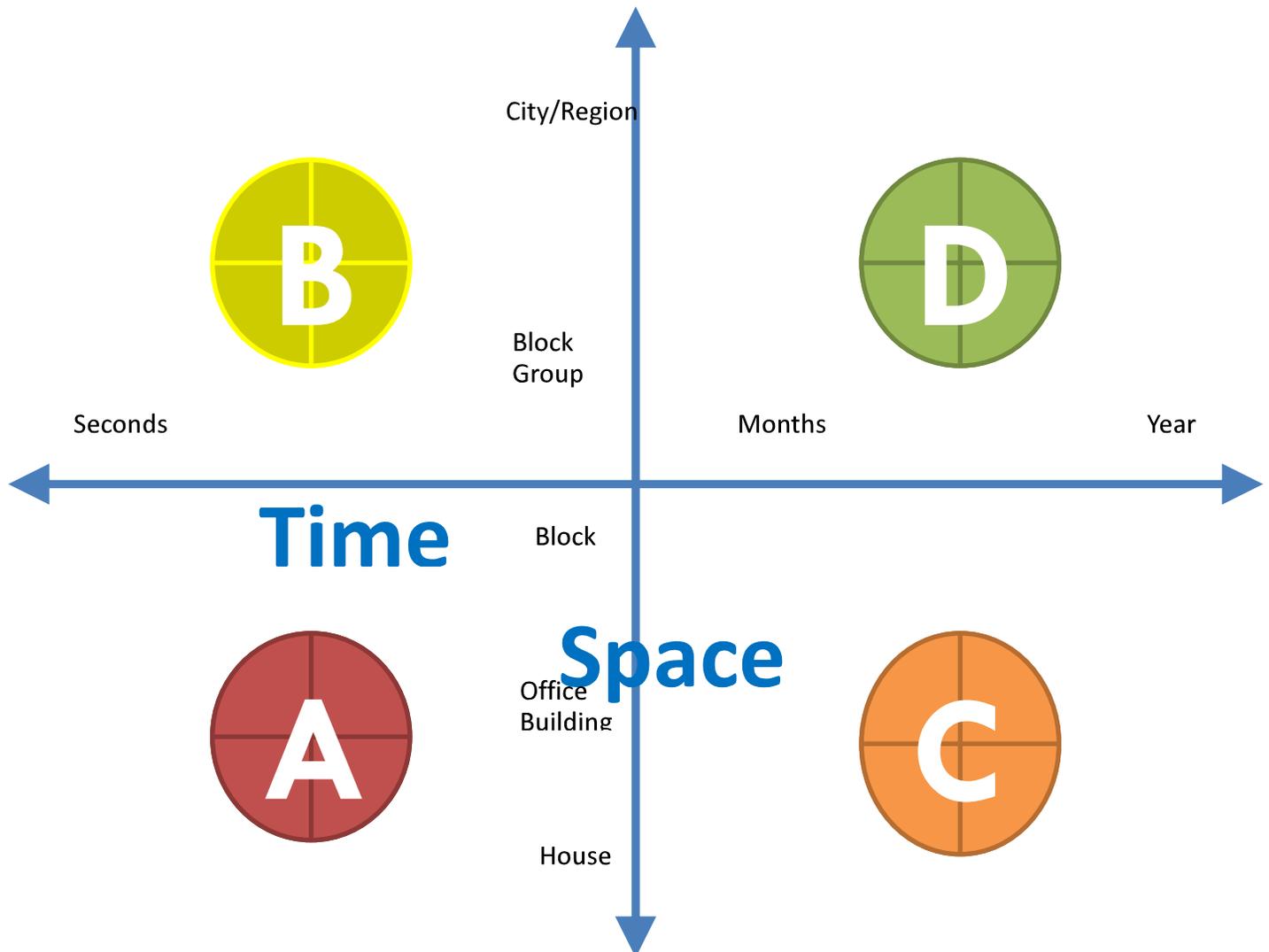
The final part of the classification is to separate the data into “covered information” which includes PII requiring customer consent for third-party access, and “anonymized information” (whether aggregated or raw “micro” data) that includes no PII directly or indirectly, by reason of “processing” of the data to remove or minimize the risk of “re-identification.” This final step in the classification then informs the policy options available to the interested parties and the Commission regarding whether the data can be made available to third-parties without violating the privacy of utility customers.

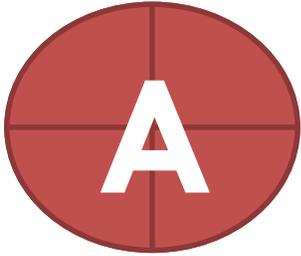
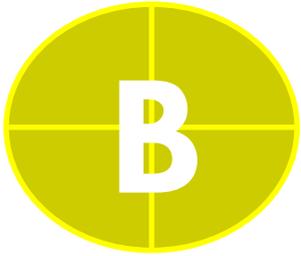
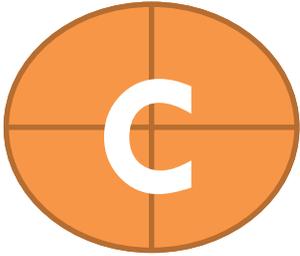
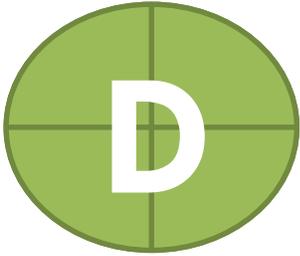
In the course of the working group meetings, LGSEC recommended a map of Energy Data Access as a suggested framework to enable clear and consistent discussion of energy usage data. The sensitivity of energy usage data varies with resolution, both geographic and temporal. An effective public policy will weigh this sensitivity alongside other key public interests recognized and prioritized in California law and policy, including effective stewardship of ratepayer investments in energy

efficiency, the energy resource loading order, public transparency, and greenhouse gas emissions mitigation.

Clear communication is essential as public concerns are weighed in the context of relevant laws. Accordingly, LGSEC recommended that the following map be used, which divides energy usage data into four 'quadrants' of resolution, labeled A, B, C, and D:

Figure 1: Energy Data Access Map. Divides temporal and geographic aggregation/resolution of energy usage data into four quadrants.



	Specific location and small time interval	Geographic aggregation and small time interval	Specific location and large time interval	Geographic aggregation and large time interval
Quadrant Label				
Sensitivity	High Clearly personally identifiable, includes details of timing, and specific activities can be exposed.	Moderate Location is not personally identifiable.	Moderate Location is identifiable. Monthly (or annual) data masks timing of specific activities, such as startup or occupancy.	Low Not personally identifiable. Monthly or annual interval masks specific activities.
Public Policy Value	Limited Contains more data than necessary for uses other than academic research or services provided with consent.	Moderate Illuminates load shape, limited use in efficiency program delivery.	High Informs priorities for investment and service delivery.	High Essential for greenhouse gas emissions tracking and city planning.
Useful to Study	<ul style="list-style-type: none"> • Limits of demand response • Customer to DR program signals • Effect of building age & shell on DR • Impact of rate design (including Critical Peak Pricing) • Plug load management • Effect of weather on residential PV output 	<ul style="list-style-type: none"> • Effect of geographically targeted measures on load shape. (Example: Intensive appliance installation in a targeted city/zone vs. a “control” area) • Demand response program design 	<ul style="list-style-type: none"> • Effect of building characteristics on energy consumption (such as building age, shell, most recent permit, etc.) • Relate energy use to demographic trends such occupant age, vulnerable population, linguistic isolation, proximity to cooling shelter for climate adaptation • Efficiency program effectiveness 	<ul style="list-style-type: none"> • Community greenhouse gas program impacts • Renewable resource planning • Effect of efficiency programs on community/ neighborhood energy use.

Suggested Protection	<p>Access only via:</p> <ul style="list-style-type: none"> • Customer consent • Academic research with NDA and protocols similar to Census protocols • Opt-out notification? 	<ul style="list-style-type: none"> • City or County aggregation: Public data (as with CSI program) • Block-group (or largest scale vulnerable to geographic disaggregation): Available to EE/renewable energy service providers under NDA, or via user interface designed to limit potential. 	<ul style="list-style-type: none"> • Available to building owner or designated representative for compliance with AB1103, CPUC benchmarking order, or local energy efficiency program/ ordinance. 	<ul style="list-style-type: none"> • Publicly accessible, published to the web, and updated annually.
-----------------------------	---	---	--	--

In reviewing the use cases below, LGSEC recommended that the Working Group consider into which quadrant each use case would fall. According to LGSEC, to the extent that parties (1) observe non-disclosure agreements and (2) are seeking data for monthly intervals, falling clearly into quadrants C and D, concerns about privacy violations should be minimized.

In the next section of the Report, these energy usage data definitions and classifications are applied to discuss and evaluate the 8 “use cases” listed in the February 27 ALJ Ruling and an additional 4 use cases provided by interested parties in the Working Group.

B. Alternative Views of Parties

SCE

As SCE will discuss in more detail in its opening comments on the Report, SCE believes that the Commission’s Privacy Rules and the statutory landscape discussed above, which should not be disturbed in this proceeding, already govern the extent to which a utility may disclose personally identifiable data without customer consent.

Creating a continuum of data sensitivity using the quadrants above complicates, rather than standardizes, the rules that are clear on their face. Moreover, SCE disagrees that NDAs can be used absent a Commission order to circumvent the requirement that PII data for secondary purposes requires customer consent for disclosure.

Solar City

We have concerns with the “factual definitions” presented in the draft report. On the graphic labeled “Information from a Privacy Perspective” the first sub-bullet under “AI (Anonymized Information) – (aka “Aggregate” Info)” includes the language “no-ability to identify customers (directly or indirectly”. We believe this language fails to acknowledge the outstanding issue of whether all customer-level usage data should be considered covered information, as discussed above. Accordingly, we respectfully request that the language be modified as follows:

UI/BI – without PII - No ability to identify customers (directly or indirectly), or data conveyed under contractual stipulations prohibiting efforts to re-identify customers”

We also have significant reservations regarding the Public Policy Value ascribed to more granular data in the LGSEC framework presented in the draft report. Keeping in mind that SolarCity’s use case does not envision third-parties receiving PII, but does envision conveyance of customer-level usage data, we believe it is problematic to deem the public policy value of this information as “limited”. To the extent that data generated from AMI can be effectively used to actually motivate customers to take action to reduce their energy consumption through the deployment of EE, DG, or other energy management solutions, we believe this far more directly serves the public policy

interests of the state than any of the other uses identified. The framework put forward appears to subordinate practical outcomes to academic endeavors and gives short shrift to the tangible benefits in which we believe the Commission was motivated when it approved the very large ratepayer investment in AMI. While academic and research activities that can also be facilitated by providing entities access to this data are indeed useful, we think it is highly problematic to declare that using the data to drive customer behavior as less important. Clearly there are tensions between using the data for this practical purpose and customer privacy, however, as we have articulated, we believe there are solutions to address these concerns that should be explored further. The balance that can be struck to enable utilization of the data to drive consumer behavior should be a priority since achievement of our greenhouse gas reduction goals fundamentally depends on customer embracing low carbon energy strategies. AMI data can be a tremendous asset in this regard.

VI. USE CASES: APPLYING DEFINITIONS AND PROTOCOLS FOR DATA SHARING TO USE CASES

A. Introduction

The February 27 ALJ Ruling listed eight “use cases” for energy usage data access, and asked that the Working Group evaluate and make recommendations on how access to energy usage data should or should not be provided under the use cases. In addition, during the course of the Working Group discussions, interested parties provided four additional use cases for evaluation. The total of 12 use cases are summarized in the table below. The full detailed description of each use case under the criteria in Attachment B to the February 27 ALJ Ruling is provided in Appendix A to this Report.

Use Case	In ALJ Ruling issued 2/27/13 (Yes/No)	Description
Use Case 1	Yes	Local Governments seeking access to aggregate data for use in creating legislatively required Climate Action Plans and implementation of energy efficiency programs.
Use Case 2	Yes	Research institutions seeking monthly billing data, which may be PII, to evaluate energy policies, including energy efficiency policies, and publishing results in aggregate, non-PII form.
Use Case 3	Yes	Research institutions seeking anonymous, individual hourly energy consumption data with other energy-related characteristics to evaluate energy policies, including energy efficiency programs and rate design, and publishing results as statistical coefficients. Thus, the data could be PII if it contained sufficient characteristics to permit reverse engineering, but the published results that describe the influence of energy-related attributes on consumption, would not be PII.

Use Case 4	Yes	Other governmental entities, like the CEC's Energy Upgrade California Program, seeking energy efficiency program participation data by customer identification number in order to cross-reference this data with other program data, and thereby evaluate government-sponsored, legislatively mandated programs, while publishing results in aggregate, non-PII form. Thus, this data is highly granular, but non-PII, while may be "reversed engineered," but the published results would be non-PII.
Use Case 5	Yes	Environmental non-governmental organizations, like the NRDC, requesting PII customer repayment history and energy consumption pre and post-retrofit for energy efficiency, to support general financial decision making on energy-efficiency investments through on-bill financing, and produce results that provide aggregate, non-PII findings that link energy usage to other relevant characteristics (e.g. geography, building characteristics, customer financial characteristics, and financing vehicle). In this case, the data is definitely PII, but the results – a decision whether a particular area, type of building, type of customer, or type of financing is viable – in non-PII.
Use Case 6	Yes	Solar installation company requesting monthly energy consumption data energy efficiency and participation in the net energy metering program, aggregated to a geographic area that protects PII, to reduce the product development and engineering costs in order to advance residential and commercial solar installations. In this case, the data, prior to aggregation, is PII, while the results – the identification of areas where solar power is financially feasible – is non-PII.

Use Case 7	Yes	Building owners and managers seeking monthly energy consumption by building to conduct building benchmarking analyses pursuant to AB 758 and AB1103, and publishing aggregate, non-PII results. In this case, raw data that is PII would likely be needed, but the results concerning the efficacy of the program, are not PII. Moreover, it may prove possible to anonymize such data via an algorithm.
Use Case 8	Yes	Energy efficiency contractor seeking CPUC-released aggregate data, similar to what the California Solar Statistics program releases, but using Energy Upgrade California data and other aggregate energy consumption data, to help validate the quality and value of energy efficiency work. Here, the raw data studied is likely PII but the program result – the validation of the energy efficiency work – does not necessarily reveal PII. Once again, it may prove possible to apply an algorithm that provides anonymization that cannot be reverse engineered.

Use Case 9 – Low income programs data sharing	No. Submitted by Department of Community Services and Development (CSD)	Governmental agencies, like CSD that implement federally-funded energy efficiency programs for low-income persons such as the Low-Income Home Energy Assistance Program (LIHEAP) and the Department of Energy Weatherization Assistance Program (DOE WAP), endeavoring to coordinate the delivery of energy services with similar services provided by IOUs under CARE and Energy Savings Assistance Program (ESAP), through the reciprocal sharing of: 1) historical, non-PII, property-centric weatherization data; 2) historical PII weatherization data; and 3) customer/ client PII, involving eligibility, account information and energy usage data, all shared with the consent of the customer/ client.
---	---	--

Use Case 10 – Title 24 Compliance	No. Submitted by Christine Awtrey from Efficiency & Renewable Energy Division California Energy Commission (CEC)	As a means of verifying compliance with the Title 24, Part 6 Building Energy Efficiency Standards as they relate to HVAC system efficiency and installation requirements, the Compliance and Enforcement Office needs to determine what HVAC systems are being imported into and sold in California for installation within the state. This determination can be made through the tracking of an HVAC’s serial number, whereby any HVAC unit sold in the state will have its serial number entered into a database so that the serial numbers in this database can be compared to the serial numbers of HVAC units installed under the permitting process in local enforcement agencies throughout the state. This information can also be used for, and should be a requirement of, any HVAC rebate program within the state, whereby a rebate will be issued only for those HVAC installations where the proper permitting by the local enforcement agency has been accomplished.
Use Case 11 – EE program implementer/contractor	No. Submitted by Robbie Addler from BPTS (aka Faraday)	Energy efficiency program implementer, contractor, consultant, research institution, city, county government, or other entity requesting PII individual energy consumption data, payment data, energy efficiency program participation, and retrofit activity to identify trends in customer participation in efficiency programs and retrofit activity. The requested data must be PII to allow linkage with other relevant data, but the results of analyses (e.g. trends) would not include PII.
Use Case 12 – DECA	No. Submitted by Aram Shumavon from Distributed Energy Consumer Advocates (DECA)	The DECA use case provides the public with a working model of the majority of California’s electricity grid, with a particular focus on the ability to model all electricity consumers’ consumption at sub-hour time interval and to tie that data to actual weather conditions, building data, etc. The use case allows for the overlaying of wholesale market data including wholesale production run simulations providing prices and emissions. Expected users of this data are policy advocates, distributed generation providers, energy efficiency marketers and evaluators, and local governments.

B. Use Case 1: Local Governments Seeking Aggregated Data

1. Description of Use Case and Benefits of Data Access

Use Case 1 was described in the February 27, 2013, ALJ Ruling as “Local Governments seeking access to aggregate data for use in creating legislatively required Climate Action Plans and implementation of energy efficiency programs.”^{34/} In the

34/ This use case does not include community choice aggregation programs that access data through California Public Utilities Code § 366.2(c)(9) and through primary purposes in accordance with D.12-08-045 and other applicable privacy rules and orders.

Working Group sessions, the Local Government Sustainable Energy Coalition (LGSEC) further described this use case as follows:

Consistent with AB 32 and the CPUC Long Term Energy Efficiency Strategic Plan, many local governments are adopting electricity, natural gas and greenhouse gas (GHG) reduction goals and action plans, and assuming a lead role in offering programs and policies aimed at achieving these goals in their communities and regions.

Achieving energy and GHG reduction goals requires access to data that enables local governments to effectively evaluate and report progress toward adopted goals and to evaluate the efficacy of specific programs and policies. Proper evaluation of programs and policies informs resource allocation moving forward and ensures the highest and best use of ratepayer and other public funds in the implementation of energy programs by local governments.

Specifically, local governments need access to three categories of energy usage data, on a monthly basis:

1. Aggregated data that illustrate the status of progress toward adopted energy and GHG reduction goals, e.g., total monthly residential energy use at the block group level
2. Aggregated data that illustrate the outcomes of a given energy program, e.g., total monthly electricity savings from the Energy Upgrade CA program at the community or sub-community level
3. Granular, anonymized data at the address level, on a monthly usage basis, that provide insight into how energy use changes as properties participate in programs, and identify unmet needs in order to plan for future programs

Local governments' ultimate objective is to effectively reduce energy consumption and the associated costs and GHG emissions in their communities. Achieving this objective saves residents, businesses, municipal governments, and utilities money. This objective can only be achieved with the provision of various forms of utility data provided to local governments in a timely, user-friendly, and consistent manner. Local and regional government entities need these data in order to meet state legislative requirements, comply with local/regional policies and ordinances, and implement programs mandated by the CPUC and paid for by ratepayers.

Local governments need aggregated energy usage data to achieve the following:

- Provide timely and consistent reporting on energy use and greenhouse gas trends to locally elected government bodies and community stakeholders
- Evaluate the efficacy and impact of energy policies and programs operating in their communities
- Identify energy program participation rates and areas within a community that are potentially underserved by a given program or programs

Local governments need granular data on a monthly basis to achieve the following:

- Evaluate meter and building-level energy consumption pre- and post-energy retrofit
- Correlate energy usage to other relevant characteristics (e.g. geography, building characteristics, and customer financial characteristics). This enables not only improved targeting of future programs, but more effective messaging to increase participation and effectiveness

- Conduct building benchmarking analyses
- Identify unmet needs to plan for future programs that will ensure the

highest and best use of ratepayer and other public funds.

Granular data would need to be protected from entering the public sphere through appropriate privacy protocols. There may be instances where local governments need data on a more frequent basis than monthly in order to account for weather.

2. Evaluation and Potential Recommendations on Use Case 1

LGSEC and the Working Group participants discussed how this use case compares to existing energy usage data access already provided by the utilities to local governments under existing utility programs, such as PG&E's "Green Communities" program and the utilities' "local government partnership" energy efficiency programs under which the utilities and local governments share energy usage data under energy efficiency partnership agreements that constitute "primary purposes" under the CPUC's privacy rules.

The Working Group also discussed and understood that, with the exception of building benchmarking programs discussed below, the type of energy usage data requested by local governments for climate planning and energy efficiency programs is not PII data, but instead more likely monthly energy consumption data that is adequately "anonymized" or aggregated at higher levels, such as zip code, Census tract, or customer class levels where appropriate and practicable. This type of data may be capable of being standardized and provided on a pre-formatted basis to local governments by all the utilities without the need for extraordinary privacy protections or

reviews. To the extent that the data is proprietary to the utilities or other third parties, or is competitively-sensitive (e.g. historical or forecast loads used for utility procurement), a standard non-disclosure agreement can be used to protect the data from unauthorized disclosure. Thus, this type of non-PII monthly energy consumption data, aggregated and pre-formatted, may be included in the streamlined utility process for data access discussed in Section VII, below, provided that the parties can agree on a standardized format that adequately protects the privacy of utility customers.

The one exception to access to this type of data discussed at length by the Working Group is building benchmarking data requested by local governments for climate planning or energy efficiency purposes that is similar to the building benchmarking data requested directly by third-party building manager and landlords from building tenants under Use Case 7, below. As discussed in more detail below, an acceptable balance between customer privacy and data access for this building benchmarking potentially could be provided by allowing access to the energy usage data of building tenants under Use Cases 1 and 7 if and only if the data is aggregated at a level of 20 or more tenants and otherwise complies with the so-called “15/15” rule used under the CPUC’s Direct Access tariffs. In addition, any building benchmarking data that is published or made public under such building benchmarking programs would be required to use data “blurring” or “processing” techniques to avoid direct or indirect disclosure of customer PII. Further details on this potential approach are provided under Use Case 7, below.

3. Alternative Views of Parties

None received as comments on draft Report.

C. Use Case 2: Research Institutions Seeking PII

1. Description of Use Case and Benefits of Data Access

Use Case 2 is described in the ALJ Ruling as “Research institutions seeking monthly billing data, which may be PII, to evaluate energy policies, including energy efficiency policies, and publishing results in aggregate, non-PII form.” In addition, the UCLA California Center for Sustainable Communities (CCSC) provided the following additional information on this use case:

CCSC seeks energy data to identify current patterns and drivers of electricity consumption, to target and evaluate energy efficiency investments, and to help the State of California achieve its energy and environmental policy objectives.

This research is intended to provide significant public benefits. To date, there is little baseline knowledge of the patterns and drivers of electricity consumption, meaning decision-makers are flying blind as they try to implement policies such as AB 32 and SB 375. Only with such baseline understanding of consumption patterns can we begin to effectively reduce consumption through targeted investments. Our research provides significant public benefits by helping minimize the implementation costs and unintended consequences of achieving such policies.

CCSC’s analysis requires monthly electricity consumption data at the individual customer account level, with each account identified by customer class (e.g. single-family residential, multi-family residential, commercial, industrial, municipal operations, etc.), for a period of at least 7 years. Commercial and industrial accounts should also be identified by NAICS code.

The data needs will vary immensely by research project across all data parameters, including temporal resolution (e.g. annual, monthly, interval), geographic resolution (e.g. ZIP, ZIP+4, census block, individual account), and whether identification by tariff or customer class is required. For this reason, a flexible approach to data provision serving public interest benefits should be pursued.

Data must include consumption information by account for at least the past 7 years, updated on an ongoing basis. This is necessary for developing an understanding of trends in energy consumption over time.

Data must be made available in an accessible electronic format (e.g. Excel, Access, comma- or tab-delimited file, etc.).

The Energy Institute at Haas and the Energy Commission also generally supported the need for access to energy usage data for public policy research and planning.

2. Evaluation and Potential Recommendations on Use Case 2

The Working Group discussed this use case as being the general use case relating to any and all requests by third-party researchers for access to energy usage data for research purposes that do not directly relate to or support utility operations or programs. In other words, under the CPUC's privacy rules, research that is for a "primary purpose" such as supporting or implementing energy efficiency or energy management programs or other utility operations, is already authorized under the privacy rules, and thus subject to appropriate privacy protections, security protocols and utility-third-party contractual agreements.

Instead, this use case covers requests for data access by researchers that may not directly support or relate to utility operations, but nonetheless may support California's overall energy and environmental policy goals, such as by researching ways to model energy usage and demand on a statewide or regional basis, rather than only on a utility-specific or program-specific basis. This type of "public interest" research can provide general "public" benefits to consumers and businesses in California, if credibly scoped and conducted.

Both CCSC and the Energy Institute at Haas cited precedents for energy usage data access under which researchers may gain access to anonymized or aggregated energy usage data under appropriate non-disclosure agreements for purposes of

conducting specific energy and environmental policy research. The type of data and level of aggregation sought under these precedents has generally been anonymized monthly energy consumption data aggregated at the zip code, zip code+4 or Census Tract level. To the extent the energy usage data needs to be transformed into a relational data base containing other data attributes, such as building size or type or income characteristics, the researchers are responsible for inserting that additional data in a manner that retains the privacy protections associated with the energy usage data. The researchers are free to publish the results of their research as long as the results do not disclose customer-specific information directly or indirectly.

Public interest research projects as proposed by CCSC and the Energy Institute at Haas potentially could be supported by the utilities through a streamlined data access process that produces standardized, pre-formatted data sets to researchers under standardized non-disclosure agreements and subject to recover of the reasonable costs of setting up and implementing the data access on a routine basis. In addition, to the extent that the researchers are affiliated with or funded by a California state agency or the University of California system, the privacy protocols for the research should comply with the applicable requirements of the California Information Practices Act. To that end, researchers potentially could be provided access to pre-formatted standard anonymized or aggregated data sets as long as (a) the utility costs and effort required are reasonable, (b) the data set is protected from re-identification, and (c) the researchers provide for reimbursement of all or a reasonable portion of the costs of the data access on a standardized fee basis, in order to ensure that utility ratepayers do not bear significant costs of research activities that provide no direct benefit to them.

3. Alternative Views of Parties

None received as comments on draft Report.

D. Use Cases 3 and 4: Research Institutions and Governmental Agencies Seeking Anonymized Data For Research or Analysis that Could Be Re-identified as PII

1. Description of Use Cases and Data Access Benefits

Use Cases 3 and 4 are similar enough to be grouped together for evaluation.

The ALJ Ruling described Use Case 3 as follows: “Research institutions seeking anonymous, individual hourly energy consumption data with other energy-related characteristics to evaluate energy policies, including energy efficiency programs and rate design, and publishing results as statistical coefficients. Thus, the data could be PII if it contained sufficient characteristics to permit reverse engineering, but the published results that describe the influence of energy-related attributes on consumption, would not be PII.” The ALJ Ruling described Use Case 4 as follows: “Other governmental entities, like the CEC’s Energy Upgrade California Program, seeking energy efficiency program participation data by customer identification number in order to cross-reference this data with other program data, and thereby evaluate government sponsored, legislatively mandated programs, while publishing results in aggregate, non-PII form. Thus, this data is highly granular, but non-PII, while [it] may be “reversed engineered,” ... the published results would be non-PII.”

The Energy Commission further described these use cases as follows:

Federal and state agencies, and local governments are tasked with formulating policies to reach energy efficiency and greenhouse gas emissions goals without having a rich set of energy use data to base their policies on. For example, knowing the

average consumption of a type of building is important, but knowing the median and standard deviation of energy use per square foot of small retail buildings built between 1970 and 1980 in the central valley is much more useful. This use case sets up the parameters by which governmental agencies can be assured access to both energy use data and the PII associated with it in a way that indemnifies the utilities supplying the data.

The objective is to allow access to energy use data, including PII for the purposes of formulating public policy. This provides value for the ratepayers by allowing for policies that are better suited to well-substantiated market conditions. The use of real world data to guide policy decisions will augment the value provided to ratepayers of such policies.

According to the Energy Commission, the current practice is that policy-setting agencies “have no access to data. They can request data and if the aggregation is large enough (i.e., entire cities’ Climate Action Plan), the utilities will often provide highly aggregated data under an NDA. The level of aggregation that is appropriate for compliance with PUC Sections 8380 and 394.4(a), however, make it impossible in the Energy Commission’s view to answer basic questions about the distribution of buildings sizes and energy use within particular climate zones or areas of construction.”

2. Evaluation and Potential Recommendations on Use Cases 3 and 4

The utilities have entered into information sharing arrangements from time to time with the Energy Commission and other governmental agencies, subject to non-disclosure agreements that protect the privacy and security of customer-specific information as required by the CPUC privacy rules and the California Information

Practices Act. However, unlike Use Cases 1 and 2, Use Cases 3 and 4 appear to assume that governmental agencies should be provided broad rights to collect customer-specific PII from utilities for purposes unrelated to utility programs or operations or regulatory oversight of the utilities, as long as the government agencies agree to protect the customer information from public disclosure. The Working Group believes that granting state government agencies such broad access to customer-specific energy usage data is premature, unnecessary and possibly in violation of the California Information Practices Act. It is premature and unnecessary because, like the research activities in Use Cases 1 and 2, the need of the Energy Commission for data access for policy and analysis can be fulfilled without identifying specific customers; energy usage data can be anonymized and aggregated and still provide the Energy Commission and other state agencies with relevant and accurate data for policymaking and analysis. Such a limitation is also consistent with the California Information Practices Act, which prohibits the Energy Commission and other state agencies from collecting customer-specific information unless customers are notified and consent in advance, or the collection of the information is for a specific, statutory regulatory purpose.

Accordingly, the energy usage data under Use Cases 3 and 4 should be provided to government agencies such as the Energy Commission on an anonymized, aggregated non-PII basis, and subject to appropriate non-disclosure and cost recovery terms similar to those applicable to energy usage data made available to researchers under Use Cases 1 and 2.

3. Alternative Views of Parties

None received as comments on draft Report.

E. Use Case 5: Environmental or other Non-Governmental Institutions Seeking PII for Energy Efficiency Programs (e.g. Financing, Building Benchmarking)

1. Description of Use Case and Data Access Benefits

The ALJ Ruling described Use Case 5 as follows: “Environmental non-governmental organizations, like the NRDC, requesting PII customer repayment history and energy consumption pre and post-retrofit for energy efficiency, to support general financial decisionmaking on energy-efficiency investments through on-bill financing, and produce results that provide aggregate, non-PII findings that link energy usage to other relevant characteristics (e.g. geography, building characteristics, customer financial characteristics, and financing vehicle). In this case, the data is definitely PII, but the results – a decision whether a particular area, type of building, type of customer, or type of financing is viable – in non-PII.”

Two parties, NRDC and Brighter Planet Technology Services/Faraday, provided additional information on this use case through the course of the Working Group Sessions. This additional information clarified that part of Use Case 5 can be considered under Use Cases 1 and 2, to the extent that anonymized, aggregated, non-PII energy usage data may be made available for research and analysis by non-governmental organizations such as NRDC and other environmental groups under terms and conditions similar to those proposed for public interest research under Use Cases 1 and 2.

In addition, NRDC and Faraday addressed the part of Use Case 5 that involves requests by third-parties for customer-specific financial and billing information for purposes of planning and conducting so-called “on-bill financing” programs for energy efficiency retrofits or other customer-directed energy management programs. The primary benefit of making this customer information available to third-parties is that the third-parties, including financial institutions, would be better able to market and solicit utility customers to enter into lending arrangements with the third-parties under on-bill financing programs.

2. Evaluation and Potential Recommendations on Use Case 5

To the extent that Use Case 5 includes access to anonymized, aggregated non-PII energy usage data for the same public interest research and policymaking purposes as Use Cases 1 and 2, non-governmental organizations such as NRDC and other environmental groups potentially could be provided access to the data under the same protocols and process as recommended for researchers under Use Cases 1 and 2.

However, non-governmental organizations and financial institutions should not be provided with customer-specific billing, credit and collection information for purposes of on-bill financing programs unless the customer authorizes access to such information as required under the CPUC’s privacy rules and the utilities’ tariffs. As also discussed below in connection with Use Cases 6 and 8, commercial or private uses of customer-specific information are fundamentally different than public interest research and governmental access authorized by statute. Utility customers have a broad expectation that the privacy of their finances and billing records with their local utility will be strictly protected, and that third-parties will not obtain access to such sensitive, confidential

data without the customers' consent or a valid, lawfully authorized order, such as a court-approved subpoena.

Accordingly, the Working Group recommends that the CPUC continue to restrict access by commercial entities to customer financial, billing, and credit and collection information, unless the customer has expressly authorized the access in accordance with CPUC precedents and utility tariffs implementing those precedents.

3. Alternative Views of Parties

None received as comments on draft Report.

F. Uses Cases 6, 8 and 11: For-Profit Commercial Entities, e.g. Solar PV Installers and Energy Efficiency Contractors, Seeking PII for Commercial Use

1. Description of Use Cases and Benefits of Data Access

Use Cases 6, 8, and 11, as discussed by the Working Group and supplemented by representatives of solar vendors and energy efficiency contractors, are sufficiently similar to be evaluated together. The ALJ Ruling describes Use Case 6 as follows: “Solar installation company requesting monthly energy consumption data energy efficiency and participation in the net energy metering program, aggregated to a geographic area that protects PII, to reduce the product development and engineering costs in order to advance residential and commercial solar installations. In this case, the data, prior to aggregation, is PII, while the results – the identification of areas where solar power is financially feasible – is non-PII.”

The ALJ Ruling described Use Case 8 as follows: “Energy efficiency contractor seeking CPUC-released aggregate data, similar to what the California Solar Statistics program releases, but using Energy Upgrade California data and other aggregate

energy consumption data, to help validate the quality and value of energy efficiency work. Here, the raw data studied is likely PII but the program result – the validation of the energy efficiency work – does not necessarily reveal PII. Once again, it may prove possible to apply an algorithm that provides anonymization that cannot be reverse engineered.” Use Case 11, based on information submitted by various parties during the Workshop Discussions, is very similar to Use Case 8 and therefore can be considered together with that use case.

Although it did not actively participate in the Working Group sessions or discussions, Solar City, a solar vendor, submitted supplemental information on Use Case 6, as follows:

Solar installation and energy efficiency companies will analyze anonymized, household level energy consumption and billing data to identify customers/households that may benefit from energy services. After analyzing energy bills, these third parties will develop proposals for these households and submit them to an Energy Data Center. Customers will have the option to select their preferred communication method (i.e. email, phone, through portal, etc). Based on the communication preferences indicated by the customer, the Energy Data Center will notify customers that trusted third-parties have developed household specific proposals, including estimates of energy and bill savings, and would like to market their services. If customers opt-in, the Energy Data Center will forward the detailed proposals from third-parties to the customer. Personally identifiable information is never revealed to any third party, unless the customer contacts the third party directly.

The objective is to analyze customer usage data to better understand opportunities to deploy distributed renewable energy and energy efficiency improvements at customer's home, reducing their energy consumption and bills. This will reduce customer acquisition costs, a major lever to facilitate more widespread adoption of distributed renewable energy and energy efficiency, by helping third party renewable energy and efficiency installers present data-driven and tailored proposals to customers who can most benefit from their services. This will also increase precision of solar and home retrofit systems, since real data helps right-size systems.

2. Evaluation and Recommendations on Use Cases 6, 8 and 11

Although the Solar City proposal assumes creation of a centralized Energy Data Center (an initiative that is not within the scope of the Working Group pursuant to the ALJ Ruling), nonetheless Use Cases 6 and 8 can be considered under the assumption that the utilities would fulfill the functions assumed by Solar City to be performed by the Energy Data Center. Under this configuration of Use Cases 6 and 8, the key issue is whether the use of a "neutral" third-party – whether the utilities or some third-party independent of the solar vendors and energy efficiency contractors, is sufficient to protect the privacy of customer-specific energy usage data made available for what is clearly a commercial, profit-making purpose. In addition, the logistics and protocols of ensuring that the third-party is genuinely "independent" and "neutral" toward the profit-making commercial motives of the solar and EE vendors is an issue.

The presence of a neutral "intermediary" between the customer-specific PII and the commercial vendors is insufficient to protect customers' expectations of privacy and probably not lawful under the privacy statutes and rules. As made clear in the

descriptions of Use Cases 6 and 8, the access to customer-specific PII may be consistent with California’s energy and environmental policies, but it is clearly for a commercial, profit-making purpose, not a governmental purpose. As such, the privacy policies and rules are also clear: The commercial, non-utility purpose of the data access is a “secondary” purpose for which express customer consent is required. (CPUC Privacy Rules 1(e) and 6(d).) In addition, making available such data to commercial entities for such a commercial purpose without customer consent is likely violative of Public Utilities Code Section 8380(b)(2), which expressly prohibits a utility from selling a customer’s electrical or gas consumption data “or any other personally identifiable information for any purpose.”

Alternatively, solar vendors, energy efficiency contractors, and other third-party commercial entities can work with the electric utilities on the implementation of the utilities’ Customer Data Access programs if and when approved by the CPUC. The CDA programs will offer third-parties with streamlined, electronic access to bulk amounts of customer-specific energy usage data under a standardized, uniform customer consent process. The CDA programs will provide third-parties with access to customer-authorized, customer-specific energy usage data as requested in Use Cases 6 and 8 without violating customer privacy.

3. Alternative Views of Parties

Solar City

In describing SolarCity’s use case, the report also appears to gloss over or ignore some important distinguishing elements, in particular the fact that under SolarCity’s use case *no PII would be conveyed to third-party entities*. SolarCity’s proposal would allow third

parties access to customer-level energy usage data, but for reasons and under conditions described above, we do not believe the conveyance of this information requires prior customer consent since we do not believe it is covered information.

Additionally, the draft report dismisses SolarCity's use case by inappropriately and prematurely stating that the use case would violate the Commission's privacy rules under which covered information cannot be conveyed for a secondary purpose without prior customer consent. We fundamentally disagree that the data that we are seeking is covered information. At a minimum, this a contested issue, and we believe it would be inappropriate for the Draft Report to take unequivocal stance on this issue.

In dismissing our use case out of hand, the draft report also makes a number of legal and factual assertions, among them "The presence of a neutral intermediary between the customer-specific PII and the commercial vendor is insufficient to protect customers' expectations of privacy and probably not lawful under the privacy statute and rules". Again, these views are highly contestable and we do not believe it reasonable to include them as reflective of the working group's views. Additionally, the Draft Report incorrectly characterizes our use case as requiring the conveyance of PII. Our use case was specifically developed to avoid the need for a third-party to receive PII. Thus, we respectfully request these corrections to the Draft Report.

Again, we sincerely appreciate the efforts by the CPUC , the IOUs and the other stakeholders that have been actively engaged in this effort. AMI data represents a significant opportunity to advance key state policies, in particular efforts to drive customer adoption of EE, DG and other energy management solutions that are fundamental to achieving the state’s greenhouse gas reduction goals. By effectively using this data, while recognizing the legitimate privacy concerns, the state can fully realize the promise of the multi-billion dollar investment it has made in AMI.

G. Use Case 7: Building Owners/Managers Seeking PII to Comply with Building Benchmarking Regulations, e.g. AB 758/AB 1103

1. Description of Use Case and Benefits of Data Access

The ALJ Ruling described Use Case 7 as follows: “Building owners and managers seeking monthly energy consumption by building to conduct building benchmarking analyses pursuant to AB 758 and AB1103, and publishing aggregate, non-PII results. In this case, raw data that is PII would likely be needed, but the results concerning the efficacy of the program, are not PII. Moreover, it may prove possible to anonymize such data via an algorithm.”

Representatives of the City and County of San Francisco provided additional information regarding the beneficial uses of building benchmarking data for compliance with building benchmarking regulatory standards, such as under CCSF’s building benchmarking ordinance and the Energy Commission’s AB 1103 statewide building benchmarking regulation. The Energy Commission also provided information demonstrating the benefits of creating a statewide and nationwide building benchmarking database for use by building owners, building design professionals, policymakers, and property managers.

2. Evaluation and Potential Recommendations on Use Case 7

The Working Group extensively discussed the privacy/data access tradeoffs inherent in building benchmarking programs. On the one hand, many buildings are owned or managed by landlords who have no routine access to the energy usage of individual tenants who are the customers of record of the electric and gas utility service providers, and thus the normal privacy rules preclude the utilities from disclosing the tenants' energy usage to the building owner or landlord if the disclosure would identify the customer without their consent. On the other hand, in many cases, where utility usage is master metered in buildings, the only way to identify whole building energy usage is through disclosure of the master-metered tenants' energy usage. Under this situation, the utilities and building owners are caught in the middle – they both want to make available the whole building usage, but they also want to protect the privacy of customer-specific energy usage data. Under the AB 1103, California's statewide building benchmarking program, utilities which receive requests from building owners for building energy usage data are required to aggregate any customer-specific or tenant-specific usage data or use other means to protect the privacy of the utility customer unless the customer affirmatively authorizes disclosure of their energy usage data.^{35/}

The normal solution to this problem is for the landlord, through its lease with the tenant or through other agreement, to obtain the tenant's consent to the disclosure of their private energy usage to the landlord for purposes of building benchmarking. The

35/ 20 Code of California Regulations, Section 1684(b) requires that "If a building has a utility or energy provider account for which the owner is not the customer of record, the utility or energy provider shall aggregate or use other means to reasonably protect the confidentiality of the customer. A utility or energy provider may verify a request or ask for clarification before releasing data."

other solution to this problem is for the utility and landlord to adequately aggregate the tenants' usage so that the customer's identity is not disclosed as part of the aggregated whole building usage. However, neither of these solutions is completely satisfactory, because either the tenants are unwilling or unavailable to consent to disclosure of their private monthly energy usage, or there are too few tenants in the building to avoid "re-identification" of the tenants' identities even when the usage is aggregated to a whole building level.

After extensive discussion on the "re-identification" risk between the representatives of CCSF and the privacy experts retained by EFF, a potential pragmatic approach was discussed that would mitigate the privacy risk to an acceptable level while at the same time making the collection of building benchmarking data more convenient and streamlined for building owners and regulators. Under this approach, privacy risks would be mitigated by allowing aggregation of tenants' usage under a slightly stricter version of the "15/15" rule. If tenant usage were aggregated at no less than 20 or more tenants, and no tenant represented more than 15 percent of the whole building usage, then such aggregation might be considered sufficient under the privacy rules and the technical standards for avoiding "re-identification." However, it should be noted that neither CCSF nor EFF and its technical experts reached agreement that an aggregation approach like this is practical enough to achieve the goals of benchmarking or technically sufficient to avoid re-identification; EFF's perspective is that additional "blurring" or "processing" of the aggregated data would still be necessary if the goal is to fully mitigate the risk. Nonetheless, the privacy risk may be considered as acceptable, given the benefits of building benchmarking and the additional privacy controls that

would be applied to the aggregated data, including a non-disclosure agreement with the landlord and the requirement that any building benchmarks that would be made available publicly would not be aggregated energy usage benchmarks, but instead comparative benchmarks that “mask” the building-specific quantitative energy usage. In addition, if the privacy risk for this type of data access and use is considered acceptable, the same modification of the 15/15 rule may be acceptable for other use cases, until “data blurring” and “data cubing” techniques are implemented.

3. Alternative Views of Parties

SDG&E

SDG&E believes there are two outstanding issues that need to be addressed by the CPUC:

1. The CPUC must determine whether, pursuant to PUC 8380(e)(3), the requirements of AB1103 allow the Utilities to provide the PII required to be provided to building owners under AB1103 without additional customer consent (i.e., constitute a primary purpose).

2. The CPUC must determine whether the utilities may release information to a requestor pursuant to an affidavit signed by the requestor indicating that he/she is the building owner of record and establishing the purpose of the request is for and shall only be used for AB1103 compliance. Proposed language for such an affidavit is provided below:

“By signing below, I represent, warrant and covenant that I am the building owner of record or an authorized agent thereof for the property(ies) I am seeking to benchmark and am duly authorized to make such benchmarking request; I am requesting

benchmarking in accordance with the requirements of California Public Utilities Code 25402.10 and Sections 1680 – 1685 of the California Code of Regulations, as amended (“CCR”); and I am required to disclose such benchmarking results under Section 1682 and 1683 of the CCR. By signing below, I covenant that I shall not disclose any information I receive through the benchmarking process that contains the confidential information of any tenant of any property I have requested benchmarking for, except as specifically required by Section 1683 of the CCR, and I shall indemnify and hold Utility harmless for any disclosure of customer confidential information included in my benchmarking reports to any third party beyond that required by the CCR.

Natural Resources Defense Council (NRDC), Institute for Market Transformation (IMT), California Center for Sustainable Energy (CCSE), and UCLA Center for Sustainable Communities

The Report does not accurately represent the working group discussion on Use Case 7, and we believe it provides the ALJ and the Commission with the wrong framework to consider the very important issues presented in Use Case 7.

Use Case 7 essentially raises the question: What conditions should apply when a utility delivers monthly whole-building aggregated usage information to a building owner to enable the owner to comply with benchmarking obligations and to engage in voluntary energy management?

We believe that utilities can implement procedures to provide monthly whole-building usage information to building owners without compromising the important privacy

interests of customers and without excluding the many building owners with a small number of tenants or a tenant that accounts for a large percentage of total usage. Any risks to the privacy interests of customers can be fully mitigated by setting reasonable conditions on the release of the usage information, such as limiting the usage information to monthly whole-building information and requiring building owners to register with the utility and agree to “terms of use” before receiving monthly whole-building usage data. Such a policy will enable building owners to fulfill the State’s benchmarking requirements and engage in energy management activities while fully protecting customer privacy interests.

The Report mistakenly urges the Commission treat a building owner’s request for aggregated monthly whole-building usage information under the same standards as it would treat a request for the same information coming from a member of the public at large or a third-party researcher. This line of reasoning leads the Draft Report authors to support a Commission policy that would exclude many building owners from receiving the needed information from the utility.

We refer the report authors, the ALJ, and the Commission to our Comment letter filed April 29, 2013, and summarize below the key points:

1. Delivering building usage information to a building owner is very different than delivering usage information to a third-party researcher or members of the public at large. In the vast majority of buildings, a building owner or manager already has

access to the information and any risks to tenants of a nefarious building owner already exist. An owner could access a tenant's utility usage information without requesting the information from the utility, such as by observing the meters located on the premises or installing metering or submetering devices. An owner could also request copies of monthly utility billing information under lease terms that require delivery.

Building owners also are in lease privity with the utility customer; know the identity of the tenant; and, routinely collect highly confidential information from the tenant including credit history, payment account information, insurance information, number of employees, and more. Moreover, any building owner would likely have a right under the terms of any lease to be notified of and approve the use of any equipment that had high electricity or gas requirements.

To the extent tenants face risks related to a building owner knowing the tenant's patterns of monthly utility usage, those risks are present today and are unchanged by a policy allowing a building owner to obtain information from the utility. Any building owner with a nefarious purpose could obtain a tenant's monthly usage information directly, without making a formal, on-the-record request for the information from the utility.

The reason for building owners' to obtain whole-building information from the utility is to enable better energy management and benchmarking – it reduces the time, cost, and difficulty of obtaining whole-building data while increasing data integrity.

2. Monthly usage information is very coarse. To the extent total monthly usage reveals any customer information to the building owner, it is highly likely the building owner would already have access to the information by virtue of its access to the premises, the lease terms, management of the building, and other similar factors.

3. Commission policies should facilitate, not inhibit, compliance with State requirements and goals. The California State legislature has directed building owners to collect usage data at the building level for the express purpose of benchmarking. The State's strong policy interest in giving owners access to whole building usage information has been clearly expressed and explicitly supports summing data to the building level. Access to monthly whole-building usage information is important to enable compliance.

4. Any risks to customer privacy that might be present can be fully mitigated with a registration process. Building owners could be required to complete an online registration with the utility and to agree to terms and conditions prior to receiving any monthly usage information. Thus, the process would be much more akin to the current utility policies for providing usage information with contractors that sign a non-disclosure agreement than it is with delivering information to researchers or the public. Utilities could require building owners to register and agree to only use the usage information for the purpose of benchmarking, energy management, and related uses, and that the owner will not attempt to "de-I.D." the data set to isolate the usage of one particular tenant or share the information with any other parties.

5. Separate policy for residential accounts. One option for the utilities, the ALJ, and the Commission is to establish different procedures for commercial accounts and residential accounts, so that buildings with residential accounts could be subject to additional limitations due to the heightened sensitivities regarding residential usage.

In summary, any policy that excludes building owners with few tenants from receiving whole-building information from the utility merely burdens and impairs building owners from complying with the California benchmarking policy and reduces building owners' ability to implement energy efficiency measures. Any building owner interested in a tenant's monthly patterns of energy usage could already obtain the information in another manner. Providing building owners with access to information from the utility does not impair the customer's privacy interests.

We understand that many large utilities in the U.S. provide building owners with whole-building usage information without excluding buildings with a small number of tenants. A description of the applicable policies in place in other jurisdictions was included in our April 29 Comment and can be found in the U.S. Department of Energy's Energy Efficient Buildings Hub report.^{36/}

We believe the facts and reasoning set-forth above strongly suggest the ALJ and the Commission should implement a policy for building owner access by examining the unique position and circumstances of building owners. A fair examination of the facts

^{36/} See Utilities Guide to Data Access for Buildings Benchmarking, located at: http://s146206.gridserver.com/media/files/IMT_Report_-_Utilities_Guide_-_March_2013.pdf.

will lead to a conclusion that a utility policy to deliver monthly usage information to the building owner could be fashioned to protect the privacy rights of customers without excluding a large number of buildings with a small number of tenants.

Commission policy should provide utilities with regulatory authority, support, and direction to implement procedures that enable building owners, including those with a small number of tenants, or a tenant with a large percentage of total energy use, to obtain monthly whole-building usage information. We urge the ALJ and the Commission to consider revising its policies for building owner access to monthly whole building information with reference to the unique circumstances of the owner/tenant arrangement and the owner's interest in the information, not in the context of use cases in which a utility is asked to share information with the public.

We also encourage the Commission to periodically revisit the policy and the risk mitigating requirements in light of building owners' and tenants' actual experience.

H. Use Case 9: Governmental agencies, such as the Department of Community Services (CSD), seeking access to customer-specific information regarding utility customers who participate in utility weatherization and low income assistance programs

1. Description of Use Case and Benefits of Data Access

Use Case 9 was not included in the ALJ Ruling, but instead was submitted for consideration by the California State Department of Community Services (CSD).

According to CSD, governmental agencies like CSD that implement federally-funded energy efficiency programs for low-income persons such as the Low-Income Home Energy Assistance Program (LIHEAP) and the Department of Energy Weatherization Assistance Program (DOE WAP), need to coordinate the delivery of energy services

with similar services provided by utilities under CARE and Energy Savings Assistance Program (ESAP). This coordination needs to take place through the reciprocal sharing of: 1) historical, non-PII, property-centric weatherization data; 2) historical PII weatherization data; and 3) customer/ client PII, involving eligibility, account information and energy usage data, all shared with the consent of the customer/ client. As a result of this information sharing, similar statewide low income assistance programs administered by CSD and the utilities can better target and reach eligible customers and save on administrative and outreach costs.

2. Evaluation and Potential Recommendations on Use Case 9

Coordination of CSD and utility low income programs is already the subject of the CPUC's pending energy efficiency proceedings, including considering how to maximize the sharing of program information that may improve the efficiency of the respective programs. As such, the CSD's "use case" is being addressed outside of this proceeding. In addition, CSD's use case is not requesting the sharing of customer-specific energy usage data, but instead the sharing of the addresses of current and historical utility customers who have received weatherization assistance from either CSD or the utilities, along with the measures installed, the date of installation, and the funding source utilized. This type of information is also outside the scope of this proceeding, which is solely addressing access to customer energy usage data.

Nonetheless, the efforts by CSD and the utilities to develop information sharing protocols which avoid duplicative or cost-ineffective weatherization services should be supported. To this end, CSD and the utilities are revising their respective customer application forms to ensure that customer data can be shared among the different

agencies prospectively, based on customer consent. In addition, CSD and the utilities are developing a joint customer data base and are considering whether certain categories of historical customer participation data, including addresses of buildings that have been previously weatherized, can be shared without a risk that the identity of the tenant or resident who resides in the building will be disclosed or “re-identified” contrary to the CPUC’s privacy rules or the California Information Practices Act. This mutual effort by CSD and the utilities should be encouraged.

3. Alternative Views of Parties

California Department of Community Services

CSD indicated that it may take issue with the conclusions and recommendations, but has submitted a detailed statement and analysis of Use Case 9 for the record separately in the proceeding.

I. Use Case 10: Energy Commission Access to Customer Specific HVAC Installation Data from Utilities for Title 24 Building Energy Efficiency Compliance

1. Description of Use Case and Benefits of Data Access

Use Case 10 also was not include in the ALJ Ruling but was submitted by the Energy Commission for consideration by the Working Group. As a means of verifying compliance with the Title 24, Part 6 Building Energy Efficiency Standards as they relate to HVAC system efficiency and installation requirements, the Energy Commission’s Compliance and Enforcement Office needs to determine what HVAC systems are being imported into and sold in California for installation within the state. This determination can be made through the tracking of an HVAC’s serial number, whereby any HVAC unit sold in the state will have its serial number entered into a database so that the serial

numbers in this database can be compared to the serial numbers of HVAC units installed under the permitting process in local enforcement agencies throughout the state. This information can also be used for, and should be a requirement of, any HVAC rebate program within the state, whereby a rebate will be issued only for those HVAC installations where the proper permitting by the local enforcement agency has been accomplished. Therefore, the Energy Commission is requesting that the utilities require their customers to provide this data as a condition of receipt of HVAC rebates and utility service.

2. Evaluation and Potential Recommendations on Use Case 10

Unfortunately, Use Case 10 does not involve energy usage data or customer-specific data, and therefore is outside the scope of the Working Group discussions. The Working Group expressed no opinion on the merits of Use Case 10.

3. Alternative Views of Parties

None received as comments on draft Report.

J. Use Case 12: Distributed Generation Providers and Other Commercial Entities Requesting Access to Customer-Specific Energy Usage Data in order to Model All Customers' Electricity Consumption at Sub-Hour Time Intervals

1. Description of Use Case and Benefits of Data Access

The Distributed Energy Consumer Advocates (DECA) submitted and extensively described a use case during the Working Group sessions relating to grid-related energy usage information to support distributed generation. DECA described its use case as providing the public with a working model of the majority of California's electricity grid, with a particular focus on the ability to model all electricity consumers' consumption at sub-hour time interval and to tie that data to actual weather conditions, building data,

etc. The use case allows for the overlaying of wholesale market data including wholesale production run simulations providing prices and emissions. Expected users of this data are policy advocates, distributed generation providers, energy efficiency marketers and evaluators, and local governments.

2. Evaluation and Recommendations on Use Case 12

DECA did not readily identify the specific distributed generation users who would benefit from Use Case 12. In the May working groups DECA provided a proposed mechanism for the public gaining access to the granular data that is the essence of Use Case 12. Specifically DECA presented a “like for like” swapping via randomization of actual sub hour meter data by meter that would be performed by the utility for a requesting party. In DECA’s proposed process a requesting entity would provide to the recipient utility a geographically bounded area for randomization of meter data. The requesting entity would attest that the bounded area contained no uniquely identifiable customers based on anomalous housing stock via a threshold mechanism. DECA proposed a threshold of at least three similarly sized houses within a geography and included easily identifiable electronic signatures such as swimming pools and hot tubs in addition to housing stock/size.

Utilities would only randomize meter/address pairs for a geography once, regardless of the number of requests for the data and would be required to keep a publicly accessible version of that area and the data it contains. DECA proposed that areas contain uniquely identifiable housing stock be aggregated with other geographies until the “like for like” threshold is met. CPUC staff would be responsible for approving

aggregated geographies. Like the homogenous geographies described above these aggregated geographies would only be randomized once to prevent re-querying.

To the extent the “working model” identified by DECA could be scoped specifically and with sufficient detail in a way similar to the research and local government planning projects in Use Cases 1 and 2, with access to similar standardized, pre-formatted aggregated or anonymized energy usage data sets and reimbursement of the costs of providing the data sets, Use Case 12 data access potentially could be provided on the same basis and under the same terms and conditions as data for research and local government projects under Use Cases 1 and 2, provided that the disclosure also complies with the rules restricting the access of “market participants” to customer-specific as well as aggregated energy usage data that could potentially be used by the market participants to manipulate prices or supplies in electricity procurement markets. See General Order 66-C, Public Utilities Code Section 583 and D.06-06-066 “Confidentiality Matrix” Rules.

3. Alternative Views of Parties

None received as comments on draft Report.

VII. IMPLEMENTATION PROTOCOLS

A. Utility “Strawperson” Process, As Modified in Response to Comments

On May 8, the utilities participating in this proceeding jointly submitted to the Working Group a “strawperson” proposal for streamlining and improving the data access process. The “strawperson” proposal is described below, with some additional details.

1. Each utility will establish a consistent, streamlined, “one-stop” process for providing authorized third-parties with energy usage data access where permitted by law and Commission privacy and ratemaking rules. The process will include the following:

a. Single point-of-contact in the utility for filing and processing of third-party energy usage data requests. The single point-of-contact will include a single email mailbox or website and other contact information to which requests for energy usage data access may be transmitted.

b. The single point-of-contact information will be provided prominently and conveniently on the utility’s website.

c. The utility’s website will provide access to an electronic input form for third-parties to request energy usage data access, comparable to the “template” provided in the Phase 3 ALJ ruling (Attachment A to ALJ Sullivan’s ruling of 2/27/13). The form will be consistent among PG&E, SCE, SDG&E and SoCalGas.

2. The utility website is expected to eventually include a “catalogue” of standard energy usage data access reports, in the most commonly requested formats among PG&E, SCE, SDG&E and SoCalGas, that can be made available to third parties at a cost-based fee. Such standard reports will be made available to third parties within e.g., 7- 10 business days of receiving a completed request form if all privacy, security and contractual controls are in place and subject to a reasonable volume of requests being processed at the same time.

3. Within e.g., 7- 10 business days of receiving a form from a third-party requesting energy usage data access, the utility will respond by phone, email or in

writing regarding whether the information on the form is complete and, if incomplete, what additional information is required for the utility to process the request.

4. Within e.g., 30 business days of receiving a complete request for energy usage data access from a third-party, the utility will respond by email or in writing regarding whether it is able to grant the request and with a proposed schedule and estimated cost for compiling and providing access to the data. If the utility responds that it cannot grant access to the data, it will provide specific reasons for why it is not providing the data or other options for providing data access (such as providing data access using a pre-approved report from the data access “catalogue” or suggested modifications to the request such that it could be granted). If the third-party disagrees with the utility’s rejection of its request for data access or the alternative options offered by the utility, the third-party may bring the dispute for informal discussion before the Energy Usage Data Access Advisory Committee established below.

5. Prior to receiving access to energy usage data, a third-party will execute a standard confidentiality agreement if required by the utility, with substantially consistent terms and conditions among PG&E, SCE, SDG&E and SoCalGas. In addition, if a pre-disclosure review of the third-party’s information security and privacy controls and protections is required by the utility, the requirement and criteria for the review will be substantially consistent among PG&E, SCE, SDG&E and SoCalGas and published in advance and available on the utilities’ websites.

6. An Energy Usage Data Access Advisory Committee should be considered, modeled on the Procurement Review Group established under the utilities’ Long Term Procurement Plans. The Advisory Committee will consist of representatives

from each of the utilities, the Commission's Energy Division, the Division of Ratepayer Advocates, representatives of consumer and privacy advocacy groups, and other interested parties. The Advisory Committee will meet at least once a quarter to review and advise on the implementation of the utilities' energy usage data access programs, and to consider informally any disputes regarding energy usage data access and make other informal advisory recommendations regarding technical and policy issues related to energy usage data access.

7. Nothing in this process requires or authorizes a utility or a third-party to violate any existing privacy or information security laws, rules or orders, including the Commission's privacy rules and the California Information Practices Act. Nothing in this process requires or authorizes a utility or a third-party to transfer, sell, or license energy usage data that consists of the utilities' intellectual property, trade secrets, or competitively-sensitive data. The transfer, sale or licensing of such intellectual property, trade secrets and competitively-sensitive data will be subject to Commission review and approval consistent with existing Commission rules and orders regarding the sale, transfer or licensing of utility assets.

8. All data outputs will be in standard formats. Data will be accessible in specified formats such as comma-delimited, XML, or other agreed-upon formats. Customized outputs or formats should be avoided or subject to higher cost fees. The Advisory Committee can review formats annually to ensure that the utilities are consistent with current technology trends for data sharing formats.

9. Mechanisms for handling data delivery for requests of all sizes in a secure manner should be standardized. Some requests are very small and require very little

effort to transmit or deliver. Others can be gigabytes in size. In addition, sensitive customer information or other confidential information must be transmitted to the third party with reasonable encryption, rather than e-mailed. By standardizing delivery mechanisms, utilities and third parties will provide pre-approved delivery methods for sensitive information, reducing risk as well as the time to transmit and receive the data.

The other interested parties in the Working Group generally supported the utilities' proposal, with some recommended clarifications and enhancements. For example, LGSEC and CCSC disagreed with the requirement that third-parties accessing customer-specific energy usage data undergo an information security review by the utility to ensure that the third-parties privacy protocols and controls are adequate. LGSCE and CCSC also requested that the processing protocols, data formats and deadlines be consistent across the utilities. DRA requested more standardization and transparency on the processing of data access requests and the formatting and transmittal of data to recipients.

B. Model Non-Disclosure Agreement

Attachment A to the February 27, 2013, ALJ Ruling included a model non-disclosure and information security agreement submitted by PG&E for consideration in the proceeding. The Working Group did not discuss the model agreement in detail during the Working Group Sessions. However, SCE raised questions about the applicability of a standard NDA absent Commission-ordered disclosure of PII. According to SCE, the Commission stated in D.11-07-056 that there is a difference between "third parties who receive data via a free interaction with the utility for a contractual purpose and those who receive data via the direction and under the supervision of the Commission or via the authorization of the customer." D.11-07-056,

p. 80. This is because “[r]esponsibility follows free contractual relationships, but responsibilities are different when data is disclosed to a third party pursuant to Commission direction or a tariff.” *Id.*

SCE viewed PG&E’s NDA as being appropriate for a vendor relationship between PG&E and parties with whom it contracts for primary utility purposes. Because this proceeding is not focused on utility-specific vendor relationships for primary utility purposes, in SCE’s view, the only context in which a standardized NDA is appropriate is one in which the Commission orders the utility to disclose data without customer authorization. To that end, SCE will prepare a revised NDA for submission in connection with its opening comments on this Report.

The Working Group expresses no opinion at this time on the details of the model agreement or SCE’s alternative to the model agreement.

C. Alternative Views of Parties

None received as comments on draft Report, other than SCE as discussed above.

VIII. IMPLEMENTATION COSTS AND COST RECOVERY

A. Potential Recommendations

The ALJ Ruling required the Working Group to assess the costs to ratepayers of implementing energy usage data access under the use cases, including costs associated with setting up and implementing a common data access process and maintaining data security protocols.

The utilities’ reasonable and incremental costs of implementing energy usage data access under the use cases should be reimbursed, either through direct reasonable fees on data access users, or through recovery from ratepayers generally,

and subject to appropriate Commission approval in a ratesetting proceeding. However, because an essential element of energy usage data access will include development and implementation of common database templates as well as new technical methods to “blur” or “process” data to avoid “re-identification,” the utilities are unable to estimate the precise costs of implementing energy usage data access at this time. Instead, the CPUC may authorize the utilities to recover their reasonable costs of implementation, subject to approval in an appropriate rate-setting proceeding.

The Commission should address all cost recovery issues before requiring the utilities to implement any new data sharing requirements in this proceeding, including not only the privacy protocols for protecting customer-specific information from re-identification, but also the administrative requirements for processing and fulfilling energy usage data access requests from third-parties.

B. Alternative Views of Parties

TURN

TURN states that there is no consensus yet on cost recovery, and that the issue of costs and methods for cost recovery were not specifically discussed in the Working Group sessions, but instead should be considered in a ratesetting phase or separate proceeding.

IX. CONCLUSION – POLICY RECOMMENDATIONS – ISSUES NEEDING RESOLUTION – NEXT STEPS

A. Potential Recommendations

Based on the findings and potential recommendations discussed above, the following next steps should take place in this proceeding, after comments on the Working Group Report:

1. The CPUC should continue this phase of the proceeding to consider three issues not resolved to date: (a) specific proposals by the utilities and interested parties to implement acceptable energy usage data “blurring” or “processing” techniques to mitigate the risk of “re-identification” of customer-specific information, which will consider the recommendations of EFF and its experts in this proceeding.; and (b) specific data access protocols based on the IOUs’ “strawperson streamlined process” (i.e., the catalogue of use case formats); (c) the appropriate scope and contents of an NDA, given the disparate positions of the parties
2. An update to this Workshop Report should be filed at the conclusion of the additional workshops identified above, with an opportunity for comment; and
3. The Commission should issue a decision adopting the consensus recommendations and resolving contested issues and approve utility applications for

recovery of reasonable, incremental costs of implementing energy usage data access, including the costs of implementing new techniques for anonymizing energy usage data to prevent re-identification.

B. Alternative Views of Parties

None received as comments on draft Report except TURN's comments on cost recovery, discussed in Section VIII, above.

APPENDIX A

Appendix A- Use Case Descriptions Provided Pursuant to February 27, 2013 ALJ Ruling

The following are detailed descriptions of the various “use cases” discussed and evaluated by the Working Group, generally in the format requested by Attachment B of the February 27, 2013, ALJ Ruling. The extent practicable, the descriptions are similar to the descriptions of the use cases in the ALJ Ruling and to the descriptions provided by sponsoring parties for those similar or additional use cases submitted as part of the Working Group discussions.

Use Case 1 – Provided by LGSEC

1. Overview

1.1 Use Case Summary

Use Case 1: Local Governments seeking access to aggregate data for use in creating legislatively required Climate Action Plans and implementation of energy efficiency programs.

Consistent with AB 32 and the CPUC Long Term Energy Efficiency Strategic Plan, many local governments are adopting electricity, natural gas and greenhouse gas (GHG) reduction goals and action plans, and assuming a lead role in offering programs and policies aimed at achieving these goals in their communities and regions. Achieving energy and GHG reduction goals requires access to data that enables local governments to effectively evaluate and report progress toward adopted goals and to evaluate the efficacy of specific programs and policies. Proper evaluation of programs and policies informs resource allocation moving forward and ensures the highest and best use of ratepayer and other public funds in the implementation of energy programs by local governments.

Specifically, local governments need access to three categories of energy usage data, on a monthly basis:

1. Aggregated data that illustrate the status of progress toward adopted energy and GHG reduction goals, e.g., total monthly residential energy use at the block group level
2. Aggregated data that illustrate the outcomes of a given energy program, e.g., total monthly electricity savings from the Energy Upgrade CA program at the community or sub-community level
3. Granular, anonymized data at the address level, on a monthly usage basis, that provide insight into how energy use changes as properties participate in programs, and identify unmet needs in order to plan for future programs

The data needs expressed in this use case are expressed under the assumption that the “15/15 rule” does not limit local governments’ access to building level data if the data is provided in a way that protects individuals’ confidentiality.

1.2 Objectives and Ratepayer Value

Local governments’ ultimate objective is to effectively reduce energy consumption and the associated costs and GHG emissions in their communities. Achieving this objective saves residents, businesses, municipal governments, and utilities money. This objective can only be achieved with the provision of various forms of utility data provided to local governments in a timely, user-friendly, and consistent manner. Local and regional government entities need these data in order to meet state legislative requirements, comply with local/regional policies and ordinances, and implement programs mandated by the CPUC and paid for by ratepayers.

Local governments need aggregated energy usage data to achieve the following:

- Provide timely and consistent reporting on energy use and greenhouse gas trends

to locally elected government bodies and community stakeholders

- Evaluate the efficacy and impact of energy policies and programs operating in their communities
- Identify energy program participation rates and areas within a community that are potentially underserved by a given program or programs

Local governments need granular data on a monthly basis to achieve the following:

- Evaluate meter and building-level energy consumption pre- and post- energy retrofit
- Correlate energy usage to other relevant characteristics (e.g. geography, building characteristics, and customer financial characteristics). This enables not only improved targeting of future programs, but more effective messaging to increase participation and effectiveness
- Conduct building benchmarking analyses
- Identify unmet needs to plan for future programs that will ensure the highest and best use of ratepayer and other public funds.

Granular data would need to be protected from entering the public sphere through appropriate privacy protocols. There may be instances where local governments need data on a more frequent basis than monthly in order to account for weather.

1.3 Actors

<i>Name</i>	<i>Role description</i>
Utility	Data owner – SCE, PG&E, SDG&E, SoCalGas
Local and regional government	Data requestor
CPUC	Regulator – Provides rules to be applied consistently across investor-owned utilities, ensure market participants can access data as deemed appropriate.
Academic institutions	Data requestor – Assist local governments in conducting studies and preparing documents, such as Climate Action Plans
3 rd party	Data requestor- 3 rd party (e.g., non-profits) working on behalf of local government or facilitating local government partnerships

1.4 Regulatory Proceedings and Rules that Currently Apply

<i>Agency</i>	<i>Description</i>	<i>Applies to</i>
CA Air Resources Board	AB 32 Climate Change Scoping Plan encourages local governments to adopt GHG reduction targets consistent with AB 32	Local governments

<i>Agency</i>	<i>Description</i>	<i>Applies to</i>
CPUC	Long Term Energy Efficiency Strategic Plan sets targets for local governments related to reducing energy in government facilities and adopting energy reduction plans and tracking achievements	Local governments
CPUC	Decision 09-09-047. The City of Irvine was awarded \$200,000 to develop a pilot GIS mapping tool with Southern California Edison (SCE).	City of Irvine, SCE
City and county governments	Locally adopted energy and GHG reduction plans require local governments to monitor and report energy and GHG trends to elected bodies and community regularly	Local governments
CA Air Resources Board	SB 375 establishes requirements for regional reductions of GHG emissions	Local governments
Attorney General regional plan mandates	Require local governments to must address a “projects” contribution to climate change. This can be done through a climate action plan, a General Plan, or through case by case analysis of development proposals (among other mechanisms)	Local governments
AB 1103	Compels the disclosure of monthly energy usage data aggregated to the level of the whole building	Utilities

2. Use Case Details

2.1 Current State Narrative

Currently, energy usage data are provided inconsistently within a utility service territory and between utility territories. Except for the highest level of data aggregation (e.g., total residential energy use), data are not provided in a format that allows local governments to manipulate them for the purposes of evaluation and analysis. For example, utilities regularly provide data in a PDF format.

In general, the IOUs do provide local governments with aggregated utility data at the “sector” level, e.g., total electricity and natural gas consumption at the residential and commercial/industrial sector levels. Local governments use these data to measure high-level energy use and GHG trends within their communities. While useful for this purpose, these data do not allow for any additional, more granular analysis related to evaluating the efficacy of specific policies and programs, let alone the efficacy of a given energy upgrade project at the building level.

In short, the data that the IOUs currently make generally (though not universally) available to local governments allows for basic, big picture reporting of trends, but not for the types of analysis and evaluation needed to actually plan and continuously improve local and regional energy efficiency efforts. Lack of access to data is a significant hindrance to local government energy efficiency efforts.

To further illustrate the current state of data provision to local governments, we provide some specific examples below.

PG&E generally provides annual community-level energy data starting with 2005. It is aggregated into residential electricity and natural gas consumption and commercial/industrial electricity and natural gas consumption. These “sector” data enable only high-level tracking of community energy use. These data do not include a single-family vs. multifamily breakout.

PG&E also provides data on aggregated average monthly energy use and number of service accounts by sector (e.g., commercial and residential). PG&E also provides aggregated KWh and therm savings by end use categories, e.g., KWh savings from lighting or therm savings from boiler and steam systems.

For non-residential energy users, PG&E also provides an illustration of the scale of consumption by market segment (e.g., hospitals, offices, biotech, etc.) and by zip code. PG&E does not provide actual energy usage by market segment, but it does provide a visual that enables one to compare the scale of use by market segment and zip code annually.

PG&E does not provide monthly or annual energy use data at a scale that is more granular than total residential and total commercial/industrial. Local governments cannot access industrial data because of PG&E’s interpretation of the 15/15 rule. Presumably because PG&E’s data counts meters, not buildings, its data are not reliable in terms of the number or types of buildings being counted, nor in most cases can the data distinguish between multi-family residential buildings and non-residential commercial buildings. Local governments cannot access data on a zip code or more granular level based on building or occupant characteristics in order to better allocate energy efficiency services and assistance. Local governments (and other service providers) also find it very difficult to get meaningful data on the outcomes of energy saving programs in their communities, such as Energy Upgrade California. Local governments also do not have access to building or meter-level data that would enable proper evaluation of energy efficiency programs or the ability to target market segments showing the greatest need and receptivity. Even when data requests are made *with customer consent*, PG&E lacks the both the technical and organizational systems to provide the data for the needed time periods.

SCE provides two free data requests a year. Beyond that, local governments have to pay for each subsequent request. Depending on the type of data requested, a local government may or may not receive this information. If the information is provided, it may or may not come in a useful format that can be analyzed/compared/calculated/totaled, input into other systems, etc.

2.2 Future State Narrative

Utility data would be provided to local governments in a timely manner, in a consistent, user-friendly format, and following consistent protocols within an IOU’s service territory and between the IOU service territories. Local governments would have access not only to aggregated energy consumption data by sector, e.g., residential and commercial/industrial,

but also to more granular data that local governments need in order to effectively evaluate and continuously improve energy efficiency projects and programs. Appropriate privacy safeguards would be put into place in order to protect customer confidentiality.

Utilities would provide a web-based system that can be queried to obtain the necessary data and to compare and analyze a series of pre-set variables. The system would produce instant reports in a useable/downloadable format. This updated information/ability to query should be available at any time to any local/regional government entity, not be cost prohibitive, and provide equal access to all local regional government entities.

In addition, the work directed as part of the CPUC Decision 09-09-047, to develop a pilot GIS mapping tool with City of Irvine and Southern California Edison (SCE), would continue and expand. The purpose of the mapping tool is to provide SCE customers access to unique maps, tables, and statistics of community utility data which in turn would assist in measuring the outcome or effectiveness of marketing energy efficiency programs. The maps are created based on energy usage, demographics (via census data) and land use (via assessor data.) No individual account information is displayed and instead, the data are aggregated to ensure a customer's confidentiality. In many cases, mapping of data is both the clearest and most persuasive way to distill and communicate large, complex sets of data. The full summary of this pilot project is listed under section 4.3 Additional Comments.

3. High Level Requirements

3.1 Data and Aggregation Requirements

Local governments need access to the following data:

- Aggregated, monthly KWh and therm consumption by residential, commercial and industrial tariffs
- Aggregated, monthly KWh and therm consumption by single-family residential vs. multifamily residential
- Aggregated, monthly KWh and therm consumption by industry sector (NAICS code)
- Aggregated utility program participation data, including unit savings and incentive values organized by market segment, types of end uses/technologies addressed, and by program (e.g., Energy Upgrade CA, Smart Lights, etc.)
- Disaggregated building and meter-level data that enables:
 - Measuring the impact of specific energy upgrades
 - Conducting building benchmarking analyses
 - Analyzing and identifying unmet needs to plan for future programs

These data should be made available for all cities, counties, and unincorporated areas. The data should be available at the following sub-community levels:

- Zip Code and zip+4
- Census tract
- City and county limits

There may also be instances where local governments require data that addresses energy within local government land use/zoning areas, or in more disaggregated forms (e.g., at the address level) to enable more detailed planning and evaluation by local government

Primary fuels used within a community (e.g., at a co-gen facility) for electricity generation should be netted out.

<i>Data Type</i>	<i>Priority (H/M/L)</i>	<i>Aggregated/Anonymized/Identifiable</i>	<i>Description/Additional Comments</i>
Aggregated by residential (including single family vs multi-family), commercial, and industrial	H	Aggregated/Anonymized	
Aggregated by industry sector (NAICS code) (General Service Commercial & Industrial Service Accounts)	H	Aggregated/Anonymized	
Aggregated utility program participation data by market segment, end use, and program	H	Aggregated/Anonymized	
Disaggregated building and meter-level data	H	Disaggregated/Anonymized	
Provide for each incorporated city and county unincorporated areas	H	Aggregated/Anonymized	
Provide for each Zip Code and Zip Code+4	H	Aggregated/Anonymized	

3.2 Functional Requirements

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
Provide data in an electronic format that enables manipulation and analysis (e.g., Excel files where data can be graphed in different ways; not pdfs)	H	

3.3 Policy & Other Requirements

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
CPUC should establish policy that various forms of aggregated data described in this use case may be made publicly available (e.g., as local	H	

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
governments report to their communities on overall progress made in reducing energy use)		
Separate policy consideration should be given to the disaggregated data described in this use case, which local governments need for effective planning and evaluation activities, but which they may be required not to publish or make publicly available.	H	

4. Barriers and Open Issues

4.1 Barriers

<i>Barrier Description</i>	<i>Priority (H/M/L)</i>	<i>Current/Anticipated</i>
Consistency between the different investor-owned utilities	H	Current
Consistency within a utility service territory	H	Current
Differing legal opinions of what information can be made available to local governments based on interpretation of customer confidentiality rules	H	Current
Cost charged to local/regional government entities for providing utility data	M	Current
Data cannot be manipulated for analysis purposes	H	Current
Timeliness of providing information to local/regional government entities	H	Current
Paperwork/NDA and process for requesting consumption and participation data	H	Current
Inability of local governments to verify the integrity of data provided	H	Current

4.2 Outstanding Issues

<i>Description</i>	<i>Proposed Next Step, if any</i>
How will variables for subsets of data be initially defined?	
How will variables for subsets of data be changed and remain flexible over time as lessons are learned and data needs change?	
How will Irvine/SCE's pilot project be incorporated into future decisions? Will the work/ratepayer funds put in to date be lost/shelved?	

<i>Description</i>	<i>Proposed Next Step, if any</i>
How will data be made available to local governments while a long-term solution/tool is being developed?	

4.3 Additional Comments

As part of the CPUC Decision 09-09-047, the City of Irvine was awarded \$200,000 to develop a pilot GIS mapping tool with Southern California Edison (SCE). The purpose of the mapping tool is to provide SCE customers access to unique maps, tables, and statistics of community utility data which in turn would assist in measuring the outcome or effectiveness of marketing energy efficiency programs. The maps are created based on energy usage, demographics (via census data) and land use (via assessor data.) No individual account information is displayed and instead, the data is aggregated to ensure a customer’s confidentiality.

The mapping tool will also be valuable in assessing a customer’s compliance with regulatory mandates such as AB 32, SB 375, AB 1103, and the California Energy Efficiency Long-term Strategic Plan by allowing the customer to not only track its energy consumption, but to also conduct an analysis of the data to see trends or patterns as well as identifying hot spot areas.

Furthermore, the mapping tool will permit customers such as municipalities to evaluate their performance and/or compliance with city-specific policies. For example, the City of Irvine has the following policies:

- Energy Plan
- General Plan with an energy component
- Green Building Ordinance based on California Green Building Standards

The City’s Energy Plan contains the following metrics to evaluate its energy efficiency performance:

- Track residential and commercial energy use
- Effectiveness of public education
- Buildings’ energy efficiency performance
- City fleet energy consumption

5. Conclusion

5.1 Conclusion

Data needs to be made available a timely manner:

- In a useable format
- Able to be sorted and reorganized
- Accurate
- Current
- Flexible in terms of the parameters and variables by which it is both provided and sorted. Needs will change both programmatically and over time as we learn that new subsets of data become useful

- Granular as needed to the building level, stripped of Personally Identifiable Information or provided with proper privacy protections, to measure the impact of specific energy upgrades, conduct building benchmarking analyses, perform analysis to identify unmet needs to plan for future programs, and enable “real world” examples to inform lessons from large scale aggregate data

5.2 Recommended Next Steps

Establish an advisory group made up of local governments and other stakeholders to define an agreed upon process, protocol, and timeline for providing the necessary data to local governments.

6. Appendix

6.1 Contact

Jody London, Regulatory Consultant to the LGSEC
510/459-0667
jody_london_consulting@earthlink.net

Meredith Reynolds, Environmental Programs Administrator
949.724.6684
mreynolds@cityofirvine.org

Timothy Burroughs, City of Berkeley
TBurroughs@cityofberkeley.info

6.2 Reference Materials – available upon request from Meredith Reynolds

City of Irvine Pilot GIS Presentation
SCE Pilot GIS Presentation

Use Case 1 – Provided by California Energy Commission

This use case establishes a framework for streamlined transfer of energy use and associated PII data from the utilities to policy making entities that clearly spells out the criteria for deciding (1) whether and (2) how, to release “electrical or gas consumption data” to policy making bodies. This clarity will release the Utilities from the burden of deciding whether to release data and the related legal liability by spelling out explicitly (1) the legal basis for this release (2) the methods that will be used for protecting confidentiality, and (3) the legal basis for the Utilities’ indemnification.

1. Overview

1.1 Use Case Summary

Federal and state agencies, and local governments are tasked with formulating policies to reach energy efficiency and greenhouse gas emissions goals without having a rich set of energy use data to base their policies on. For example, knowing the average consumption of a type of building is important, but knowing the median and standard deviation of energy use per square foot of small retail buildings built between 1970 and 1980 in the central valley is much more useful. This use case sets up the parameters by which governmental agencies can be assured access to both energy use data and the PII associated with it in a way that indemnifies the utilities supplying the data

1.2 Objectives

Allow access to energy use data, including PII for the purposes of formulating public policy. This provides value for the ratepayers by allowing for policies that are better suited to well-substantiated market conditions. The use of real world data to guide policy decisions will augment the value provided to ratepayers of such policies.

1.3 Actors

Name	Role description
Federal Agency (DOE)	Maintains the secure data repository Standard Energy Efficiency Data Platform (SEED), taxonomy, and data access protocols. Provides an Application Programming Interface (API) framework that allows for aggregated access to data by non-agency entities and granular data by approved agencies. Maintains a record of data access and usage.
Policy Setting Agency (State or Local)	Accesses the data in a controlled fashion under a Non Disclosure Agreement (NDA) to formulate policy for energy efficiency, demand response, or energy management.
Utilities	Provides data to the repository, maintains NDA records

1.4 Applicable Statutes and Regulatory Rules

Agency	Description	Applies To	Purpose
Public Utilities Commission	Public Utility Code (PUC) Section 8380	Electrical or Gas Corporations, Contractors, third party implementers of EE and IDSM programs, State or Federal agencies, Customers	Outlines permissible use for data collected via an advanced metering infrastructure (including PII)
Public Utilities Commission	Public Utility Code Section 394.4(a)	governing body of a public agency or electrical service providers for residential and small commercial customers, customers	Establishes minimum standards for parties providing electricity to residential and small commercial customers

2. Use Case Details

2.1 Current Data Practices

Current practice is that policy-setting agencies have no access to data. They can request data and if the aggregation is large enough (i.e., entire cities' Climate Action Plan), the utilities will often provide highly aggregated data under an NDA. The level of aggregation that is appropriate for compliance with PUC Sections 8380 and 394.4(a), however, make it impossible to answer basic questions about the distribution of buildings sizes and energy use within particular climate zones or areas of construction.

2.1 Requested Data Practices

The requested data practice is that utilities release all data that is descriptive of the building, its energy use, and the efforts that have been made in outreach to improve its energy performance. This data should be released into the requesting agency's SEED database under an NDA that clearly states the - parameters that are required for any release of aggregated data.

The SEED database will be detailed, granular, and secure. This will make it useful for answering policy agencies' questions on the likely effects on the market of a new or changed policy. This level of granular data will not be accessible to anyone outside of the agency, and the access to the data will be logged and stored. Any data that will be made publicly available will be both aggregated and anonymized.

There have been repeated references to breaches of confidentiality of data in non-energy industries. What those breaches have in common is that they require a secondary data set to tie anonymized data to specific users. In order for this to be possible, individual users must be characterized precisely enough for an overlapping data set to be used to identify users. This problem is prevented by giving proxy data in place of precise data, defining static geographic boundaries, and ensuring a minimum sample size for any query. This approach was outlined in Aram Shumavon's presentation on January 16, 2013. An illustrative example of the data proxy approach is given below.

Data Field	Raw Data	Data Proxy
Building Identifier	CEC Building	Building Number 00572
Street Address	1516 9 th street, 95814	9581X
Demand*	245 kW	125-250 kW
15 minute Energy Use	67 kWh	50-75 kWh
Electrical Energy Use Intensity	12.08 kWh/ft ² /year	10-15 kWh/ft ² /year
Year Constructed*	1980	1980-1990
Construction type*	Concrete	Concrete
Building Use	State Government Office - Large	Large office

*Only disclosed if there are enough similar buildings in the data set – the API will regulate the conditions of data release.

3. High Level Requirements

3.1 Data Granularity Requirements and Data Use

Data Type	Priority (H/M/L)	Aggregated/ Anonymized/Identifiable	Description/Additional Comments
Demand	H	Identifiable	
Service address	H	Identifiable	This is essential to allow for cross referencing of the building with secondary data sets within the secure environment
15 minute energy consumption	H	Identifiable	Allows multivariate market analysis to determine which programs and factors have lead to success in which customer segments, and what the energy effects have been
EE Program History	M	Identifiable	
Customer Industry	M	Identifiable	
Rate structure	M	Identifiable	Allows for segregation of electric only, agricultural, and other specialized

The above is a small subset of the data that would be useful in formulating policy. The Commission should direct the utilities to make a list of the data fields they maintain available for this proceeding.

3.2 Data Collection and Maintenance Requirements

Requirement	Priority (H/M/L)	Additional Comments
Usage/demand data	H	Updated according to billing cycle
15 minute consumption data	M	This data is useful for identifying temporal use patterns in specific segments. It will never be released on a disaggregated basis.
Other Data Fields	M	As indicated by data type

3.3 Required Policy & Other Determinations

Requirement	Priority (H/M/L)	Additional Comments
Reporting potential confidentiality breaches to PUC	H	
Reporting potential confidentiality breaches to utility of origin	H	
Timely review of process for releasing aggregated data to public	H	This must have a hard deadline for response and resolution of issues in order to avoid undue delays in data release

4. Current Data Obstacles and Other Issues

4.1 Barriers

Barrier Description	Priority (H/M/L)	Current/Anticipated
Criteria for aggregation	H	Current
Criteria for anonymization	H	Current
Data transfer IT protocols	M	Anticipated

4.2 Outstanding Issues

There is no documentation of which data fields are being maintained and populated by which utilities. Not knowing what data exist hampers our ability to establish appropriate parameters for a use case utilizing this data

Description	Proposed Next Step, if any
Data fields unknown	Utilities to submit a list of data fields that have been and are currently maintained on customers

4.3 Additional Comments

None

5. Conclusion

5.1 Conclusion

In order to formulate the most effective policy, agencies should have access to the most comprehensive data available. This use case ensures that agencies will be well prepared to formulate effective policy without compromising the confidentiality of customer data.

5.2 Recommended Next Steps

Utilities should make a list of data fields available. Agency stakeholders, CPUC, and CEC should work together to reach consensus on what constitute reasonable levels of aggregation and anonymization to protect customer confidentiality. The model NDA should be amended to reflect a use case where no consideration is being provided, as there is no individual or corporate entity that will benefit monetarily. CPUC should issue an order as described in PUC 8380 (e)(3) directing the utilities to upload granular data to policy setting agencies in the manner described above. CPUC and stakeholders should come to consensus on what constitutes a reasonable threshold for aggregation and anonymization. Details of how the data transfer to SEED may be affected should be documented for each of the utilities' billing and customer databases.

Appendix

Contact

*Christine Awtrey
California Energy Commission
(916) 651-1227*

Reference Materials

See Public Utilities Code below

Use Case 2 – Provided by UCLA Center for Sustainable Communities

Here is a draft of use case 2 from the perspective of the California Center for Sustainable Communities at UCLA:

1.2: Objectives:

CCSC seeks energy data to identify current patterns and drivers of electricity consumption, to target and evaluate energy efficiency investments, and to help the State of California achieve its energy and environmental policy objectives.

This research is intended to provide significant public benefits. To date, there is little baseline knowledge of the patterns and drivers of electricity consumption, meaning decision-makers are flying blind as they try to implement policies such as AB 32 and SB 375. Only with such baseline understanding of consumption patterns can we begin to effectively reduce consumption through targeted investments. Our research provides significant public benefits by helping minimize the implementation costs and unintended consequences of achieving such policies.

3.1: Data granularity requirements and data use:

CCSC's analysis requires monthly electricity consumption data at the individual customer account level, with each account identified by customer class (e.g. single-family residential, multi-family residential, commercial, industrial, municipal operations, etc.), for a period of at least 7 years. Commercial and industrial accounts should also be identified by NAICS code.

We note that data needs will vary immensely by research project across all data parameters, including temporal resolution (e.g. annual, monthly, interval), geographic resolution (e.g. ZIP, ZIP+4, census block, individual account), and whether identification by tariff or customer class is required. For this reason, a flexible approach to data provision serving public interest benefits should be pursued.

3.2: Data collection and maintenance requirements:

Data must include consumption information by account for at least the past 7 years, updated on an ongoing basis. This is necessary for developing an understanding of trends in energy consumption over time. Data must be made available in an accessible electronic format (e.g. Excel, Access, comma- or tab-delimited file, etc.).

Regards,
Sinnott

Sinnott Murphy
Researcher
Institute of the Environment and Sustainability, UCLA
smurphy@ioes.ucla.edu
(310) 825-3778

Use Case 3 – No Additional Proposal Provided

Use Case 4 – Provided by California Energy Commission

ATTACHMENT B State-Owned Building Data Use Case

1. Overview

1.1 Use Case Summary

The State is following Executive Order B 18-12 to drastically improve the energy efficiency of its 8,000+ buildings.

Under AB758, the Energy Commission is recommending that public buildings “lead by example”; for example, benchmarking and disclosing the energy use of its buildings and then educating and marketing this effort to local governments, private businesses, and homes with the goal that voluntary benchmarking and disclosure programs improve over time.

1.2 Objectives

The objectives are two-fold:

- Meet the targets identified in EO B 18-12
- Serve as a “lead by example” model to save energy by improving the energy efficiency of local governments, private businesses, and homes

Value to ratepayers is provided by improving access to data and allowing entrepreneurial businesses to mine the data for business opportunities that could help improve energy efficiency of California, and create jobs.

1.3 Actors

Name	Role description
State	State would download energy information for data benchmarking and disclosure; energy project analysis; energy monitoring and verification

1.4 Applicable Statutes and Regulatory Rules

Agency	Description	Applies to
Governor’s Office	EO B 18-12	State
CPUC/ CEC	AB758	Existing Buildings

2. Use Case Details

2.1 Current Data Practices

For benchmarking and disclosure: 33 separate state departments call their utility account managers and ask for monthly data to be uploaded to Energy Star Portfolio Manager; the data is then corrected as necessary and posted in an aggregated format at green.ca.gov.

For energy efficiency improvement and monitoring and verification: This is currently a totally bottoms-up process where individual departments, or their hired consultants, request various types of hourly, monthly data, or bills in order to evaluate the potential of improving efficiency, or to install solar PV, or to use as a check on the installed savings of projects.

2.1 Requested Data Practices

A central data repository including all electric, gas and water usage data from the IOU's and the POU's in the State in the form of billing data, monthly data, 15 minute data, rebate and permit data, and perhaps more granular data (e.g., 1 minute data, 6 second data) for a minimum of 7 years. Data should be filterable in many ways.

The goal is to allow access to state owned-building data to any interested party at some time in the future. Until then,

- State data should be accessible through Freedom of Information Act.
- State data should be accessible to State employees through login and password.

3. High Level Requirements

3.1 Data Granularity Requirements and Data Use

Data Type	Priority (H/M/L)	Aggregated/ Anonymized/Identifiable	Description/Additional Comments
Billing	M	Identifiable	Billing questions
Rebate and Permits	M	Identifiable	Track rebates and permits to accounts
Monthly	M	Identifiable	Benchmarking
15 minute	M	Identifiable	Energy analysis for EO B 18-12
Audited Energy Savings	L	Identifiable	See Barriers
Estimated Installed Energy Savings	L	Identifiable	See Barriers
Verified Energy Savings	L	Identifiable	See Barriers

3.2 Data Collection and Maintenance Requirements

Requirement	Priority (H/M/L)	Additional Comments
Billing	H	Timely bill payment
Rebate and Permits	M	Upload quarterly or faster
Monthly	M	Upload quarterly or faster
15 minute	M	Upload quarterly or faster
Audited Energy Savings	L	Upload quarterly or faster
Estimated Installed Energy Savings	L	Upload quarterly or faster
Verified Energy Savings	L	Upload quarterly or faster

Note: Save all data for at least 7 years.

3.3 Required Policy & Other Determinations

Requirement	Priority (H/M/L)	Additional Comments
Develop a single database that includes audited, installed, and verified energy savings	H	See Barriers
Tie rebates and permits closer together	H	

4. Current Data Obstacles and Other Issues

4.1 Barriers

Database: The IOU's (and POU's) perform hundreds of energy audits each year, rebate hundreds of millions of dollars in energy savings projects and verify the savings on some projects, yet all of this data is collected in hundreds of disparate spreadsheets, brand x databases, and brand y databases. In addition, the data that is collected is not standardized and is very difficult to combine upstream.

As a start, the IOU's should be required to develop a standardized open source, energy project database- it could certainly include supplementary fields that one IOU finds useful, but others do not. Over time POU's could opt-into using the database and perhaps Energy Upgrade California contractors would be required to use the database.

The Department of Energy's Standardized Energy Efficiency Database (SEED) and Building Performance Database (BPD) could perhaps serve as a foundation for this process. SEED contains PII information but is linked to BPD; BPD could provide tremendous amounts of useful information to entrepreneurs that could help make AB758 more successful.

4.2 Outstanding Issues

Description	Proposed Next Step, if any
None seen other than lack of use of standardized database and tying rebates and permits closer together	

4.3 Additional Comments

5. Conclusion

5.1 Conclusion

The state is a public entity asking for access to all of its energy data in one database. Given its public status, PII is not applicable to the state and should not be a stumbling block in implementing a data repository with public access. As the data repository moves beyond alpha and beta status the state will encourage making the data available to interested parties that request it.

5.2 Recommended Next Steps

Implement a data repository and energy project database and populate it with state data as quickly as possible.

Appendix

Contact

Christine Awtrey

916-651-1227

christine.awtrey@energy.ca.gov

Reference Materials

Use Case 5 – Provided by Faraday/Brighter Planet Technology

1 Overview

1.1. Use Case Summary

Environmental non-governmental organizations, like the NRDC, requesting PII customer repayment history and energy consumption pre- and post-retrofit for energy efficiency, to support general financial decision-making on energy-efficiency investments through on-bill financing, and produce results that provide aggregate, non-PII findings that link energy usage to other relevant characteristics (e.g. geography, building characteristics, customer financial characteristics, and financing vehicle). In this case, the data is definitely PII, but the results – a decision whether a particular area, type of building, type of customer, or type of financing is viable – is non-PII.

1.2. Objectives

Improve and increase financing opportunities for energy efficiency work by providing greater certainty around energy savings and in turn financial return.

1.3. Actors

<i>Name</i>	<i>Role description</i>
Electric utility	Data owner
NGOs, financing entities, and program implementers	Data requestor

1.4. Applicable Statutes and Regulatory Rules

<i>Agency</i>	<i>Description</i>	<i>Applies to</i>
State of California	P.U. Codes 8380 and 8381	PII energy use data from an advanced metering infrastructure. Note that this statute does not limit access to data from a conventional metering infrastructure.

2 Use Case Details

2.1 Current Data Practices

<This section can quickly summarize how the process for this use case takes place today. It can be helpful in getting people grounded.>

2.2 Requested Data Practices

<This section should focus on the desired “to-be” state, without necessarily spelling out the technical solution. In other words, it should capture the process through which the parties want to interact, but not necessarily the tools and all the policies that need to be in place. If consensus can’t be reached, this section can summarize the options.>

3 High Level Requirements

3.1 Data Granularity Requirements and Data Use

<i>Data Type</i>	<i>Priority (H/M/L)</i>	<i>Aggregated/ Anonymized / Identifiable</i>	<i>Description/ Additional Comments</i>
Monthly usage	H	Identifiable	Customer-specific monthly energy use. Must include service address to allow linkage with other data.
Monthly billing	H	Identifiable	Customer-specific monthly bill amount.
Efficiency program participation	H	Identifiable	Any customer-specific data such as program name, participation date, acquisition channel, action taken, incentive received, etc. along with general program details such as dates, promotion channels, incentives, participation rate, etc.
Retrofit data	M	Identifiable	Any customer-specific data such as retrofit date, type, incentive, acquisition channel, contractor name, etc.

3.2 Data Collection and Maintenance Requirements

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
Data must cover the last 6 months at a minimum. Ideally data would cover the last 12 to 24 months.	H	A time series is critical to avoid biases from weather-related, seasonal, or random effects and identify patterns that precede and follow energy efficiency program participation.
Data must be machine-readable and accompanied	H	Any machine-readable format such as fixed-width text, delimited text, csv, xml, xls, etc. is

by a schema.		acceptable. This is critical for efficient processing and error avoidance.
Data must be transferred and stored securely. Third party must not share data or use for any other purpose, and must delete data after use.	H	To protect PII and comply with rules and regulations.
Data should be updated quarterly.	M	New data will allow further refinements of analyses and keeps results up to date.

3.3 Required Policy & Other Determinations

<This section should outline high-level policy requirements, e.g. if there is a need to have CPUC approve release of data. Anything else should also go here.>

4 Current Data Obstacles and Other Issues

4.1 Barriers

<This section should summarize all the barriers that currently exist or are anticipated by the stakeholders.>

4.2 Outstanding Issues

<This section should summarize any issues or open questions that the team wasn't able to resolve.>

4.3 Additional Comments

<Anything that didn't fit anywhere else can go here.>

5 Conclusion

5.1 Conclusion

<Conclusions about this use case.>

5.2 Recommended Next Steps

<Proposed next steps.>

Appendix

Contact

Ian Hough – Brighter Planet Technology Services
ian@brighterplanet.com
1.802.458.0441 x 316

Robbie Adler – Brighter Planet Technology Services
robbie@brighterplanet.com
1.802.458.0441 x 306

Use Case 5 – Provided by Harcourt Brown & Carey

Background

This memo summarizes two type of data collection and dissemination activities related to energy efficiency financing. These two activities are quite different from one another and can broadly be summarized as:

- “Backward-looking” utility bill payment performance history data to help financial institutions better understand how a portfolio of financial instruments collected on the utility bill might perform, using that account performance history as a proxy for likely financing performance. This data has a short-term time horizon for completion.
- “Forward-looking” energy efficiency project-level, energy and finance data that will be used to assess the effectiveness and performance of the new energy efficiency financing pilots. This activity has a longer-term time horizon since data will be collected and analyzed as the pilots are implemented and operated.

This memo summarizes each of these two separate activities in turn.

Backward-Looking Historic Performance Data Request

Backward-looking historic performance data. Financial institutions have requested information on the historical performance of utility customer accounts. Financial institutions will use this data to assess whether they will participate in the on-bill repayment (OBR) pilots that have been ordered by the CPUC and what financial product terms (i.e. interest rate, security, maturity) & underwriting standards (i.e. utility bill repayment history, property loan-to-value ratio) they will use.

This section summarizes the data that financial institutions have indicated would be useful to them in assessing the value of on-bill repayment (OBR).¹

What Data	Description
Delinquency and default frequency rates at market level (geographies)	The level of detail at which this data is provided is to be negotiated but in no case is data requested that would be customer-specific. A minimum level of detail is data broken down by customer class. Financial institutions also indicated that data would be useful (meaning it would make them more willing to participate in a financing program and to do so at more advantageous terms) if such data were

¹ This information is synthesized from discussions with New Resource Bank, One Pacific Coast Bank, Environmental Defense Fund and Citibank.

	provided at the level of: <ul style="list-style-type: none"> • Geography (eg. Zip code) • NAICS code • Customer size (by energy use or square footage)
Partial payment trends	Level of detail desired is the same as in the above description.
How often do meters/properties “go dark” or drastically reduce usage?	Overall customer class data
Disconnection	Data that describes by customer class: <ul style="list-style-type: none"> • Frequency of disconnection by customer class • Result of disconnection (pay full arrearages, payment over time of arrearage).
A written description of utility standard collection procedures in the event of utility bill under- or non-payment.	

Forward-Looking Pilot Performance Needs

Forward-looking pilot performance data. The CPUC has instructed that the energy efficiency financing pilots operated in the 2013-2014 program cycle be accompanied by a robust data collection effort that can help to increase confidence in the performance of financial products specifically targeted at energy efficiency (e.g. do loans for energy efficiency default less frequently than loans for cars or televisions?), and how different program elements influence this performance (e.g. does bill neutrality reduce customer financing default rates?). This data set is focused on collection and dissemination of:

- Pre-installation energy and financial information (borrower, property, project, financial instrument, energy consumption and projected savings)
- Post-installation (financial instrument performance and actual energy savings).
- Energy project data such as measure type, expected energy savings and similar data

We are happy to provide a list of expected data needs, however we feel strongly that this element of the financial data initiative should be coordinated closely and linked with the California Energy Data Center Initiative. Specifically, this coordination and linkage should be tied to:

- Treatment of personally identifiable information
- Storage of such data
- Dissemination of such data to all parties

We believe that in many cases, protocols and procedures for these activities exist in the financial industry through a long history of dealing with mortgage backed securities and related financial instruments, however any process developed for a financing initiative must be carefully linked and coordinated to a broader CPUC effort on data privacy.

Use Case 5 – Provided by NRDC

- Use Case 5 describes research on customer-level, address-level utility customer data.
- Contemplates ability to combine CEUI with other data sets (e.g., appraisal records) to enable new insights. Data must be matchable by a field such as customer name or address.
- Results are built-up from pii CEUI, but results only show non-pii metrics, averages, and summaries.

The question is how to enable analysis & obtain useful results while maintaining required confidentiality of customer data.

Use Case 5 could occur in either scenario	<i>Primary scenarios to consider</i>	<i>Questions to address</i>
→	<i>Disclose results based on CEUI</i>	What level of aggregation / obfuscation is needed? See LGSEC two-axis model (5/13/13).
→	<i>Utility permits researchers to access CEUI</i>	What are terms for eligibility, MOU/NDA, timing, and other to enable access for researchers? See "Utility Process" draft memo (5/8/13). See Energy Data Center memo, Appendices.
	<i>Utility delivers whole building info to building owner</i>	How to verify status/identity of owner, and what terms of use and commitments are required to deliver data? See NRDC/IMT Comment Letter of 4/29/13.
	<i>Utility delivering CEUI to a designated party with customer permission</i>	<i>[Outside scope of this proceeding]</i>

Options to consider for Use Case 5

1. Standard reports (produced by the utility)

- *Process that is responsive to market needs*
- *Method to look at data across utilities*
- *Method to match to other data sets*
- *Cost-based fees?*

See IOU's
"Utility
Process" draft
memo 5/8/13

2. Process and tools to enable designated researchers to access utility systems (subject to terms of use, qualification, etc.).

3. Central "Data Center" with tool that permits queries against CEUI but only shows researcher anonymous results.

- *Compare to CoreLogic, Fidelity, others with loan data*

Contemplated
in original
CPUC Staff
Memo

**Submitted for discussion purposes only Philip Henderson
Natural Resources Defense Council
May 22, 2013**

Use Case 6 – Provided by Solar City

1. Overview

1.1 Use Case Summary

This use case envisions an electronic data center (EDC) with the following characteristics:

- *EDC will host anonymous and homeowner-level energy consumption data.*
- *Third-parties will have the ability to register with the EDC to analyze data and submit proposals after meeting specific criteria like business licenses, safety certifications, etc.: .*
- *EDC will have the ability notify energy consumers, informing the homeowner that a third-party has developed a proposal.*
- *If the customer opts-in, EDC will have the ability to present third-party proposals to homeowner via email or online portal.*

Solar installation and energy efficiency companies will analyze anonymized, household level energy consumption and billing data to identify customers/households that may benefit from energy services. After analyzing energy bills, these third parties will develop proposals for these households and submit them to the EDC. Customers will have the option to select their preferred communication method (i.e. email, phone, through portal, etc) Based on the communication preferences indicated by the customer, the EDC will notify customers that trusted third-parties have developed household specific proposals, including estimates of energy and bill savings, and would like to market their services. If customers opt-in, the electronic data center will forward the detailed proposals from third-parties to the customer. Personally identifiable information is never revealed to any third party, unless the customer contacts the third party directly.

1.2 Objectives

Analyze customer usage data to better understand opportunities to deploy distributed renewable energy and energy efficiencyimprovements at customer’s home, reducing their energy consumption and bills.

Reduce customer acquisition costs, a major lever to facilitate more widespread adoption of distributed renewable energy and energy efficiency, by helping third party renewable energy and efficiency installers present data-driven and tailored proposals to customers who can most benefit from their services.

Increase precision of solar and home retrofit systems, since real data helps right-size systems.

1.3 Actors

<This section should describe the participants in this process. At a minimum, this should specify the data owner and the data requestor. This may end up being the same across all of the use- cases, but maybe different.>

<i>Name</i>	<i>Role description</i>
Utility Organization	Collects and provides billing and smart meter data to Energy Data Center, run by a third-party host.
CPUC	Defines the Energy Data Center's database funding, structure policies and Processes.
Academic institution	
Energy Data Center	Warehouses customer data; provides household level, anonymized data to 3 rd Party Service Providers; notifies customers of proposal availability; if customer expresses interest, EDC provides detailed proposal to customer, including 3 rd Party Service Provider contact information.
3 rd party Service Provider	Develops household specific proposals; notifies EDC which households (e.g. based on a household reference # assigned by the EDC), sends proposals to Energy Data Center; engages with end customer if contacted to implement proposal/perform work. Provides data on specific upgrade measures proposed and performed, and the estimated distributed generation or energy savings of these so there
End-Use Customer/Households	Evaluates information provided by EDC and determines if interested in receiving full proposal details; contact 3 rd Party Service Provider to perform work.

1.4 Applicable Statutes and Regulatory Rules

<This section should describe any specific rules or regulations that already apply to this use case, e.g. if there are requirements that stem from a specific CPUC mandated program.>

<i>Agency</i>	<i>Description</i>	<i>Applies to</i>
CPUC		
Other		

2. Use Case Details

2.1 Current Data Practices

Today, there is no way for energy service companies to analyze anonymous household-level data to identify customers/households who can most benefit from these solar and energy efficiency services. Currently, in order to obtain this data, customers must opt-in, and elect to send their monthly energy bills or Green Button Data to solar and energy efficiency companies. This presupposes that customers are aware of the opportunities they have to deploy energy efficiency and distributed generation technologies to reduce their energy bills. It also presupposes customers have the technical knowledge and time to download their greenbutton data and provide it to third parties. It also presupposes that the customer's utility implemented Green Button Connect. An approach that allows energy service companies to see anonymized individual household data and develop proposals using household specific data would help ensure that customers that stand to see substantial energy and bill savings are made aware of these opportunities and assess the market potential of specific energy efficiency technologies or products.

2.1 Requested Data Practices

The Energy Data Center would host 13 months of 15-minute interval electricity and natural gas consumption for every homeowner in the state in a database accessible to approved third parties. The data would keep the homeowner's name and address anonymous to third parties, with each household tagged with a unique randomized ID. Third parties would be given access to this database to allow them to run queries and develop specific proposals. Third parties would then submit proposals to the Energy Data Center and alert customers of the availability of a proposal along with high level information regarding estimated energy and bill savings. The Energy Data Center would have the ability to forward these proposals to customers, if customers authorized the EDC to do so..

3. High Level Requirements

3.1 Data Granularity Requirements and Data Use

<This section should summarize the type of data that we are talking about for this specific use case. Not every data element should be spelled out at this level – just the type/categories of data.>

<i>Data Type</i>	<i>Priority (H/M/L)</i>	<i>Aggregated/Anonymized/Identifiable</i>	<i>Description/Additional Comments</i>
13 months of 15 minute, household level interval data	H	Anonymized	This information is fundamental to the development of proposals to be responsive to specific customer opportunities. 13 months of kWh and BTU consumption recorded by a customer's meter at a 15 minute interval. 13 months allows seasonal
Building and Occupancy Characteristics	L	Anonymized	Provide the opportunity for customers to self-report basic information that drives their energy use and more accurate analysis (# occupants, type of business, Square feet, single family, electric vs gas water heater, pool?
Climate Zone	M		Important for purposes of understanding environmental context of households and likely key
13 months of household-level customer billing data	H	Anonymized	Monthly bill costs and relevant Utility, and tariffs under which the customer takes service. This is critical driver of project

3.2 Data Collection and Maintenance Requirements

<This section should outline high-level functional requirements (technical and non-technical). For example, any requirements about how frequently data needs to be updated, what format it needs to be in, security specifications etc.>

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
<p>EDC should be capable of securely storing household level consumption and billing data, along with other relevant data (climate zone) in a format that can be easily queried and utilized by third party service providers to develop proposals while maintaining the anonymity of the specific households. The data should cover the past 13</p>	H	
<p>The Electronic Data Center must include functionality that allows 3rd parties to submit a web link to promotional materials to a clearing house. The promotional materials must be tailored to an individual home and based on the individual home's data. The EDC must have functionality to prevent SPAM.</p> <p>When a third-party generates a proposal for the homeowner, the Electronic Data Center will notify the homeowner using the homeowner's preferred method of communication. The communication will include high level estimates of energy and bill saving. If the customer opts-in, the EDC will provide the customer access to the full proposal.</p> <p>The customer's personally identifiable information will never be revealed to anyone but the Electronic Data Center, until such time as the customer determines they wish to contact a 3rd Party Service Provider in response to a</p>	H	

3.3 Required Policy & Other Determinations

<This section should outline high-level policy requirements, e.g. if there is a need to have CPUC approve release of data. Anything else should also go here.>

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
The CPUC shall approve the release of any data that could be personally identifiable, as well as determine what information the EDC will house and make available to 3 rd Party Service Providers.	H	
Funding levels and sources of funding for		
Selection of and EDC contractor		
Qualifications of third parties to access the EDC.		

4. Current Data Obstacles and Other Issues

4.1 Barriers

<This section should summarize all the barriers that currently exist or are anticipated by the stakeholders.>

<i>Barrier Description</i>	<i>Priority (H/M/L)</i>	<i>Current/Anticipated</i>
Utilities must either transfer data to a common state database OR make their data available in a standard format for a common state web service to access.	H	
Common data format – all household usage and billing data will need to be in a common format to facilitate inclusion in the envisioned EDC database	H	

Protocols to address how and what information from 3 rd Party Proposals will be provided to customers and in what format will	H	
--	---	--

4.2 Outstanding Issues

<This section should summarize any issues or open questions that the team wasn't able to resolve.>

<i>Description</i>	<i>Proposed Next Step, if any</i>
e.g. there wasn't enough information about how XYZ is being done today	

4.3 Additional Comments

<Anything that didn't fit anywhere else can go here.>

5. Conclusion

5.1 Conclusion

<Conclusions about this use case.>

5.2 Recommended Next Steps

<Proposed next steps.>

Appendix

Contact

<May want to include the list of people who participated in the development of the use case or who to contact with questions.>

Reference Materials

<Reference Materials.>

Use Case 7 – Provided by LGSEC and San Francisco Department of the Environment

1. Overview: Use Case 7: Benchmarking Whole Building Annual Energy Performance with Monthly Data

1.1 Use Case Summary

Pursuant to AB 758, AB 1103, SB 1476, and local energy efficiency policies such as San Francisco’s Existing Commercial Buildings Energy Performance Ordinance, provide building owners and managers with automated access to whole-building monthly energy consumption data to conduct building benchmarking analyses using ENERGY STAR Portfolio Manager. Purposes of benchmarking with ENERGY STAR Portfolio Manager include measuring, managing, and disclosing energy performance. Managing energy performance includes identifying, valuing, and tracking gross monthly savings from energy efficiency upgrade projects.

”Monthly whole building data for benchmarking annual energy performance” in this case is:

- Aggregated to monthly temporal resolution (only),
- Inclusive of all metered energy sources serving the building (electricity and natural gas – whether metered by Smart Meter or other. “Whole building monthly energy consumption data” may also include energy sources such as district steam and on-site renewables.),^{2/}
- Inclusive of all energy customers – including any number of separately metered tenants for non-residential, multifamily, and mixed-use buildings.
- Suggested to be aggregated to the level of the whole building for each energy commodity delivered to the building.

1.2 Objectives and Ratepayer Value

The main objectives of providing whole building monthly energy consumption data are to:

- Enable building owners to comply with laws designed as energy efficiency programs which require benchmarking, without separate written consent to obtain energy use data reflecting, or summarizing, tenant energy use.^{3/}
- Enable building owners (commercial and multifamily) to benchmark and track the

2/ Though district steam, on-site renewable energy delivered to meter(s) serving the building, and non-smart-metered electricity & gas might not be directly addressed in this proceeding, we suggest that any rules or procedures stemming from the use case explicitly acknowledge and such non-smart-metered sources and establish common procedures that can be applied to such sources by whatever entity serves such resources used by a given building. An Investor Owned Utility may not be the sole party to provide benchmarking data pursuant to AB 758, AB 1103, or a complimentary local ordinance.

3/ The technical ability of utilities to sum the energy use for all meters serving a building may vary at this time, but adopted regulations require building owners subject to AB 1103 to disclose annual whole building energy performance

energy performance of their buildings over time, enabling them to identify under-performing buildings and supporting investment in energy upgrades.

- As a reasonable data privacy protection, preferably enable building owners to gain access to monthly whole building data for benchmarking as the sum of the monthly energy used in the entire building for each commodity served to the building, without directly individual tenant meter’s energy consumption.
- Enable property owners to disclose building-level energy information to prospective tenants, lenders, and/or potential buyers. Energy information disclosure enables these parties to analyze the full cost of their decisions and to compare costs across buildings.
- Enable cost-effective delivery of energy efficiency resources – including incentives, technical assistance, and financing – targeting buildings with the greatest relative energy use and greatest relative carbon emissions.
- Enable systematic benchmarking as a low cost mechanism for building owners and local governments to measure the outcomes of energy efficiency upgrade projects

The provision of these data is of significant value to ratepayers. Ultimately, monthly whole building data for benchmarking annual energy performance will increase the effectiveness of ratepayer energy efficiency investments by:

- Identifying under-performing buildings
- Demonstrating the outcomes of building-level upgrades and of programs designed to increase energy efficiency in existing buildings
- Enabling prospective tenants, lenders, and buyers to analyze the true cost of investment/decisions, and to empower
- Reducing the time required for owners, tenants, and utilities to communicate about, navigate, and administer a cumbersome consent procedure in order to complete a state-mandated process.

1.3 Actors

Name	Role description
Utility	Provide whole building monthly energy consumption data in a manner consistent with state law and adopted regulations, that explicitly protects the utility in the course of aiding customers’ fulfillment of relevant requirements, provides reasonable protection of entities’ privacy, and enhances all parties’ ability to deliver energy efficiency benefits to ratepayers.
CPUC	Establish and monitor a system for the provision of whole building energy use data
CEC	Observe compliance with statutory energy efficiency requirements based primarily or substantially on energy use data, including AB 1103 and AB 758.
Building owners and managers	Request and employ whole building energy use data in order to monitor and disclose energy performance, comply with state and local laws and inform efficiency investment decisions.

Name	Role description
Local Governments	Enforce mandatory benchmarking and limited disclosure of summary statistics, where applicable. Deliver voluntary energy benchmarking programs to multiple tenant buildings to support better management of energy resources and cost-effective achievement of local and state energy efficiency and greenhouse gas emissions goals and requirements.
EPA ENERGY STAR Portfolio Manager	Partner with utility to aggregate individual tenant data into building-level energy use information and distribute whole building data, weather normalized energy use intensity and benchmarking score to building owners and managers and other parties.
3 rd party	Partner with utility, building owner, and/or local government to provide supplemental analysis based on monthly whole building data for benchmarking annual energy performance, and use this data, for example, to identify underperforming buildings and confirm eligibility for program resources.

1.4 Regulatory Proceedings and Rules that Currently Apply

Agency	Description	Applies to
CEC	AB 758	Existing buildings in California
CEC	AB 1103	Existing non-residential buildings of 5,000 gross square feet or larger at whole building transaction (sale, lease, or refinance).
CPUC	- SB 1476/PUC 8380 (e) (2) allows utility customer data disclosure in the implementation of energy management, energy efficiency - SB 1475/PUC 8380(e)(3) Allows utility customer data disclosure required or permitted by state or federal law, or an order of the commission.	Utility customer data used in the course of duly established laws and energy efficiency programs.

Agency	Description	Applies to
San Francisco	<p>Existing Commercial Buildings Energy Performance Ordinance requires annual energy benchmarking with ENERGY STAR Portfolio Manager with limited public disclosure of summary statistics about whole building annual energy performance.</p> <p>Data reported to the city is limited to statistics that summarize annual whole-building performance, including:</p> <ul style="list-style-type: none"> - kBTU consumed per square foot per year - 1-100 ENERGY STAR rating - Pounds of CO2 emitted <p>Commodity energy consumption is not reported to the city.</p> <p>The ordinance also requires an energy audit from a qualified professional (defined therein) every 5 years.</p>	Applies annually to existing non-residential buildings of 10,000 conditioned gross square feet or larger.
CEC, CPUC	Cal Civil Code 1798.24(e) authorizes disclosure of an individual's personal data when the information is necessary to fulfill statutory duties – such as compliance with AB 1103.	Utility customer data

Agency	Description	Applies to
AB 32 Scoping Plan	Establishes goals and mechanisms for achievement, applicable to utilities, CPUC (oversight), CEC (oversight), and local governments.	

1.5 Use Case Details

1.5.1 Current State Narrative

Currently, in order to benchmark a building with ENERGY STAR Portfolio Manager, a commercial or multifamily building owner must obtain written consent from each utility customer (i.e., each individual tenant in facilities where tenants purchase energy from a utility) and provide service agreement IDs (SSID) and meter numbers to the utility to connect the usage data with the pertinent account in Portfolio Manager. The utility then uploads all available data connected to the specific meters into Portfolio Manager through a web service termed the “Automated Benchmarking System.” Obtaining written consent from each tenant is time-consuming and is a significant barrier to building-level energy monitoring and management.

Consequently, consent is most readily obtained in buildings with fewer tenants, and is increasingly logistically prohibitive as the number of tenants grows. Therefore it is currently impractical to benchmark large multifamily buildings where tenants are responsible for electricity or gas use, increasing the cost of delivering energy efficiency benchmarking assistance, as well as concomitant incentives and services, to such buildings. Similarly, common non-residential uses such as multi-tenant retail are logistically challenging to reach for the same reason, preventing building owners from getting recognition for high performing buildings as well as expediently complying with AB 1103, building labeling provisions enabled by AB 758, San Francisco’s energy efficiency ordinance, and voluntary benchmarking programs – all dependent on the building owner having reasonable yet limited access to monthly whole building energy consumption.

One tool that may help in the future, but will not resolve this problem in a consistent and timely way, is lease language. Commercial leases of 5 years with options to extend are common in commercial uses other than retail, and even 10 years is not uncommon. In buildings with long average lease duration, decades may be required before tenant turnover occurs throughout the entire building.

1.6 Future State Narrative

Utilities should be able to aggregate whole building data from all meters providing services and provide it to the building owner/manager without the need for service IDs

and written authorization from account holder. If the utility cannot aggregate all meters associated with a building address, the owner can provide the meter IDs. Opt out provisions could be made available to tenants with national security or other barriers to utility use disclosure. Data must be provided in a format that can be manipulated, as described in LGSEC use case 1, above.

Benchmarking allows building owners to effectively manage and reduce energy use. Management expert Peter Drucker said, “What’s measured improves.” Drucker’s wisdom is supported by a 2012 analysis by US EPA of all buildings using ENERGY STAR Portfolio Manager – the common tool required by all US cities with benchmarking policies. EPA found that just 35,000 buildings had used Portfolio Manager to consistently benchmark energy consumption for 4 years from 2008-2011, and that those 35,000 buildings had reduced their energy consumption by an average of 2.4% per year.

http://www.energystar.gov/ia/business/downloads/datatrends/DataTrends_Savings_20121002.pdf

Similarly, a 2012 report for the California Public Utilities Commission found benchmarking was highly correlated with energy efficiency improvements, and a strong catalyst for participation in rebate and incentive programs.

http://www.calmac.org/publications/Statewide_Benchmarking_Process_Evaluation_Report_CPU0055.pdf

San Francisco’s 2011 Municipal Energy Benchmark Report found that benchmarked facilities total energy use decreased 1.1% and carbon footprint decreased 2.3%.

<http://www.sfwater.org/index.aspx?page=71>

The *California Long Term Energy Efficiency Strategic Plan* of 2008 calls for “100% of existing multi-family homes have a 40% decrease in purchased energy from 2008 levels” by 2020 (page 19). Achieving this goal will require a high frequency of significant physical energy efficiency upgrades in multifamily buildings, most of which have individually metered units. Building owners must make the decision to invest in these upgrades. Owners are much more likely to make these decisions if they have whole building data enabling them to participate in energy benchmarking and programs that target resources to under-performing buildings, or that might require benchmarking as a prerequisite to program participation as a demonstration of owner commitment.

1.7 High Level Requirements

1.7.1 Data and Aggregation Requirements

Data Type	Priority (H/M/L)	Aggregated/Anonymized/Identifiable	Description/Additional Comments
Monthly usage data, aggregated to the whole building	H	Aggregated yet Identifiable in some non-residential cases.	Balances legislative requirements to benchmark energy use in support of energy efficiency with requirements to provide reasonable privacy protections.

1.7.2 Functional Requirements

Requirement	Priority (H/M/L)	Additional Comments
Data shall include consumption data for the previous 12 months at a minimum.	H	Ratepayer investment in data delivery infrastructure is either complete (PG&E) or required by AB 1103. Data would continue to be delivered to Portfolio Manager.
Sum energy use for affected meters to the level of the whole building without exception for identifiability of a non-residential tenant.	M	
Data shall be inclusive of energy use for prior 12 months by former owner/tenants.	H	Partial building tenant turnover is more common than whole building tenant turnover. To be able to obtain 12 months of data for disclosure, all energy use for the period must be available. (Rare) exceptions for public safety or national security are likely to be necessary.

1.7.3 Policy & Other Requirements

Requirement	Priority (H/M/L)	Additional Comments
Pursuant to SB 1475/PUC 8380(e)(2) and PUC 8380(e)(3), and consistent with: the requirements of AB1103 as well as the intent of AB 758 and AB32; the Big, Bold energy efficiency objectives and programs of the CPUC, as well as local agencies and the State of California; CPUC shall order utilities to release monthly whole building data for benchmarking annual energy performance to a building owner via electronic upload to ENERGY STAR Portfolio Manager upon request. This order must apply even when limited personally identifiable information will be shared with the building owner.	H	Viable solution to balance privacy intent and requirements with building owner access to data in order to better manage energy use.

1.8 Barriers and Open Issues

1.8.1 Barriers

In a multi-tenant building (both commercial and multi-family), the meter configuration and bill payment responsibilities are divided between the property owner and the multiple tenants. The IOUs currently require that a data release authorization form to be completed and signed by each tenant. Obtaining this information is difficult for many reasons that relate more to logistics than privacy:

Barrier Description	Priority (H/M/L)	Current/Anticipated
Number of Meters: Tracking of multi-tenant meters requires capacity to handle a large volume of meters. For example, in a 40-unit building with individual meters for both gas and electricity, there could be more than 82 meters that must be tracked for a comprehensive whole-building perspective. In most communities, over 90% of buildings are individually metered for electricity, and around half are individually metered for gas.	H	Utilities have managed the challenges of “big data” for: - Billing - PG&E proxy benchmarking program
Accessing tenants – property owners/managers must find each tenant to request the data release authorization;		

Barrier Description	Priority (H/M/L)	Current/Anticipated
tenant schedules vary making this task difficult. The time investment required to reach tenants can be estimated at 15 – 30 minutes per tenant on average for a multi-family building. For the 40-unit example above, this would equal 10 – 20 hours of the property owner’s time. The time is typically inconveniently spread out over several weeks.		
Tenant transience – each time a unit turns over, the property owner must obtain a new data release authorization form from the new tenant. With some properties experiencing up to 50% annual turnover, this becomes a significant on-going task. In theory, adding data authorization to the lease is one potential solution to this obstacle. In practice, however, tenants do not typically receive their meter-specific SAID – which is required for the data release authorization – until up to a month into their tenancy. This means they will not be able to complete the form at time of lease signing, and the property owner must follow up with each tenant at a later date.		
Tenant contentiousness – the landlord/tenant relationship often has some contentious dynamics. Some tenants will not sign data release authorizations, and they may not be motivated to do so by a rebate program if they perceive the efficiency improvement as benefitting the property owner. Requiring release of this data as a condition of occupancy may be impossible in some cases, such as federally subsidized housing. Whole building monthly summary data for benchmarking annual energy performance would not be subject to these limitations.		
Tenant lack of interest: Rather than seeking to protect monthly energy consumption summary data (<i>not 15 minute interval data</i>) , it is more common for both non-residential and multifamily tenants to wish to decline to sign or discuss anything,		

Barrier Description	Priority (H/M/L)	Current/Anticipated
or to consent to avoid further discussion.		

1.9 Outstanding Issues

Description	Proposed Next Step, if any
Utilities may need to update data systems to be able to associate all meters collocated at the same address, for the purpose of aggregating and providing whole building data.	
Utilities may need to update data systems to share meter-level data with another entity such an energy data center.	

1.10 Additional Comments

1.11 Conclusion

1.11.1 Conclusion

The Commission must ensure that utilities aggregate whole building data from all meters providing services and provide it to the building owner/manager without the need for service IDs and written authorization from account holder.

1.11.2 Recommended Next Steps

Establish an advisory group made up of building owners or their representatives, local governments implementing energy ordinances and energy efficiency programs, and other stakeholders to define an agreed upon process, protocol, and timeline for providing the necessary data to building owners.

1.12 Appendix

1.13 Contact

Jody S. London, regulatory consultant to the LGSEC
 Oakland, California 94609
 510/459-0667
jody.london_consulting@earthlink.net

Barry Hooper

Green Building Program, Private Sector

San Francisco Department of the Environment
 1455 Market Street, Suite 1200, San Francisco, CA 94103
barry.hooper@sfgov.org (415) 355-3753

Use Case 7 – Provided by California Energy Commission

1. Overview

The rulemaking (08-12-009 filed December 18, 2008) for the Phase III Energy Data Center included an initial use case that identified the AB1103 disclosure as a potential end user. The intent is for a building owner to go the Energy Data Center, 'copy' energy use data for a building and 'paste' that into the United States Environmental Protection Agency's Energy Star Portfolio Manager (Energy Star Portfolio Manager). From there, the building owner will produce the required AB 1103 disclosure report. The use case goes on to discuss that this energy use data can be "anonymized" in some way to render it as non-personally identifiable information (PII).

Use Case 7

Building owners and managers seeking monthly energy consumption by building to conduct building benchmarking analyses pursuant to AB 758 and AB 1103, and publishing aggregate, non-PII results. In this case, raw data that is PII would likely be needed, but the results concerning the efficacy of the program, are not PII. Moreover, it may prove possible to anonymize such data via an algorithm.

What can be gleaned from the Energy Data Center information, which may be useful for the implementation of AB 1103, is default energy use data by building type, size and climate zone (or county). The regulations for AB 1103 allow the building owner to use a "safe harbor" option, which permits the use of an approximation of the energy use data if the actual energy use data is not available (California Code of Regulations (CCR) § 1684(e)). However, it can be difficult for building owners to generate this approximate data on their own, but the Energy Data Center can be used as the basis for a report for building owners who take the safe harbor option.

California Code of Regulations § 1684(e)

If there is information missing from a disclosure, and if the owner has made a reasonable effort to ascertain the missing information, the owner may then use an approximation of the information, provided that the approximation is identified as such, is reasonable, is based on the best information available to the owner, and is not used for the purpose of circumventing or evading this article.

1.1 Use Case Summary

The parameters for a report that building owners may reference when implementing the AB 1103 "safe harbor" option must be determined through careful review of the available data and its efficacy. This *Safe Harbor Report* will be based on PII data for all

building types and sub-types, sizes (as established by the available data), and county or climate zone (most building owners will not know their climate zone, but will know their county). To protect PII data, no single value reported should be an aggregate of less than 15 PII data sources, and none of the 15 PII data sources should individually represent more than 15 percent of the total resulting energy use (utilities refer to this as the '15/15 Rule').

1.2 Objectives

For building owners to have access to energy use data to report to the Energy Star Portfolio Manager and file the AB 1103 Disclosure Report.

1.3 Actors

Name	Role description
State	State (or their consultant) would download PII energy data, location data, building type data, and building use data to be consistent with data benchmarking and disclosure requirements.
Building Owners of California Nonres-Buildings	Building owners of California nonresidential buildings would use the resulting <i>Safe Harbor Report</i> to fulfill their obligations when using the safe harbor provision under CCR§ 1684(e).
Other Real Estate Professionals	As other real estate professionals as involved in the lease, sale or financing of California nonresidential buildings, they may have necessity to perform the same function as the building owners.

1.4 Applicable Statutes and Regulatory Rules

Agency	Description	Applies to
Energy Commission	AB 1103 CCR §1684(e)	The sale, whole building lease or whole building finance of nonresidential buildings in California.

2. Use Case Details

2.1 Current Data Practices

CCR §1684(e) requires that nonresidential building owners or operators disclose Energy Star Portfolio Manager benchmarking data and ratings, for the most recent 12-month period, to a prospective buyer, lessee, or lender.

The regulations also require electric and gas utilities to maintain records of the energy consumption data of all nonresidential buildings to which they provide service, in a format compatible for uploading to the Energy Star Portfolio Manager for at least the most recent 12 months. Upon written or secured electronic authorization of a nonresidential building owner or operator, the utility is required to upload all of the

energy consumption data for a building to the Energy Star Portfolio Manager in a manner that preserves the confidentiality of the customer.

Presently, utilities require a signed release from the customer (typically the current lessee) before releasing non-aggregated data to the building owner. The utilities are firm in their position that anything less than the 15/15 Rule to release data requires their customer’s express consent and release.

Since many building owners are unable or unwilling to gather the necessary releases from lessees, they are turning to the safe harbor option. Under this option the building owner must generate an approximation of the missing information, provided that the approximation is identified as such, is reasonable, and is based on the best information available to the building owner.

2.1 Requested Data Practices

AB 1103 created a new requirement for building owners and a new obligation for utilities and Energy Star Portfolio Manager. These three entities would prefer that another governmental body, that can access the necessary PII data, produce the ESPM 0 to 100 score as the disclosure without releasing the actual energy use data or average energy use per square foot. However, this government body was not provided for in statute, nor was only the release of the ESPM 0-100 score as the disclosure.

3. High Level Requirements

3.1 Data Granularity Requirements and Data Use

Data Type	Priority (H/M/L)	Aggregated/ Anonymized/Identifiable	Description/Additional Comments
Energy Use	H	Identifiable	Monthly
Building Type & sub-type	M	Identifiable	Based on the occupancy permit/authorization
County & Climate Zone	M	Identifiable	Based on address.

3.2 Data Collection and Maintenance Requirements

Requirement	Priority (H/M/L)	Additional Comments
Energy Use Data	H	Updated annually
Building type and location	L	Updated as needed
Data should be available over a number of years (5-10).	L	

3.3 Required Policy & Other Determinations

Requirement	Priority (H/M/L)	Additional Comments
The Energy Commission and the CPUC should enter into a non-disclosure agreement.	H	
The <i>Safe Harbor Report</i> should not release PII data.	H	

4. Current Data Obstacles and Other Issues

4.1 Barriers

Barrier Description	Priority (H/M/L)	Current/Anticipated
Cost/Time of contractor to prepare Safe Harbor Report.	M	The Energy Commission would contract out for this report.

4.2 Outstanding Issues

Description	Proposed Next Step, if any
None	

4.3 Additional Comments

None

5. Conclusion

The proposed *Safe Harbor Report* would make use of the PII data, but would not release PII data to the public. The Report would be a significant and immediate use to building owners of nonresidential buildings as they are required to comply with AB 1103 disclosure regulations.

5.2 Recommended Next Steps

Propose project through management approval process.
Review available data in Energy Data Center to determine appropriate deliverables for eventual contract proposal.

Appendix

Contact

Christine Awtrey
California Energy Commission
916-651-1227
christine.awtrey@energy.ca.gov

Reference Materials

AB1103
CCR1684

Use Case 8 – No Additional Proposal Provided

Use Case 9 – Provided by California Department of Community Services

The Department of Community Services and Development (CSD) submits the present comment on the California Public Utilities Commission (CPUC) ruling of February 27, 2013, establishing next steps for the four major utilities and working group participants, concerning the provision of energy usage data and the collaborative process to identify data “use cases” where personally identifiable information (PII) may be involved.

Specifically, CSD proposes that a supplementary “use case” be developed to address data sharing in connection with the coordination of the *low-income* customer programs of the Investor Owned Utilities (IOUs) and the federally-funded *low-income* client programs of CSD. To the extent that this data use falls outside the proposed scope of the present rulemaking phase⁴ of 08-12-009 – with regard to *non-usage* and *non-PII* data in particular – CSD proposes that consideration be given to the issuance of a similar ruling for the purpose of effecting greater coordination and reducing duplication of effort between the programs.

Proposed Use Case

Governmental agencies, like CSD that implement federally-funded energy efficiency programs for low-income persons such as the Low-Income Home Energy Assistance Program (LIHEAP) and the Department of Energy Weatherization Assistance Program (DOE WAP), endeavoring to coordinate the delivery of energy services with similar services provided by IOUs under CARE and Energy Savings Assistance Program (ESAP), through the reciprocal sharing of: 1) historical, non-PII, property-centric weatherization data; 2) historical PII weatherization data; and 3) customer/ client PII, involving eligibility, account information and energy usage data, all shared with the consent of the customer/ client.⁵

Background and Analysis

On March 17, 2009 the CPUC and CSD entered into a Memorandum of Understanding (MOU) “to leverage and coordinate existing programs for low-income energy efficiency and utility assistance to maximize the energy efficiency and health and safety of low-income households, and reduce the energy burden of economically vulnerable Californians...”⁶

⁴ The proposed use case does not involve either the Smart Grid System or Energy Data Center, but rather applies to a special purpose data base that will be developed for the CPUC and CSD low-income programs.

⁵ CSD and the IOUs are working jointly to establish a framework to facilitate the sharing of customer, utility account and service activity information with the consent of the customer/client. Due to the reliance on customer consent, this approach will only apply to information associated with future program participants and would exclude information associated with previously assisted customers and weatherized residences.

⁶ MOU, p. 1.

A primary objective of the MOU was to coordinate the low-income energy efficiency programs of the IOUs⁷ with CSD's energy and community service programs that improve the quality of life for the low-income population of California. The MOU referenced CPUC policy goals and targeted outcomes that were to be achieved through "increased collaboration and partnerships between itself and the IOUs and other federal, state and local agencies and community based organizations providing services to the low-income community, to leverage and coordinate existing programs, services, tools and funding."⁸ For its part, CSD was to urge local service providers (LSPs) "to coordinate with other low-income utility programs where possible, including the CARE program, the LIEE⁹ program..."¹⁰

The MOU contemplated attaining program coordination through cooperative agreements and facilitating collaboration between the IOUs and the LSPs. Among the many objectives specified in the MOU was the following:

"Development of a database of information about scheduling and service delivery that both LIEE providers and LSPs can use to coordinate services to eligible homes where possible and coordinate funding streams to maximize the number of energy saving and health and safety measures installed in low income households."¹¹

In the past year the CPUC, IOUs and CSD have initiated pilot programs designed to establish a model framework for attaining the various goals and objectives set out in the MOU, including a data sharing endeavor. A data sharing task force, composed of representatives of CSD and the IOUs, has undertaken efforts to develop a comprehensive statewide, low-income program database, perhaps modeled on a prototype low-income database tool developed by the Southern California Gas Company and Southern California Edison. The ultimate configuration of the statewide low-income database will depend upon: 1) the nature of the data collected; 2) the data sharing objectives; 3) who will provide the data; 4) who will have accesses to the data; and 5) the manner in which it will be used. Those same factors will inform and shape privacy issues and customer/ client PII data security concerns, as well the consequent obligations and liabilities of those providing, managing and utilizing the data.

The proposed use case would provide guidelines for the sharing of data through the statewide low-income database tool, to include the identification of data elements in which PII data *is not* involved as well as data elements that do involve PII data. To better understand the underlying issues, a closer look at the nature of the data and the proposed uses is required.

⁷ See the California Long-Term Energy Efficiency Strategic Plan (decision D.08-09-040) and the California Investor Owned Utilities' 2009-2011 Low Income Energy Efficiency (LIEE) portfolios (decision D.08-11-031).

⁸ MOU, p. 2.

⁹ The LIEE (Low Income Energy Efficiency) program was the predecessor program to ESAP.

¹⁰ MOU, p. 4.

¹¹ MOU, p. 6.

The MOU provides that the coordination and leveraging of the CPUC and CSD low-income programs through cooperative agreements such as the statewide low-income database "...will be designed to maximize opportunities for program and administrative coordination between the cash assistance, energy efficiency and weatherization services..."¹² of each party. The data associated with these functions varies considerably, and each intended use will carry its own considerations with respect to privacy, data security, and terms and conditions under which data may be shared.

For example, the underlying objectives sharing *weatherization* data are twofold:

1. Avoiding duplication of effort in order to optimize the utilization of public and ratepayer resources and to minimize waste; and
2. Leverage resources through cooperation between service providers in order to increase efficiency, effectiveness and to optimize benefit to the public.

To coordinate and leverage weatherization services under the ESAP and LIHEAP/ DOE WAP, service providers need to have access to the following historical¹³ data:

- Addresses of weatherized properties
- Date weatherization services were provided
- Measures installed
- Program or funding source utilized

It is important to note that customer/ client energy *usage* data is *not* a component of historical weatherization data needed for present purposes, and therefore it is not clear that the privacy and security protections codified in the Public Utilities Code¹⁴ and the Information Practices Act (IPA)¹⁵ apply.¹⁶ There is disagreement among legal counsel as to whether the mere mention of an address or residence constitutes PII even when *usage* data is not involved (with respect to the PU Code) or when it does not disclose information "in a manner that would link the information to the individual to whom it pertains"¹⁷ (with respect to the IPA). It is also unclear that an address used in the context of historical weatherization data is truly "customer-provided" information¹⁸ or otherwise associated with a particular low-income customer or client when the occupant of a weatherized property is, in a very large percentage of cases, not the occupant of the property at the time the weatherization service was provided.

¹² MOU, p.5

¹³ Historical data consists of information retained in IOU, CSD or service provider files, which typically is not readily accessible to other potential data users.

¹⁴ See PU Code § 394.4(a) and § 8380(a), the former cited at CPUC Ruling of February 27, 2013, p. 12.

¹⁵ California Civil Code § 1798 et. seq.

¹⁶ See also Commission Decision 11-07-056 of July 2011, which confines privacy and security protections to "Energy Usage Data."

¹⁷ California Civil Code § 1798.24.

¹⁸ Would an address cease being "customer-provided," if weatherized properties were identified and filed by parcel number (a data item that is not contained in any application form) rather than by street address?

This begs the question whose PII is being protected? Is it the tenant who qualified for services and occupied the property five years ago? The current tenant? The property owner who authorized the work? The current property owner? Further, a substantial portion of CSD weatherization work requires building permits, which are public records that reference property addresses as well as descriptions of the work conducted. Can information already in the public domain be considered PII?

It is submitted that historical weatherization data is *property-centric*, not customer/client-centric, and therefore is not PII that requires privacy protections.¹⁹ Accordingly, the use case should reflect this distinction and/ or a rulemaking should provide that in the context of low-income program weatherization services, an address alone, absent other customer-specific information, cannot be deemed PII that need be protected under the PU Code and CPUC decisions.

The statewide low-income database has application and potential benefit to the CPUC, CSD, the IOUs, and the LSPs beyond weatherization, namely with respect to the utility assistance and subsidized utility payment programs (“assistance programs”). In order to effect and optimize the provision of benefits and availability of services to qualifying low-income individuals, significant exchanges of customer information are required, including names, addresses, account information, customer/ client qualification information, including income, energy usage data, etc. There is no doubt that this type of data is PII that must be protected in accordance with statute, and further that such information should not be shared without informed customer/ client consent. Legal counsel are in agreement on this point. Questions remain, however, about the specific nature and breadth of the required consent and the manner in which the PII is gathered, stored and accessed, as well as the permissible uses.

In the context of assistance programs, customer/ client PII is provided to the IOU or the LSP with the understanding and the expectation that the sharing of data is essential to receiving the benefits and subsidies sought. Accordingly, this use is distinct from other use cases in which customer information is utilized by third parties for purposes other than for the customer/ client’s direct personal benefit. Such third party use, while for a worthy purpose, is not initiated by the customer/ client, and it typically does not benefit the customer/ client directly. Consequently, it can be argued that the PII involved is subject to higher and more rigorous standards of protection than is true with respect to assistance programs.

The distinction is one of expectation. Customer/ client PII exchanged between an LSP and an IOU to enable the LSP to pay the individual’s utility bill cannot be subject to the rigorous security protocols and non-disclosure requirements as would be expected in other use cases, if the programs are to be effective. Limitations and restrictions on PII collection, retention and sharing that impede the efficient provision of benefits and services defeat the purpose of the program. The expectations of the customer/ client

¹⁹ As noted, not all legal counsel subscribe to this interpretation. Counsel for the IOUs consider “address” and “residence,” terms referenced in the code sections, to be personally identifiable information, irrespective of context, and despite a lack of nexus between the property and an individual with a protectable interest.

make the present use case unique in that regard, and, accordingly, the use should be subject to different, less restrictive, rules and procedures.

As noted, informed customer/ client *consent* is the sine qua non of data sharing in the context of assistance programs. Once consent is given, reasonable restrictions on access and use of the PII are still in order. But the applicable standards and procedures must enable the efficient and effective implementation of the assistance programs, while at the same time maintaining a modicum of security and protection commensurate with the customer/ client's reasonable expectations.

Another distinct, but related, use of statewide low-income customer/ client data should be mentioned. Program funding impacting energy efficiency services, to include carbon emissions programs and the like require comprehensive granular energy usage data. The collection of such data would be both aggregated and customer/ client-specific, depending on the use and the needs of funding sources. Clearly, customer/ client informed consent for access and use of such PII data would be required. To the extent the statewide low-income database is utilized for such purposes, the standards and procedures associated with the proposed use case would be of relevance.

Note: Changes in federal program requirements will also impact CSD's need for customer/ client energy usage data from the IOUs. Beginning in Federal Fiscal Year 2014 the U.S. Department of Health and Human Services (HHS), funding agency of LIHEAP, will require states, through the applicable energy vendors, to provide client energy usage data in the annual LIHEAP Performance Measures Report, submitted to Congress. Among the many required items of data are the following: 1) the annual main heating fuel bill *of each assisted household*; 2) the annual main heating fuel consumption *of each assisted household*; 3) annual electric bill *of each assisted household*, when the main heating fuel is non-electric, and there is cooling; and 4) annual electric consumption *of each assisted household*, when the main heating fuel is non-electric, and there is cooling. This data requirement may or may not fall within the scope of the Smart Grid System, the Energy Data Center, nor indeed within the scope of this proposed "use case," but it is illustrative of the need for close coordination and cooperation between CSD and the CPUC regulated utilities, if there is to be a comprehensive statewide approach to low-income customer/ client services.

Conclusion

The proposed statewide database for CPUC/ IOU and CSD/ LSP low-income energy efficiency and customer/ client utility assistance programs – currently under development pursuant to the MOU of 2009 between the CPUC and CSD – poses unique issues and challenges with respect to PII protections and security. To the extent the proposed collection, management, access and use of non-PII, PII and consented authority for the sharing of data falls within the scope and purview of the current rulemaking, it is suggested that an additional or *special use case* is required to meet the needs of service providers and customer/ clients. If, on the other hand, the statewide low-income program database is deemed to fall outside the scope and purview of

rulemaking concerning the Smart Grid System and the associated Phase III Energy Data Center, it is suggested that a separate CPUC procedure be initiated to address the unique PII protections and security issues implicit in the project.

Department of Community Services and Development
May 2013

Use Case 10 – Provided by California Energy Commission

1. Overview

1.1 Use Case Summary

As a means of verifying compliance with the Title 24, Part 6 *Building Energy Efficiency Standards* as they relate to HVAC system efficiency and installation requirements, the Compliance and Enforcement Office needs to determine what HVAC systems are being imported into and sold in California for installation within the state. This determination can be made through the tracking of an HVAC's serial number, whereby any HVAC unit sold in the state will have its serial number entered into a database so that the serial numbers in this database can be compared to the serial numbers of HVAC units installed under the permitting process in local enforcement agencies throughout the state. This information can also be used for, and should be a requirement of, any HVAC rebate program within the state, whereby a rebate will be issued only for those HVAC installations where the proper permitting by the local enforcement agency has been accomplished.

1.2 Objectives

The objective of this (serial number tracking) program is to have the ability to compare HVAC units imported into and sold within California to those HVAC units installed under local enforcement agency permitting process, so that we can determine what HVAC systems are not being installed under the required permit process and as a result are not getting the required efficiency verification tests done to ensure properly-installed and efficiency-operating HVAC systems. Those HVAC systems that are determined to be imported into or sold within the state but whose installation has not been permitted through a local enforcement agency must be assumed to not meet the requirements of the *Building Energy Efficiency Standards*. For this reason, those HVAC systems that are shown to be imported into or sold within the state but not shown to be permitted by a local enforcement agency should not be eligible for efficiency rebates or other incentives.

1.3 Actors

<i>Name</i>	<i>Role description</i>
CEC	Data Requestor (HVAC Serial Numbers, Enforcement Agency Permit Data)
CPUC	Incentive information coordinator
CSLB	Data Requestor (HVAC Serial Numbers, Enforcement Agency Permit Data)
Utilities	Incentive Program information, review permit data for compliance
HVAC Manufacturers/Distributors	Supply distribution and sales data and serial numbers of HVAC units, and names of contractors purchasing the HVAC equipment

1.4 Applicable Statutes and Regulatory Rules

<i>Agency</i>	<i>Description</i>	<i>Applies to</i>
CEC	Title 24, Part 6; <i>Building Energy Efficiency Standards</i>	Residential and Non-Residential Buildings, Newly-Constructed and Existing
Other		

2. Use Case Details

2.1 Current Data Practices

There is currently no system in place to verify that HVAC units imported into or sold within the state are installed using the required local enforcement agency permit process, and in turn comply with the Building Energy Efficiency Standards requirements.

2.1 Requested Data Practices

A database that contains both the HVAC serial numbers of equipment that has been imported into or sold in the state of California, the serial numbers of the HVAC systems for which permits were pulled in all local enforcement agencies within California, and the contractor purchasers of the HVAC equipment. A comparison of these data sets would allow Compliance & Enforcement to determine the approximate compliance rate with the *Building Energy Efficiency Standards* and to improve compliance with these standards. The data set containing the serial numbers of permitted HVAC systems will be used by entities issuing incentives to determine the eligibility of installed HVAC systems to receive such incentives.

3. High Level Requirements

3.1 Data Granularity Requirements and Data Use

<i>Data Type</i>	<i>Priority</i>	<i>Aggregated/Anonymized/Identifiable</i>	<i>Description/Additional Comments</i>
HVAC serial number for imported/sold HVAC systems	H	Identifiable	Each HVAC system has a unique identifying serial number that can be used to track the HVAC system through the distribution channel.
HVAC serial number, installation location, and installing contractor for permitted installation	H	Identifiable	The serial number of the permitted HVAC system installation can be compared to the master list of all HVAC systems imported/sold in the state.

3.2 Data Collection and Maintenance Requirements

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
Monthly update of serial numbers and purchasers for HVAC units shipped to and sold in the state of California by manufacturers and distributors.	H	This data should be submitted to the CEC in Excel format. This data is critical to evaluate compliance with the requirements in Title 24, Part 6, <i>Building Energy Efficiency Standards</i> .

Monthly update of HVAC installation building permits from local enforcement agencies	H	This data needs to be provided so as to compare HVAC installation permits with the HVAC units distributed/sold in the state as tracked by their serial numbers (above). This data is critical to evaluate compliance with the requirements in Title 24, Part 6. This information should also be used by entities providing incentives for HVAC installations to insure that
--	---	---

3.3 Required Policy & Other Determinations

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
CSLB and CEC to require manufacturers and distributors of HVAC equipment, through the subpoena process if necessary, to submit to CSLB and CEC a monthly listing of serial numbers of HVAC units imported into or sold within the state. Also required is	H	CSLB and CEC, working with local enforcement agencies, will use this serial number data in conjunction with permit data, to determine which HVAC equipment has been properly installed using the permit process and which units have no record of being installed through the permit
CSLB and CEC to require local enforcement agencies to submit permit data, including HVAC serial numbers, to these agencies for use in comparison with the master list of HVAC serial numbers provided by manufacturers and distributors.	H	Building permits will need to contain the serial numbers of HVAC equipment installed.

4. Current Data Obstacles and Other Issues

4.1 Barriers

<i>Barrier Description</i>	<i>Priority (H/M/L)</i>	<i>Current/Anticipated</i>
HVAC Mfrs/Distributors refuse to give serial number data and purchaser of units imported/sold.	H	Anticipated
Lack of centralized database and manpower to collect and organize HVAC serial number, building permit, and rebate information	H	Current
CPUC only has oversight of IOUs, but we need incentive data (and permit checking) from all California Utility companies.	H	Current

4.2 Outstanding Issues

<i>Description</i>	<i>Proposed Next Step, if any</i>
None	

4.3 Additional Comments

Many projects that receive rebates also require permits. The utilities are doing a good job at collecting permit numbers for such projects however the information is not fed back to the permitting agencies for verification. IOU's and POU's should be required to identify permit numbers, permitting agency and project details on a secure website that is accessible only to regulators and permitting agencies, so that the permitting agency can confirm that a permit was procured for the project.

5.1 Conclusion

5.2 Recommended Next Steps

Appendix

Contact

Christine Awtrey
California Energy Commission
916-651-1227
Christine.Awtrey@energy.ca.gov

Reference Materials

Use Case 11 – Provided by Faraday/Brighter Planet Technology Services

1 Overview

1.1. Use Case Summary

Energy efficiency program implementer, contractor, consultant, research institution, city, county government, or other entity requesting PII individual energy consumption data, payment data, energy efficiency program participation, and retrofit activity to identify trends in customer participation in efficiency programs and retrofit activity. The requested data must be PII to allow linkage with other relevant data, but the results of analyses (e.g. trends) would not include PII.

1.2. Objectives

The objective of this use case is to improve energy efficiency program effectiveness. A thorough understanding of program trends would help utilities and program implementers increase participation rates and lower acquisition costs by uncovering ratepayer cohorts deserving additional focus. Ratepayers would benefit from the decreased cost and increased effectiveness of energy efficiency programs.

1.3. Actors

<i>Name</i>	<i>Role description</i>
Electric utility	Data owner
Efficiency Program Implementer, Contractor, Research Institution, City, County Government, or other entity	Data requestor

1.4. Applicable Statutes and Regulatory Rules

<i>Agency</i>	<i>Description</i>	<i>Applies to</i>
State of California	P.U. Codes 8380 and 8381	PII energy use data from an advanced metering infrastructure. Note that this statute does not limit access to data from a conventional metering infrastructure.

2 Use Case Details

2.1 Current Data Practices

<This section can quickly summarize how the process for this use case takes place today. It can be helpful in getting people grounded.>

2.2 Requested Data Practices

<This section should focus on the desired “to-be” state, without necessarily spelling out the technical solution. In other words, it should capture the process through which the parties want to interact, but not necessarily the tools and all the policies that need to be in place. If consensus can’t be reached, this section can summarize the options.>

3 High Level Requirements

3.1 Data Granularity Requirements and Data Use

<i>Data Type</i>	<i>Priority (H/M/L)</i>	<i>Aggregated/ Anonymized / Identifiable</i>	<i>Description / Additional Comments</i>
Monthly usage	H	Identifiable	Customer-specific monthly energy use. Must include service address to allow linkage with other data.
Monthly billing	H	Identifiable	Customer-specific monthly bill amount.
Efficiency program participation	H	Identifiable	Any customer-specific data such as program name, participation date, acquisition channel, action taken, incentive received, etc. along with general program details such as dates, promotion channels, incentives, participation rate, etc.
Retrofit data	M	Identifiable	Any customer-specific data such as retrofit date, type, incentive, acquisition channel, contractor name, etc.

3.2 Data Collection and Maintenance Requirements

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
Data must cover the last 6 months at a minimum.	H	A time series is critical to avoid biases from weather-related, seasonal, or random effects

Ideally data would cover the last 12 to 24 months.		and identify patterns that precede and follow energy efficiency program participation.
Data must be machine-readable and accompanied by a schema.	H	Any machine-readable format such as fixed-width text, delimited text, csv, xml, xls, etc. is acceptable. This is critical for efficient processing and error avoidance.
Data must be transferred and stored securely. Third party must not share data or use for any other purpose, and must delete data after use.	H	To protect PII and comply with rules and regulations.
Data should be updated quarterly.	M	New data will allow further refinements of analyses and keeps results up to date.

3.3 Required Policy & Other Determinations

<This section should outline high-level policy requirements, e.g. if there is a need to have CPUC approve release of data. Anything else should also go here.>

4 Current Data Obstacles and Other Issues

4.1 Barriers

<This section should summarize all the barriers that currently exist or are anticipated by the stakeholders.>

4.2 Outstanding Issues

<This section should summarize any issues or open questions that the team wasn't able to resolve.>

4.3 Additional Comments

<Anything that didn't fit anywhere else can go here.>

5 Conclusion

5.1 Conclusion

<Conclusions about this use case.>

5.2 Recommended Next Steps

<Proposed next steps.>

Appendix

Contact

Ian Hough – Brighter Planet Technology Services
ian@brighterplanet.com
1.802.458.0441 x 316

Robbie Adler – Brighter Planet Technology Services
robbie@brighterplanet.com
1.802.458.0441 x 306

Use Case 12 – Provided by DECA

1. Overview

1.1 Use Case Summary

The DECA use case provides the public with a working model of the majority of California's electricity grid, with a particular focus on the ability to model all electricity consumers' consumption at sub-hour time interval and to tie that data to actual weather conditions, building data, etc. The use case allows for the overlaying of wholesale market data including wholesale production run simulations providing prices and emissions. Expected users of this data are policy advocates, distributed generation providers, energy efficiency marketers and evaluators, and local governments.

1.2 Objectives

The purpose of the DECA use case is to provide the public with a tool for accurately measuring the potential for optimization of California's electricity grid from an integrated resource perspective. It allows for market development of geographically targeted energy efficiency, demand response, and distributed generation infrastructure, scientific assessment of rate impacts, and optimization of energy production and conservation based on transparent costs and benefits.

Ratepayer benefit from the DECA use case in a number of ways. First, billions of dollars in programs can be made more efficient by bringing transparency to the localized avoided costs from CPUC programs, rather than relying on statewide programs that treat the return on an investment as unrelated to geography. Second, by enabling third party actors to have better transparency regarding potential markets, ratepayers benefit by learning about how their electricity consumption can be most efficiently addressed. Third, clarity regarding program potential can be better integrated into CAISO backstopping and CPUC procurement planning to reduce the overall inefficiencies in the grid's operation by providing geographically accurate descriptions of program participation and vehicle for the quantification of program potential by geography.

1.3 Actors

<This section should describe the participants in this process. At a minimum, this should specify the data owner and the data requestor. This may end up being the same across all of the use- cases, but maybe different.>

Name	Role description
Utility Organization	All three IOUs provide usage data by customer including customer class and feeder line identification.
CPUC	CPUC can utilize these data internally as well see third party
Academic institution	The DECA use case provides incalculable opportunities for academic research.
3 rd party	Local government, 3 rd party service providers, environmental & policy

1.4 Applicable Statutes and Regulatory Rules

<This section should describe any specific rules or regulations that already apply to this use case, e.g. if there are requirements that stem from a specific CPUC mandated program.>

Agency	Description	Applies to
CPUC		
Other		

2. Use Case Details

2.1 Current Data Practices

There is no access to this data currently. Some aggregated forms of this data are available to entities like the CAISO subject to non-disclosure restrictions. The CPUC has a multi-year history of trying unsuccessfully to obtain these data directly from the utilities for procurement planning purposes.

2.1 Requested Data Practices

DECA believes that synthetic data and related obscuring techniques (such as randomized like for like substitution) can be used to prevent re-identification of customers within a reasonable probability. By this DECA means that the dataset

should include data that, while it may include PII, cannot be used to say with a known percentage of certainty/probability (e.g. less than 66%) to be actual PII vs. synthesized or randomized data.

DECA emphasizes that like for like substitution must be adjusted for geography at a sub climate zone level, socioeconomics, housing stock, and socioeconomic factors to ensure that the data remain useful.

3. High Level Requirements

3.1 Data Granularity Requirements and Data Use

<This section should summarize the type of data that we are talking about for this specific use case. Not every data element should be spelled out at this level – just the type/categories of data.>

<i>Data Type</i>	<i>Priority (H/M/L)</i>	<i>Aggregated/Anonymized/Identifiable</i>	<i>Description/Additional Comments</i>
5 to 15 minute interval usage data	H	Identifiable if not blurred	Data that's being recorded by a customer's meter at a 15 minute interval. Provided by day. dadavdayfor se

3.2 Data Collection and Maintenance Requirements

DECA supports the release of this data as datasets not subject to any maintenance other than server space and bandwidth. Data should be released quarterly or annually

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
Consumption information in kW averaged by increment for 5 to 15 minute increments for 12 months	H	All related data can be built from this level of granularity

3.3 Required Policy & Other Determinations

<This section should outline high-level policy requirements, e.g. if there is a need to have CPUC approve release of data. Anything else should also go here.>

<i>Requirement</i>	<i>Priority (H/M/L)</i>	<i>Additional Comments</i>
The CPUC shall approve the synthesis and randomization process as well as any redaction methodology	H	Redaction methodology should be limited to cases where other data sources may reveal very large customer data

4. Current Data Obstacles and Other Issues

4.1 Barriers

<This section should summarize all the barriers that currently exist or are anticipated by the stakeholders.>

<i>Barrier Description</i>	<i>Priority (H/M/L)</i>	<i>Current/Anticipated</i>

4.2 Outstanding Issues

<This section should summarize any issues or open questions that the team wasn't able to resolve.>

<i>Description</i>	<i>Proposed Next Step, if any</i>
e.g. there wasn't enough information about how XYZ is being done today	

4.3 Additional Comments

<Anything that didn't fit anywhere else can go here.>

5. Conclusion

5.1 Conclusion

The DECA use case operates on the assumption that the probability of electricity usage at a given location is a public good and that PII is not revealed if the data released includes actual usage data that has a sufficiently low probability of being the actual usage data.

5.2 Recommended Next Steps

DECA recommends development of metrics by which probability can be reduced to one standard deviation and that computer scientists, statisticians, and data users develop a methodology for defining like for like substitutions.

Appendix

Contact

<May want to include the list of people who participated in the development of the use case or who to contact with questions.>

Reference Materials

<Reference Materials.>

APPENDIX B

**Appendix B - “Legal Considerations for Smart Grid Energy Data Sharing,”
EFF and the Samuelson Law, Technology & Public Policy Clinic at the
University of California, Berkeley, School of Law, April 1, 2013**

**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)
Phase III Energy Data Center

M E M O R A N D U M

To: Participants of Working Group organized pursuant to Administrative Law Judge's Ruling Setting Schedule To Establish "Data Use Cases," Timelines For Provision Of Data, And Model Non-Disclosure Agreements, from Rulemaking Proceeding No. 08-12-009

From: Electronic Frontier Foundation and the Samuelson Law, Technology & Public Policy Clinic at the University of California, Berkeley, School of Law

Date: April 1, 2013

Re: Legal Considerations for Smart Grid Energy Data Sharing

INTRODUCTION

This memorandum is one of two memoranda offered by the Electronic Frontier Foundation (EFF) and the Samuelson Law, Technology & Public Policy Clinic at the University of California, Berkeley, School of Law to aid in the parties' discussions during the Working Group meetings outlined in Judge Sullivan's February 27, 2013 ruling, titled *Administrative Law Judge's Ruling Setting Schedule to Establish "Data Use Cases," Timelines for Provision of Data, and Model Non-Disclosure Agreements* ("Ruling").

This memorandum covers legal background relevant to this proceeding, providing a brief explanation of important laws that apply to energy usage data sharing, as well as a brief background of the legal landscape covered in the proceeding to date. The other memorandum, titled *Technical Issues with Anonymization & Aggregation of Detailed Energy Usage Data as Methods for Protecting Customer Privacy*, offers some technical background on aggregation and

anonymization models for protecting privacy.

The proceeding thus far has established both basic principles and a targeted legal framework—in the form of the Rules Regarding Privacy and Security Protections for Energy Usage Data (“Privacy Rules”),¹ adopted by the California Public Utilities Commission (“Commission”) in D. 11-07-056 (“2011 Decision”)² and set forth in Attachment D to that Decision—for managing customer data collected by smart meters. In 2012 the Privacy Rules were extended to customers of gas corporations, community choice aggregators, as well as residential and small commercial customers of electric service providers.³ It now presents an opportunity to apply this framework in establishing effective, secure protocols for more streamlined access to the rich and highly sensitive information captured by smart meters.

Following the Ruling, the Working Group is expected to discuss definitions of “aggregate” and “anonymous” data, as well as standards for achieving optimal aggregation or anonymization and reasonable protocols for sharing those categories of data. In order to fulfill these goals, Working Group participants must have the legal landscape on which we are operating firmly in hand. Further, understanding the legal contours of smart grid data sharing will enable more productive discussions of the validity and/or scope of the proposed “use cases” set out in the Ruling.

DISCUSSION

During this proceeding, the Commission has established that smart grid data can reveal a great deal of private information about life inside a premises, including: how many inhabitants are home or away at a given time; when those inhabitants go to bed, wake up, take showers, or cook dinner; and what devices inhabitants use, including personal medical devices.⁴ Known privacy and security risks include, among others:

¹ *Rules Regarding Privacy and Security Protections for Energy Usage Data*, in *Attachment D*, Decision Adopting Rules to Protect The Privacy And Security of the Electricity Usage Data of the Customers of Pacific Gas & Electric Company, Southern California Edison Company, And San Diego Gas & Electric Company, Rulemaking 08-12-009 (July 29, 2011) [“Privacy Rules”].

² Decision Adopting Rules to Protect The Privacy And Security of the Electricity Usage Data of the Customers of Pacific Gas & Electric Company, Southern California Edison Company, And San Diego Gas & Electric Company, Rulemaking 08-12-009 (July 29, 2011) [“2011 Decision”].

³ D. 12-08-045 (August 23, 2012).

⁴ See Statement from Martin Pollock of Siemens Energy, in Gerard Wynn, *Privacy Concerns Challenge Smart Grid Rollout*, REUTERS, June 25, 2010, available at: <http://uk.reuters.com/article/idUKTRE65O1RQ20100625>. See also

- Data breach (hacking) or data leaks (inadvertent disclosure to the public);
- Re-identification of aggregated and/or anonymized data to reveal personally-identifying information; and
- “Mission creep,” the potential future expansion of access to energy usage data to include additional users or uses of the data beyond what was initially contemplated (e.g., for law enforcement).

This proceeding has also already established the applicability of a variety of laws intended to protect Californians’ data privacy interests. Many of these laws are already discussed in the 2011 Decision and are reflected in the Privacy Rules. In the Privacy Rules phase of the proceeding and in his presentation at the January 15th Workshop, Chris Warner of Pacific Gas & Electric provided a list of the laws and regulations relevant to the collection, maintenance, use, and disclosure of smart grid data.⁵ Additionally, in its Opening Comment on the Proposed Energy Data Center (“EDC”), EFF raised questions regarding the applicability of existing state law, including the Information Practices Act of 1977 (“IPA”),⁶ to EDC proposals. Parties participating in the January 15th and 16th Workshops identified as the IPA as a relevant topic for further review.⁷

To aid this phase of the proceeding, this memorandum further discusses some of these laws as applied to the disclosure of customer energy usage data. Specifically, it briefly reviews the California Constitution, the Fair Information Practices Principles (“FIPPs”), and Public Utilities Code Section 8380 (commonly referred to as “SB 1476”) as important foundations for the Privacy Rules. It then provides further review of the IPA and its applicability to agency sharing of energy usage data. Finally, the memorandum reviews for the Working Groups the key provisions of the Privacy Rules themselves, which implement SB 1476, other relevant law, and the FIPPs for smart meter data. With a foundational understanding of these laws, the Working Groups will be better equipped to devise solutions for smart grid data sharing that comply with these existing laws.

Mikhail A. Lisovich, Deirdre K. Mulligan & Stephen B. Wicker, *Inferring Personal Information from Demand-Response Systems*, IEEE SECURITY & PRIVACY (Jan.–Feb. 2010).

⁵ *Appendix A: List of Current Statutes, Regulations, Decisions and Protocols Related to Customer Privacy Applicable to California Energy Utilities*, Attachment B from Ruling D. 11-07-056; Slide presentation by Christopher J. Warner, *Existing Energy Data Sharing Protocols: A Potential Consensus Approach*, CPUC Workshop (Jan. 15, 2013), available at ftp://ftp.cpuc.ca.gov/13011516_EgyDataWorkshop/.

⁶ Opening Comments of the Electronic Frontier Foundation, at 10–11 (Dec. 17, 2012) [hereinafter EFF Opening Comment].

⁷ Slide presentation by Christopher J. Warner, *Existing Energy Data Sharing Protocols: A Potential Consensus Approach*, CPUC Workshop (Jan. 15, 2013), available at ftp://ftp.cpuc.ca.gov/13011516_EgyDataWorkshop/.

Before commencing the Working Groups, participants should understand that these laws require us to propose definitions and implement “use case” solutions that are dynamic and adaptable. This is because the legal landscape governing data sharing varies—and can change dramatically—depending on a number of factors: (1) the identity of the data custodian; (2) the identity of the data requester; (3) the purpose of the data disclosure; and (4) the level of granularity of the data requested. The proposed use cases represent different permutations of these variables, so the law necessarily treats them differently. Understanding the legal obligations that attach to each data-sharing scenario will enable more accurate evaluation and more effective problem-solving.

A. California Law

1. The California Constitution

Article I, Section 1 of the California Constitution recognizes each individual’s right to privacy. There is general agreement among the judicial, scholarly, legislative, and regulatory communities that the data collected by smart meters reveals intimate details about the lives of California citizens. As such, the California Constitution establishes a baseline obligation to protect energy usage data from harmful disclosure or use.

The same interests that motivated California citizens to enact Section 1 by ballot amendment in 1972 still apply today: (1) the overbroad collection and retention of unnecessary personal information by government and business interests; and (2) 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.⁸

Representative of the high value the California public places on privacy, the California Constitution imposes an obligation to protect consumer privacy on all parties—including private parties—engaging in smart grid data sharing. As such, addressing privacy issues are necessarily central to this proceeding, and Working Group participants should bear in mind adequate protections against unauthorized use or disclosure of personal information when addressing definitions and use cases.

/

⁸ *White v. Davis*, 13 Cal. 3d 757, 775 (1975).

2. *Information Practices Act*

The IPA (California Civil Code section 1798 *et seq.*) governs the manner in which state agencies, as defined in the IPA, disclose personally identifiable data that they collect and maintain. The statute applies to state-wide agencies, including the Commission and the California Energy Commission (CEC).⁹ Should the Commission designate one of these agencies as a custodian of smart grid data, the IPA will apply to that agency's disclosure of the data.

The IPA protects energy usage data that “identifies or describes an individual”—in this context, an individual utility customer.¹⁰ The IPA offers a non-exhaustive list of example types of “personal information” that might be used to identify or describe an individual, including an individual's “name, social security number, physical description, home address, home telephone number, education, financial matters, and medical or employment history.”¹¹ At the January Workshop, Professor Ashwin Machanavajjhala asserted that additional types of information, such as sex, birthdate, and zip code, operate as “quasi-identifiers,” capable of re-identifying an individual when linked to other available data. The IPA's open-ended list of identifiers would include that information as well.

As a general rule, state agencies are not permitted to disclose any personal information “in a manner that would link the information disclosed to the individual to whom it pertains.”¹² However, a number of exceptions apply, subject to varying protocols and approval procedures depending on the data recipient. For example, Section 1798.24 authorizes disclosure of an individual's personal data in the following pertinent scenarios, among others:

- With the prior written voluntary consent of the individual, Cal. Civ. Code § 1798.24(b);
- To persons, or another state agency, such as the CEC, for whom the information is necessary to fulfill statutory duties, Cal. Civ. Code § 1798.24(e);
- Where the CPUC is required by law to disclose the information to a local government (or federal government) entity,¹³ Cal. Civ. Code § 1798.24(f);
- Disclosure to a researcher, if (1) he provides assurance that the information will be used solely for statistical research or reporting purposes, and (2) he does not

⁹ Cal. Civ. Code § 1798.3.

¹⁰ Cal. Civ. Code § 1798.3(a).

¹¹ The IPA also includes “statements made by, or attributed to, the individual” within its list of identifiers. Cal. Civ. Code § 1798.3(a).

¹² Cal. Civ. Code § 1798.24.

¹³ We note that there are two separate exceptions relating to warrant and subpoena requirements.

receive the information in a form that will identify the individual, Cal. Civ. Code § 1798.24(h); and

- Disclosure to a researcher within the University of California system, provided that the request is approved by the Committee for the Protection of Human Subjects, Cal. Civ. Code § 1798.24(t).

Of particular relevance to Working Group discussion is Section 1798.24(h), which specifically addresses disclosure for research purposes. This provision underscores the California legislature's commitment to protecting the privacy of the individual(s) to whom the data pertains by explicitly limiting disclosure of personally identifiable information to researchers, while allowing research. We additionally note that Section 1798.24(e) also practically limits the scope of agency disclosures to only those specifically and directly authorized by statute, lest the exception swallow the rule.

One of the fundamental privacy concerns motivating the enactment of the IPA was the risk of data breach, a problem that is prevalent and well-documented among all institutions, including California institutions. An important obligation the IPA imposes on third party data recipients working within the University of California system is that requests for disclosure of personal information must first be approved by the Committee for the Protection of Human Subjects (CPHS), or another institutional review board that has written authorization from the CPHS. Although Section 1798(t) appeared in the original 1977 version of the statute, the specific language requiring approval from the CPHS was added in 2005 to ensure that the UC satisfies minimum standards for data security.¹⁴

This amendment responds to a high-profile computer hacking incident and data breach that occurred in August 2004, in which a UC Berkeley researcher inadvertently disclosed names, addresses, social security numbers, birthdates, and phone numbers for nearly 1.3 million people residing in California.¹⁵ Data breaches continue to plague the UC system, giving credence to the state legislature's concern about security protocols at public research institutions. For example, in December 2006, UCLA alerted approximately 800,000 current and former students, faculty,

¹⁴ See Stats. 2005, c. 241 (S.B. 13) § 1 (“The Legislature recognizes the research community has legitimate needs to access personal information to carry out research . . . the provisions of this bill are not intended to impede research but rather to require and set minimum standards for careful review and approval of requests.”).

¹⁵ EFF Opening Comment, at 11. See also Senate Bill Analysis, Third Reading, Stats. 2005, c. 241 (S.B. 13) (Aug. 17, 2005). In that case, the researcher requested data from the Department of Social Services (DSS) about participants in the In-Home Supportive Services (IHSS). Although the researcher needed only a random sample of IHSS data, the DSS made the entire IHSS database available for download. Shortly thereafter, a hacker broke into the researcher's computer system, causing a massive data breach.

and staff that a sophisticated computer hacker had broken into its systems and accessed a restricted database containing their personal information.¹⁶ More recently, in 2011, the UCLA Health System notified over 16,000 patients that their names, birthdates, addresses, and medical information had been stolen during the burglary of a physician's home.¹⁷ Although the physician had stored the data on an encrypted external hard drive, the password for the hard drive was written on a piece of paper kept near the computer that was found missing after the incident.

As such, the IPA provides both legal requirements binding on relevant agencies and overall guidance as to how California has thus far approached data risks for California citizens. Accordingly, although the IPA is not binding on utility companies, academic or local government researchers, or other parties who cannot be characterized as state agencies, it nevertheless provides useful guidance in this situation because it approximates how California law might treat the disclosure of energy usage data more generally.

B. The Privacy Rules

In the smart grid context, statewide concern in California with consumer privacy has culminated in the Commission's adoption of the Privacy Rules, which specifically address the sharing of energy usage data held by investor-owned utilities ("IOUs"). The Privacy Rules most directly address the type of data sharing at issue in this phase of the proceeding: (1) they specifically regulate energy usage data collected by smart meters, and (2) they concern disclosure by the IOUs to third party data requesters. As such, they provide the governing general authority on energy usage data sharing by the IOUs.

Accordingly, the Privacy Rules are the primary source of legal guidance as the Working Groups determine how to manage any disclosure of such data, and comprise the central feature of our discussion on relevant law. Part 1 of this section provides a brief background to the Privacy Rules, adopted in 2011, and their implementation of the provisions of SB 1476 and the FIPPs. This background provides a fuller understanding of the Privacy Rules for those participants not previously involved in the proceeding. Part 2 explains the standards and requirements for disclosure of covered information set forth in the Privacy Rules.

¹⁶ *UCLA Warns of Unauthorized Access to Restricted Database*, UCLA NEWSROOM (Dec. 12, 2006), <http://newsroom.ucla.edu/portal/ucla/UCLA-Warns-of-Unauthorized-Access-7571.aspx?RelNum=7571>.

¹⁷ *UCLA Medical Officials Say Patient Information Data Stolen*, L.A. TIMES BLOG (Nov. 4, 2011), <http://latimesblogs.latimes.com/lanow/2011/11/ucla-patient-identification-stolen.html>.

1. Brief Background to the Privacy Rules: SB 1476 and the FIPPs

In 2010 the California legislature passed **SB 1476**, now codified as Public Utilities Code Section 8380, to regulate the use and disclosure of utility customer data collected by smart meters. SB 1476 applies both to “electrical corporations and gas corporations.” Subject to some exceptions, SB 1476 generally prohibits disclosure of “electrical or gas consumption data . . . available as part of an advanced metering infrastructure, [including] the name, account number, or residence of the customer.”¹⁸ Under Section 8380 (b)(1) “an electrical corporation or gas corporation shall not share, disclose, or otherwise make accessible to any third party a customer’s electrical or gas consumption data, except as provided in subdivision (e) or upon the consent of the customer.” The Privacy Rules implement these restrictions and their exceptions with regard to the IOUs.

In addition to implementing the requirements of SB 1476, the Commission established that the sharing of energy usage data should follow **Fair Information Practice Principles** (FIPPs), a widely accepted international framework for handling electronic information in a privacy-protective manner. In the 2011 Decision, the Commission explicitly adopted the FIPPs as California’s policy for smart grid privacy. Thus, the foundational principles set forth in the FIPPs provide guidance to the Working Groups as participants determine how to most effectively implement the Privacy Rules.

The eight principles embodied in the FIPPs can inform privacy discussions in the upcoming Working Groups in a number of ways. For example:

1. *Transparency*: Any new repository of data that is separate from the IOUs would make it more difficult to provide notice to individual utility customers about the use or dissemination of their personal information
2. *Individual Participation*: The Commission should continue to use consent measures to involve individual utility customers in processes for data collection, use, dissemination and maintenance. Unlike typical consumers, many utility customers have no choice when buying energy. As a result, foregoing consent for disclosure is not bargained for in the relationship with the utility.
3. *Purpose Specification*: Requesting parties must be required to specify the purpose underlying the request prior to authorization for disclosure.
4. *Data Minimization*: Only the data actually necessary for the particular purpose identified should be disclosed. The FIPPs’ minimization principle helps in developing

¹⁸ Pub. Util. Code § 8380(a).

data handling practices that limit data breach and other risks before they happen, and helps data handlers decide on data needs in an efficient manner.

5. *Use Limitation*: There must be mechanisms to ensure that the disclosure of information is used solely for the specified purpose(s).
6. *Data Quality and Integrity*: If multiple parties were permitted to collect and store energy usage data, it would be harder to ensure that the data is accurate, relevant, timely, and complete. The problems associated with one data set may be multiplied across parallel data sets.
7. *Security*: Any data collected from the IOUs and stored pursuant to security protocols that are less rigorous than those utilized by the IOUs may be susceptible to loss, unauthorized access, destruction, modification, or unintended disclosure.
8. *Accountability and Auditing*: Mechanisms are already in place to enforce IOUs compliance with the FIPPs. It will be of utmost importance during the Working Groups to ensure that any other entity collecting and maintaining smart grid data be accountable for customer privacy in the same manner.

Both the FIPPs and SB 1476 were at the forefront when the Commission ultimately decided to adopt the Privacy Rules.

2. Privacy Rules, adopted in D. 11-07-056 (Attachment D)

Recognizing the need to more directly operationalize the FIPPs and the requirements of SB 1476 to protect consumer privacy in smart meter data,¹⁹ the Commission adopted the Privacy Rules, which regulate the disclosure of energy usage data by IOUs. As noted above, last year the Privacy Rules were extended to cover gas utilizes, community choice aggregators, electric service providers, and other “load serving” entities.²⁰ The Privacy Rules determine the extent to which an IOU may disclose energy usage data to third parties, depending on the purpose for which the data will be used. It covers all energy usage data captured by smart meters that, “when associated with any information . . . can reasonably be used to identify an individual [utility customer]”²¹ Data that cannot reasonably be re-identified are excluded from the Privacy Rules.²²

¹⁹ 2011 Decision, at 19–21.

²⁰ D. 12-08-045 (August 23, 2012).

²¹ The exact language of the Privacy Rules reads:

“Covered information” does not include usage information from which identifying information has been removed such that an individual, family, household or residence, or nonresidential customer cannot reasonably be identified or re-identified. Covered information, however, does not include information provided to the Commission pursuant to its oversight responsibilities.

The Privacy Rules categorize various potential uses into two categories. “Primary purposes” are uses of the data that directly serve utility operations, are specifically authorized by the utility company or the Commission in connection with an energy-related program, or are for services required by state or federal law. “Secondary purposes,” cover all other uses. Each category comes with its own list of obligations and security protocols relating to data transfer. The Rules impose these obligations on both the IOU disclosing the data and the third party recipients of the data.²³

a. Primary Purpose

Under the Privacy Rules, a covered entity may only disclose covered information without customer consent if the data will be used for a “primary purpose.” The Privacy Rules identify four limited purposes that fit within this category:

- (1) [to] provide or bill for electrical power or gas,
- (2) [to] provide for system, grid, or operational needs,
- (3) [to] provide services as required by state or federal law or as specifically authorized by an order of the Commission, or
- (4) [to] plan, implement, or evaluate demand response, energy management, or energy efficiency programs under contract with an electrical corporation, under contract with the Commission, or as part of a Commission authorized program conducted by a governmental entity under the supervision of the Commission.²⁴

Privacy Rules § 1(b). Further, for the purposes of “analysis, reporting or program management,” disclosure of “aggregated usage data that is removed of all personally-identifiable information” is permissible, “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.” Privacy Rules § 6(g).

²² As explained in our accompanying memo titled *Technical Issues with Anonymization & Aggregation of Detailed Energy Usage Data as Methods for Protecting Customer Privacy*, which covers recent scientific advancements in re-identification, no level of basic anonymization and aggregation provides a guarantee against re-identification. The Commission should pursue more robust solutions.

²³ The Privacy Rules govern “covered entities,” a category that includes:

- (1) [A]ny electrical corporation, or any third party that provides services to an electrical corporation under contract, (2) any third party who accesses, collects, stores, uses or discloses covered information pursuant to an order of the Commission, unless specifically exempted, who obtains this information from an electrical corporation, or (3) any third party, when authorized by the customer, that accesses, collects, stores, uses, or discloses covered information relating to 11 or more customers who obtains this information from an electrical corporation.

Privacy Rules § 1(a). The Commission’s authority to create regulations binding on third parties derives from the language of SB 1476, which conferred upon the Commission “broad powers and a legislative mandate” to take regulatory action to protect consumer interests. 2011 Decision, at 33–35.

²⁴ Privacy Rules § 1(c).

Section 6(b) further clarifies which entities may access, collect, store and use covered information for primary purposes without customer consent:

- An electrical corporation
- A third party acting under contract with the Commission to provide energy efficiency or energy efficiency evaluation services authorized pursuant to an order or resolution of the Commission
- A governmental entity providing energy efficiency or energy efficiency evaluation services pursuant to an order or resolution of the Commission.²⁵

According to the 2011 Decision, “[t]o the extent other governmental organizations, such as the California Energy Commission or local governments, may seek Covered Information in a manner not provided in these rules, the Commission will determine such access in the context of the program for which information is being sought absent specific Legislative direction.”²⁶

Accordingly, where the Privacy Rules do not explicitly provide for a certain form of disclosure, the Commission will determine on a case-by-case basis whether the disclosure is appropriate, and whether it is permissible under relevant legislation, such as the IPA. Please see above for more information about the IPA.

Sections 6(c)(1)(a–b) provides additional insight as to what qualifies as a “primary purpose,” and how disclosures must be carried out. Under these provisions, an IOU may share covered information with a third party without customer consent (a) if “explicitly ordered to do so by the Commission” or (b) if the disclosure serves “a primary purpose being carried out under contract with and on behalf of the electrical corporation disclosing the data.”²⁷ These provisions indicate that the Commission intended for the “primary purpose” category to cover a fairly narrow selection of disclosure scenarios, largely directed to IOU operations (such as billing, maintenance, and the like by contractors), along with the noted services, when under direct Commission oversight.

“Primary purpose” disclosures create a chain of obligations that carry down to subsequent custodians of “covered information.” When disclosure occurs for a “primary purpose,” the covered entity disclosing the data “shall, by contract, require the third party to agree to access, collect, store, use, and disclose the covered information under policies, practices and notification requirements no less protective than those under which the covered entity itself operates as

²⁵ Privacy Rules § 6(b).

²⁶ See 2011 Decision at 47-48.

²⁷ Privacy Rules §§ 6(c)(1)(a–b).

required under this Rule, unless otherwise directed by the Commission.” Thus, a “primary purpose” recipient of covered information must employ at least the same privacy and security measures as those implemented within the IOU from which it collected the data. The Privacy Rules attach to all data that originates with the IOUs, regardless as to whom ultimately takes possession of it.²⁸

b. Secondary Purpose

Any purpose that does not fall within one of the above categories is considered a “secondary purpose” under the Privacy Rules.²⁹ IOUs are prohibited from disclosing covered information for any secondary purpose without the “prior, express, written authorization” of each utility customer represented in the data.

Three limited exceptions to this requirement exist. A covered entity may only disclose smart grid data without customer consent in the following situations: (1) disclosure pursuant to a certain types of legal process (such as a warrant or court order); (2) disclosure in “situations of imminent threat to life or property; and (3) disclosure “authorized by the Commission pursuant to its jurisdiction and control.”³⁰ Again, without an authorization order from the Commission, third parties not working on behalf of the utility company likely cannot obtain covered information without the prior, express, written authorization from utility customers.

c. Data Minimization Requirements

Under Section 5(c), covered entities must limit the disclosure of smart grid data to only that which is “reasonably necessary or as authorized by the Commission” to carry out the specific purpose permitted under the Privacy Rules. For data uses constituting “secondary purposes,” this means that the covered entity may not disclose more information than is

²⁸ Privacy Rules § 6(c)(1). Rule 6(c)(2) reinforces the recursive nature of the Privacy Rules:

Any entity that receives covered information derived initially from a covered entity may disclose such covered information to another entity without customer consent for a primary purpose, provided that the entity disclosing the covered information shall, by contract, require the entity receiving the covered information to use the covered information only for such primary purpose and to agree to store, use, and disclose the covered information under policies, practices and notification requirements no less protective than those under which the covered entity from which the covered information was initially derived operates as required by this rule, unless otherwise directed by the Commission.

Privacy Rules § 6(c)(2).

²⁹ Privacy Rules § 1(e).

³⁰ Privacy Rules §§ 6(d)(1–3).

reasonably necessary to carry out the specific purpose authorized by the customer in writing. As noted above, data minimization requires entities to consider, in advance of disclosure, what data is reasonably necessary for the agreed-upon purpose before disclosing the data.

d. Data Security and Breaches

Section 8 of the Privacy Rule establishes the minimum security requirements that covered entities must employ when in possession of covered information. “Covered entities shall implement reasonable administrative, technical, and physical safeguards to protect covered information from unauthorized access, destruction, use, modification, or disclosure.”³¹ Furthermore, when a breach has been detected, a covered third party must notify the disclosing IOU within one week, and the utility must notify the Commission of all breaches affecting one thousand or more customers.³² Utility companies are additionally obligated to file an annual report at the end of the each calendar year, chronicling all security breaches affecting covered information that year.

e. Enforcement and Recourse for Privacy Rule Violations

If a recipient party fails to comply with its contractual obligations to handle the covered information in a manner “no less protective” than those under which the originating entity operates—a “material breach” under the Privacy Rule—“the disclosing entity shall promptly cease disclosing covered information to such third party.”³³

CONCLUSION

The laws and regulations described above each bear heavily on the data sharing scenarios contemplated within this proceeding. As such, it will be important for participants to enter the Working Group discussions with a firm understanding of their relevant provisions, with the Privacy Rules front and center.

Among the California state Constitution, the IPA, the FIPPs, SB 1476, and the Privacy Rules, utility customers receive legal protections for the privacy of their energy usage data.

³¹ Privacy Rules § 8(a).

³² Privacy Rules § 8(b). The Commission may also request that the utility company provide notification of any other breach for which notification is not already compulsory.

³³ Privacy Rules § 6(c)(3).

These protections, in various ways, bind the IOUs, the Commission, and other state agencies handling smart meter data, as well as third parties who obtain energy usage data from the utilities. At this stage of the proceeding, keeping these laws and regulations in mind will better position the Working Groups to devise solutions that are appropriately tailored to each disclosure scenario and are consistent with applicable law.

Respectfully submitted this April 1, 2013 at San Francisco, California.

/s/ Jennifer Urban _____

JENNIFER URBAN, Attorney
Samuelson Law, Technology & Public Policy Clinic
University of California, Berkeley School of Law
396 Simon Hall
Berkeley, CA 94720-7200
(510) 642-7338
Attorney for ELECTRONIC FRONTIER
FOUNDATION

/s/ Lee Tien _____

LEE TIEN, Attorney
Electronic Frontier Foundation
454 Shotwell Street
San Francisco, CA 94110
(415) 436-9333 x102
Attorney for ELECTRONIC
FRONTIER FOUNDATION

APPENDIX C

**Appendix C - "Technical Issues with Anonymization & Aggregation of Detailed Energy Usage Data as Methods for Protecting Customer Privacy,"
Electronic Frontier Foundation and the Samuelson Law, Technology & Public Policy Clinic at the University of California, Berkeley, School of Law
April 1, 2013.**

**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)
Phase III Energy Data Center

M E M O R A N D U M

To: Participants of Working Group organized pursuant to Administrative Law Judge’s Ruling Setting Schedule To Establish “Data Use Cases,” Timelines For Provision Of Data, And Model Non Disclosure Agreements, from Rulemaking Proceeding No. 08-12-009

From: Electronic Frontier Foundation and the Samuelson Law, Technology & Public Policy Clinic at the University of California, Berkeley, School of Law

Date: April 1, 2013

Re: Technical Issues with Anonymization & Aggregation of Detailed Energy Usage Data as Methods for Protecting Customer Privacy

INTRODUCTION

This memorandum is one of two memoranda offered by the Electronic Frontier Foundation (EFF) and the Samuelson Law, Technology & Public Policy Clinic at the University of California, Berkeley, School of Law to aid in Working Group discussions outlined in Judge Sullivan’s February 27, 2013, titled *Administrative Law Judge’s Ruling Setting Schedule to Establish “Data Use Cases,” Timelines for Provision of Data, and Model Non-Disclosure Agreements*, No. 08-12-009 (“Ruling”). This memorandum addresses the technical issues surrounding aggregation and anonymization of customer data. The other memorandum covers particular privacy rules and laws that apply to the disclosure of energy consumption data.

Thus far, this proceeding has established basic principles and a targeted framework—in the form of the Rules Regarding Privacy and Security Protections for Energy Usage Data

(“Privacy Rules”),¹ adopted by the California Public Utilities Commission (“Commission”) in D. 11-07-056 (“2011 Decision”)² and set forth in Attachment D to that Decision—for managing customer data collected by smart meters. This proceeding has already established the serious implications for privacy in the home that come from releasing customer energy consumption data.³ Accordingly, the Privacy Rules adopted by the Commission govern the release of “covered information:” customer usage data that can identify the customer or be re-identified after some identifying information has been removed. The Privacy Rules are discussed in further detail in our companion memo *Legal Considerations for Smart Grid Energy Data Sharing* regarding applicable law.

In this next phase, the proceeding aims to implement the Privacy Rules and other relevant legal requirements, in part by devising effective, secure protocols for manipulating customer energy data so that it can be shared with third parties without unduly compromising customer privacy. We offer this memorandum to help the Working Group understand the practical realities of known aggregation and anonymization techniques in light of computer science research demonstrating the characteristics of these techniques in protecting customer privacy, including their limitations. We also explain the need to involve technical experts working in the fields of data privacy and re-identification in order to develop protocols that effectively protect customer privacy and provide useful data to researchers.

This phase of the proceeding has thus far focused its attention on protecting privacy through anonymization and aggregation techniques. Unfortunately, a known set of technical problems that come with these techniques can make them highly vulnerable to re-identification of individual households or ratepayers included in the data set. While the terms “anonymization” and “aggregation” have not yet been clearly defined in the proceeding,⁴ individual methods that have been discussed—including the “15/15 Guideline,” zip code aggregation, and census-tract aggregation—are all vulnerable to these threats.

¹ *Rules Regarding Privacy and Security Protections for Energy Usage Data*, in *Attachment D*, Decision Adopting Rules to Protect The Privacy And Security of the Electricity Usage Data of the Customers of Pacific Gas & Electric Company, Southern California Edison Company, And San Diego Gas & Electric Company, Rulemaking 08-12-009 (July 29, 2011) [hereinafter Privacy Rules].

² Decision Adopting Rules to Protect The Privacy And Security of the Electricity Usage Data of the Customers of Pacific Gas & Electric Company, Southern California Edison Company, And San Diego Gas & Electric Company, Rulemaking 08-12-009 (July 29, 2011) [hereinafter 2011 Decision].

³ Decision Adopting Rules To Protect The Privacy And Security Of The Electricity Usage Data Of The Customers Of Pacific Gas And Electric Company, Southern California Edison Company, And San Diego Gas & Electric Company. D. 11-07-056.

⁴ See Ruling No. 08-12-009 at section titled “Definitions.”

The first Working Group is expected to discuss various threshold definitions, including definitions for “aggregate” and “anonymous” data. The Working Group has also been charged with proposing standards for data anonymization and aggregation that “ensure the anonymity of data, protect customer privacy, and prevent the reverse engineering of the aggregated data.”

In order to effectively engage with these tasks, Working Group participants first need to consider existing and ongoing research in the computer science community. To help with this task, we have consulted with technical experts in the field, and requested analysis from them. As part of this analysis, we are pleased to attach as Appendix A to this memorandum a paper titled *Privacy Technology Options for Protecting and Processing Utility Readings*, written as background for the Working Groups by computer security and privacy expert George Danezis. Unfortunately, analysis of the existing research demonstrates that existing techniques for anonymization or aggregation of data, taken alone, are insufficient protections for customer privacy. Anonymizing data (removing identifiers) and aggregating data (processing data and releasing only sums or patterns) have proven inadequate for protecting customer privacy because attackers and researchers can manipulate these data sets to re-identify individuals. As the Privacy Rules explicitly limit the release of data that can be re-identified, these proven workarounds must be taken into account when deciding what protocols to put in place for protecting customer privacy.

Accordingly, to devise the appropriate measures for protecting customer privacy without the risk of data re-identification, we believe that it is critical for the Working Groups to consult technical experts to help develop more robust solutions, beyond mere aggregation and anonymization (see, for example, the suggestions under “Robust Privacy Technology Options” in Appendix A). More robust solutions will help to prevent re-identification of “covered information,” as required by the Privacy Rules, and to provide researchers with useful data that contributes to valuable energy research.

DISCUSSION

A. Disclosure of the Detailed Customer Energy Consumption Data Collected from Smart Meters Creates Serious Risks to Customer Privacy.

Since the late 1980s, scientists have reported the ability to derive detailed behavioral information about a household or other premise from electrical meter readings.⁵ For example, Non-intrusive Appliance Load Monitoring (NALM) “use[d] temporally granular energy consumption data to reveal usage patterns for individual appliances in the house.”⁶ These usage patterns revealed, for example, time away from one’s home, cooking and sleeping habits, or the number of inhabitants in a particular household. Not long after its development in 1989, scientists described this technology as capable of remotely identifying patterns based on externally available meter information. In a 1989 paper, NALM creator George Hart simultaneously noted that identifying these patterns created the potential for invasions of private information.⁷ By tracking the daily energy usage of a household, it is possible to create a consumption profile and deduce behavior for that household.⁸ It exposes not only energy consumption patterns overall, but also intimate behavioral information that most customers would not suspect is being shared, including travel, sleeping, and eating patterns, occupational trends, and even detailed information such as when children are home alone.⁹ This type of profiling is attractive for a number of purposes, from behavioral research to marketing. For an example of such consumption profiling used in the retail industry, Target Corporation used data on women’s shopping habits to develop a pregnancy detection method so reliable that it often

⁵ According to one employee of Siemens Energy:

We, Siemens, have the technology to record [energy consumption] every minute, second, microsecond, more or less live. From that we can infer how many people are in the house, what they do, whether they're upstairs, downstairs, do you have a dog, when do you habitually get up, when did you get up this morning, when do you have a shower: masses of private data.

Quote from Martin Pollock of Siemens Energy in Gerard Wynn, “Privacy Concerns Challenge Smart Grid Rollout” *Reuters*, June 25, 2010, *available at*: <http://uk.reuters.com/article/idUKTRE65O1RQ20100625>.

⁶ Jennifer M. Urban, *Privacy Issues in Smart Grid Deployment*, at 6-7, in SMART GRID AND PRIVACY (forthcoming 2013).

⁷ Hart, George W. (1989), ‘Residential Energy Monitoring and Computerized Surveillance via Utility Power Flows’, *IEEE Technology and Society Magazine*, 8 (2), 12-16 at 13; F. Sultanem (1991), “Using Appliance Signatures for Monitoring Residential Loads at Meter Panel Level,” *IEEE Transactions on Power Delivery*, 6 (4), 1380, 1381, col. 2 (showing load graphs of various appliances and a fluorescent light). The reader can find a lay introduction to NALM technology in Quinn, Elias L. (2009) ‘Privacy and the New Energy Infrastructure’, *Social Science Research Network*, 09 at 21-25.

⁸ D. 11-07-056.

⁹ *Id.*; See also, Presentation of Chris Vera at January 15 workshop (slides available at ftp://ftp.cpsc.ca.gov/13011516_EgyDataWorkshop).

allowed for targeted advertisements before a woman had even revealed her pregnancy to others.¹⁰ Similar predictive algorithms can be used to extend noticeable trends in energy consumption data, such as using real-time data to determine when an occupant is at home for solicitation by the utility or some third party. To continue with family formation as an example, an occupant's consumption profile might indicate a new baby in the house. This would violate the home occupants' privacy and create risks of leaking personal information that the customer had not even considered exposed in the first place.¹¹

Working Groups will need to consider both existing profiling capabilities and those that are likely to arise in the near future. More recent scientific research on techniques for ascertaining information from energy data describes the developing ability to discern what video content is being viewed on a television or computer monitor. Known as "use-mode detection," this method relies on collecting energy data in real time. Lab scientists tested multiple television sets to determine that the content viewed on those devices left uniquely identifying energy signatures, known as electro-magnetic interference (EMI). The same video content would produce the same repeatable EMI traces, even across different television sets. Under laboratory conditions, researchers were able to identify 1200 movies at a 92% accuracy rate by reviewing these trace EMI patterns.¹²

Given the present and developing abilities to use energy data to detect appliance usage, discern regular household habits, and review the in-home consumption of video content or online information, the Working Groups must implement protections that guard such personal information and align with the requirements of the Privacy Rules.

B. Known Limits to Anonymization and Aggregation as Methods for Preventing Re-identification and Protecting Privacy.

As described further below and in Appendix A, scientists now recognize that aggregating or anonymizing data to sufficiently prevent re-identification of an individual is almost impossible. As such, instead of relying directly on these techniques, instances of re-identification have prompted new efforts among computer science and privacy experts to "balance the risks

¹⁰Presentation of Ashwin Machanavajhala at January 15 workshop (slides available at ftp://ftp.cpuc.ca.gov/13011516_EgyDataWorkshop).

¹¹ Presentation of Lee Tien, EFF at January 15 Workshop (slides available at ftp://ftp.cpuc.ca.gov/13011516_EgyDataWorkshop)

¹² Jawurek, et. al., "SoK: Privacy Technologies for Smart Grids – A Survey of Options" at 5, *available at* <http://research.microsoft.com/pubs/178055/paper.pdf>.

and value of data sharing in a de-identification regime.”¹³ Existing and developing re-identification capabilities must inform the Working Group’s decisions on the dynamic definitions of aggregated/anonymized data to give privacy-protecting protocols any value.

In this section, we summarize for the Working Group some of the research shared in the workshops and previous proceedings, from consulting with experts, and from scientific literature, showing that these techniques fail to effectively protect customer privacy, and that data that have been anonymized or aggregated remain subject to the Privacy Rules, which cover all information about the customer that is “reasonably re-identifiable.” For more detail, please see George Danezis’ analysis in Appendix A.

1. Anonymization

Anonymization techniques attempt to protect anonymity of data subjects by removing personal identifiers, such as names and addresses, from the data. Although anonymized data do not, on their own, point to specific individuals, numerous examples demonstrate that re-identification can be achieved by comparing anonymized data with external information that contains corresponding data points. See, for example, Appendix A, which offers the example of cross-referencing a customer’s load profiles against external information about that customer’s occupancy, allowing someone to re-identify the individuals referenced in the data.¹⁴ It explains that a customer’s (sometimes public) travel schedule, mobile phone location records, or even a short period of observation of the customer’s house might be enough external information to match the anonymized load profile to a particular utility customer.

As evident in the case studies below, the removal of key identifiers, such as the data subject’s name, address and birthdate, is insufficient to protect customer privacy.

a. Examples: Netflix and AOL Research Datasets

Professors Jennifer Urban and Ashwin Machanavajjhala both noted the Netflix Prize privacy breach at the January workshop. Netflix offered a prize for the contestant who could develop the best algorithm for matching users to films and released anonymized, customer-specific data to get them started. University of Texas-Austin researchers Arvind Narayanan and

¹³ Paul Ohm, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,” 57 UCLA Law Review 1701 (2010); Jane Yakowitz, “Tragedy of the Data Commons” (March 18, 2011). Harvard Journal of Law and Technology, Vol. 25, 2011. Available at SSRN: <http://ssrn.com/abstract=1789749>.

¹⁴ George Danezis, *Privacy Technology Options for Protecting and Processing Utility Readings*, Mar. 1, 2013, p. 3.

Vitaly Schmatikov, however, combined the data with available information from the Internet Movie Database, allowing them to re-identify users.¹⁵ This brought Netflix under legal process and the scrutiny of the FTC; ultimately, Netflix chose not to pursue further similar competitions.

Professor Machanavajjhala also highlighted a privacy breach experienced by AOL as a further example. In 2006, AOL decided to publish search logs, containing user search queries, to help researchers communities improve searching algorithms. AOL user IDs were replaced by random numbers. No names or other traditional identifying information was included with the search queries. Within two hours, researchers were able to reveal a photograph of a particular user, based on review of the search queries. The fact that the anonymization attempt was broken in only two hours demonstrates how trivial it would be for an attacker to identify specific households within an “anonymized” energy usage data set with a small amount of external information about that customer’s energy consumption. Disclosure of supposedly anonymized data for energy research purposes, such as to multiple third parties to assess energy efficiency programs, could create similar problems for the utilities, the Commission, or researchers, highlighting the need to address these risks in developing data protocols.

b. Example: Massachusetts Government Health Data

Professor Machanavajjhala additionally noted the Massachusetts government breach involving medical information. In 1997 the Massachusetts government began making “anonymized” health records of state employees available to researchers. Patients’ names, addresses, and SSNs were removed from the health records, which otherwise remained intact. The governor assured his citizens that it would be impossible to re-identify individual patient information. Within two days, an MIT graduate student was able to identify the Governor’s health records by cross-referencing them against voter registration records. She mailed the Governor’s health records to him in an envelope.¹⁶

Professor Machanavajjhala referred to data points shared with data from external sources—like the voter registration records the researcher used here—as “quasi-identifiers” because they can identify an individual, but require comparison with other data sets in order to

¹⁵ Arvind Narayanan and Vitaly Shmatikov “Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset),” Feb. 5, 2008, U. Tex. at Austin, *available at* <http://arxiv.org/pdf/cs/0610105v2.pdf>.

¹⁶ Erica Klarreich, “Privacy by the Numbers: A New Approach to Safeguarding Data,” in *Scientific American*, at 1 December 31, 2012 (*available at* <http://www.scientificamerican.com/article.cfm?id=privacy-by-the-numbers-a-new-approach-to-safeguarding-data>) (Hereinafter Klarreich)

do so. In the energy world, a number of other data points could qualify as quasi-identifiers, including sets of appliances, devices, or vehicles, patterns of appliance usage, sleep patterns, and potentially a variety of other information. At the January workshop, some presentations included intentions to compare energy data to external sources, such as state-wide and county assessor maps, as well as data on building characteristics.¹⁷ Knowing that researchers seeking anonymized energy use data intend to combine that data with additional information sources highlights the need for Working Group members to take seriously the potential risk to utility customer privacy that could occur via re-identification techniques.

c. Example: Amazon Purchase History

In 2011, researchers showed that it is possible to determine an online shopper's personal purchase history simply by studying the displays on Amazon.com's product recommendation feature. The researchers noticed that the aggregate-level statements—"Customers who bought this item also bought A, B and C"—changed over time, based on a shopper's own purchase history. By cross-referencing the product recommendations with customers' public reviews of purchased items, the researchers could successfully infer that a particular customer had bought a particular item on a particular day, even before the customer had posted a review of the item.¹⁸

Energy data similarly changes over time, allowing for noticeable patterns to appear. Unique energy signatures become personally identifying characteristics when compared to external information with shared data points. In addition, many of the same characteristics, such as name, address, birthdate, etc., are collected by utilities, as were in the Massachusetts government health data breach or by online service providers like Amazon, Netflix, and AOL. Further, many of these characteristics are available to the public on other databases, making it possible to identify an individual through linking other data.

These examples, among others, explain why anonymizing data by removing a few key identifiers unfortunately does little to prevent re-identification. In some cases, it was only a matter of hours before data considered "anonymized" was cross-referenced with external data and re-identified, compromising the data subject's privacy. As such, data that has been "anonymized" is often easily re-identifiable. Accordingly, data that has been processed with

¹⁷ See Presentations of Lauren Rank, Mike McCoy, and Paul Matthew from January 15 workshop. (slides available at ftp://ftp.cpuc.ca.gov/13011516_EgyDataWorkshop)

¹⁸ Klarreich at 3.

these types of anonymization techniques, without additional protective steps, would still be considered “covered information” under the Privacy Rules. As a result, it can only be released with consent or otherwise pursuant to the Rules, and without additional steps in place, could expose customers to re-identification risks

2. *Aggregation*

The use of the term “aggregated data” has not been consistent throughout this proceeding. Based on the scientific literature in this area, we understand aggregated data not to include micro-data—i.e., the underlying, discrete records about individuals from which the aggregation is derived. Unlike attempts to anonymize data, for example by removing certain identifiers from individual records, aggregating data requires processing it such that there are no individual-level records, for example by computing the sum or the average of a group of individual households’ energy usage information. For our purposes, “aggregated data” would not include the total annual or average annual energy usage for an individual household, precisely because the data pertains to a specific household.

Despite excluding micro-data, aggregated data can still leak private information. Traditional privacy protections for aggregation, such as the 15/15 Guideline, are sometimes referred to by computer scientists as “naïve aggregation rules” because of the uncomplicated techniques for circumventing their restrictions.

To use an historical example, this one from as far back as World War II, it is now well-known that re-identification of naively aggregated Census Bureau data helped the U.S. military locate and transfer Japanese-Americans to internment camps during World War II. Although naïve aggregation was considered an acceptable privacy policy in the 1940s, today’s Census Bureau employs a series of complex data-blurring techniques to promote data integrity but maintain heightened security in response to such re-identification risks.¹⁹

The 15/15 Guideline is the most prominent “aggregation” model in this proceeding.²⁰ Although burying an individual’s data within a larger data set like this may seem like a reasonable means to protect privacy, the shortcomings of this approach are well documented.

¹⁹ 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.

²⁰ The 15/15 Guideline is a model that permits a database to generate query results, only if the results represent an aggregate data set consisting of 15 or more individual utility customers and no one utility customer in the set constitutes 15% or more of the total aggregated data.

Specifically, a carefully crafted series of queries can generate aggregate results that, when looked at together, reveal customer-specific information. A brief explanation of how queries can work around the limits imposed by the 15/15 Guideline is given below, followed by an example of the risks of cross-referencing aggregated data with external sources. Please see Appendix A for further discussion of data security issues with the 15/15 Guideline.

a. *Likely Smart Grid Data Leaks from Naïve Aggregation Rules*

The 15/15 Guideline and similar well-intentioned standards unfortunately exhibit fundamental flaws that render them unable to effectively defend customer privacy. Numerous researchers have addressed how a combination of queries can enable the re-identification of individuals represented in aggregate data, even though neither query on its own infringes the individual's privacy.²¹

To illustrate, imagine a quantitative query system²² under a standard like the 15/15 Guideline, which ignores requests when the number of results is less than a particular threshold. In such a case, one need only ask two questions that meet that threshold to obtain an answer otherwise forbidden by the rule:²³

The first question:

How many people in this database exhibit power usage patterns consistent with using a television and video games in the afternoon, but patterns consistent with additional appliances, electric vehicles, and lights in the evening?

²¹ Salil Vadhan, et. al. Comment on “Advance Notice of Proposed Rulemaking: Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators” HHS-OPHS-2011-0005 at 6.

[In an] interactive system designed to answer queries about the health care expenses of the Harvard faculty, which allows queries of the form “how many Harvard faculty satisfy X” where X is a search criterion that can involve attributes like age, health care expenses, and department. While “how many” questions may seem relatively safe when computed over a population of 2000+ individuals, they are not. By asking the question “How many Harvard faculty are in the computer science department, were born in the U.S. in 1973, and had a hospital visit during the past year?,” it is possible to find out whether one of the authors of these comments (S.V.) had a hospital visit during the past year (according to whether the answer is 0 or 1), which is clearly a privacy violation. A common “solution” to this sort of problem is to only answer queries whose answers are sufficiently large, say at least 10. But then, by asking two questions --- “how many Harvard faculty had hospital visits during the past year?” and “how many Harvard faculty, other than those in the computer science department and those born in the U.S. in 1973, had hospital visits during the past year?” --- and taking the difference of the results, we can obtain an answer to the original, privacy-compromising question.

²² For example, how many individuals in this data set have characteristic X?

²³ Klarreich at 2.

The second question:

How many people in this database who exhibit power usage patterns consistent with using a television and video games in the afternoon, but patterns consistent with additional appliances, electric vehicles, and lights in the evening, do not live at 100 Main Street?

Although both questions provide aggregated results, the combination of these two questions has effectively "leaked" information about 100 Main Street. The first question essentially asked for the total number of homes where children are likely to be home alone in the afternoon. The second question sought the same information but excluding 100 Main Street. If the answers to these two questions are the same, then one can reasonably infer that there are no latchkey children at 100 Main Street; if the answers differ by 1, then one can reasonably infer that there are. See Appendix A for further detail regarding problems with the 15/15 Guideline.

Unfortunately, it is very difficult for computer programs to detect the query combinations that breach customer privacy in advance.²⁴ Professor Machanavajjhala pointed out at the January workshop that energy data is dynamic, not static. If aggregated data changes, then individuals can be uniquely identified in ways that computers were not programmed to protect against. For example, if data shows a new house on the block, then an attacker can look at changes in the neighborhood's energy consumption and subtract the new information to attribute change to the new home.

Because this simple, two-query process for overcoming the 15/15 Guideline defeats its protective purpose, data masked in this manner is likely to remain re-identifiable. As such, like data that has been subjected to basic anonymization techniques, data aggregated according to these techniques would still be considered "covered information" under the Privacy Rules, and would expose customers to re-identification risks if released without additional protective protocols in place.

b. Attacks Using Pre-existing Information about an Individual

If an attacker or researcher has background information about an individual represented in an aggregated data set, re-identification becomes even easier. For example, in 2008, a research team, led by Nils Homer, then a graduate student at the University of California at Los Angeles,

²⁴ Klarreich, at 2.

showed that in many cases, knowing a person’s genome can help determine, beyond a reasonable doubt, whether that person had participated in a particular genome-wide test group.

Homer’s research team demonstrated the risks of disclosing aggregate information from genome-wide association studies, one of the primary research vehicles for uncovering links between diseases and particular genes. These studies typically involve sequencing the genomes of a test group of 100 to 1,000 patients who have the same disease and then calculating the average frequency in the group of something on the order of 100,000 different mutations. If a mutation appears in the group far more frequently than in the general population, that mutation is flagged as a possible cause or contributor to the disease.²⁵

After Homer’s paper appeared, the National Institutes of Health reversed a recently instituted policy that had required aggregate data from all NIH-funded genome-wide association studies to be posted publicly.²⁶ In this example as in others, the comparison of supposedly “safe” data to external, background data led to re-identification.

Energy data is susceptible to the same sorts of attacks on other types of personal data. If an attacker knows the unique combination of appliances that a utility customer has in their kitchen, he can examine aggregate energy usage patterns to determine if the data signature corresponding to that combination of appliances fits the aggregate profile, which would lead to an inference that the customer was or was not included in the data.

Accordingly, with certain background information and data manipulation, data aggregated according to these techniques, as well, can easily be re-identified—especially as researchers, marketers, or others combine datasets—and would still be considered “covered information” under the Privacy Rules.

The Working Groups will need to consider carefully protocols to protect energy usage data in order to find methods that take attacks like those we have described into account. As noted next, we believe specific technical expertise is required in order for the Working Groups to sufficiently consider the issues and develop appropriate approaches.

²⁵ Klarreich at 2–3.

²⁶ Klarreich at 3.

C. Technical Expertise Is Required to Develop More Robust Privacy Solutions Because Anonymization and Aggregation Techniques Alone Fail to Protect Private Customer Data

We hope this background is helpful to the Working Groups. As made clear during our analysis and in the examples above, when devising protocols for the disclosure of customer data, Working Group participants should be aware that neither aggregation nor anonymization can be defined or evaluated in static terms if privacy is to be protected. Re-identification is a dynamic concept. Each time there is an influx of publicly available data, an advance in computer technology, or additional collection of personally identifying characteristics, re-identification strategies will evolve. This means that the techniques required for the “safe” release of smart grid data will likely also change. Any definitions adopted by the Working Groups will need to accommodate this reality. In order to do this, the Working Groups need to consult experts in the fields of computer science, consumer privacy, and data security at each stage of developing data disclosure procedures, in order to understand the unfortunate, but genuine challenges in securely sharing data and to develop feasible solutions that overcome the known shortfalls of anonymization and aggregation.

D. Summary and Next Steps

In summary, we hope this memorandum has supplied the Working Group with useful background information to move forward in this proceeding, acknowledging that:

- ❖ Both scientific research and live, real-world examples show that basic techniques for anonymizing or aggregating data do not by themselves provide sufficient protections to customer privacy.
- ❖ Unfortunately, the 15/15 Guideline and similar well-intentioned aggregation standards cannot be relied on to protect customer specific data because of simple workarounds that neither human beings nor computer programs can reliably predict.
- ❖ The dynamic nature of energy data and the constantly developing technologies for de-identification and re-identification should each be considered by the Working Groups in developing definitions and proper disclosure procedures.

- ❖ Consultation with technical experts is necessary at all stages of this proceeding to determine:
 - What types of data can be released or should not be released under the requirements of the Privacy Rules;
 - What privacy solutions have been shown from experience to adequately or inadequately protect customers' private information; and
 - What feasible solutions can the Commission use to impart sufficiently robust protections of customer privacy while still providing useful energy data for valuable research purposes. (See, for example, the suggestions under "Robust Privacy Technology Options" in Appendix A.)

Respectfully submitted this April 1, 2013 at San Francisco, California.

/s/ Jennifer Urban

JENNIFER URBAN, Attorney
Samuelson Law, Technology & Public Policy Clinic
University of California, Berkeley School of Law
396 Simon Hall
Berkeley, CA 94720-7200
(510) 642-7338
Attorney for ELECTRONIC FRONTIER
FOUNDATION

/s/ Lee Tien

LEE TIEN, Attorney
Electronic Frontier Foundation
454 Shotwell Street
San Francisco, CA 94110
(415) 436-9333 x102
Attorney for ELECTRONIC
FRONTIER FOUNDATION

Appendix A

PRIVACY TECHNOLOGY OPTIONS FOR PROTECTING AND PROCESSING UTILITY READINGS

George Danezis
Paris, Friday, 1 March 2013

SCOPE OF THE DOCUMENT

This document discusses the privacy concerns surrounding the collections and processing of granular readings from next generation utility architectures, such as smart electricity grids. New generation distribution systems rely partially on computerised meters installed in households and businesses that record more information than previous electromechanical meters, and have facilities to transmit them regularly to the energy operators and distributors. A modern smart meter is capable of recording consumption of electricity, as well as production, at a very fine granularity, close to “real time.” Most deployments in the US²⁷ and Europe²⁸ are presently working toward readings every 15 minutes to 30 minutes respectively (48 or 96 readings per day) uploaded as a single “load profile” about once a day. These are collated with other readings from the same household to build larger load profiles over months or years. This document is concerned with the management and privacy of those detailed readings – other information such as billing details, demographics and subscriber information are broadly similar to information already gathered and benefit from established processes to ensure their security and privacy.

The management of the electricity grid is special, compared to water and gas, in that production and consumption has to be balanced very carefully at all times. Some production requires significant planning to start or stop, and the use of renewables adds uncertainty as to the capacity. These make forecasting and demand response mechanisms important. On the other hand, gas and water provision is also undergoing computerization in its control and distribution, since better recording of consumption could be used to optimize the delivery of those services (like detect leaks). Those attempting to manage privacy issues in smart grids, and the regulatory and technical solutions applied, should therefore foresee that they will create a precedent for the management of other utility data. Furthermore those undertaking privacy impact assessments for managing and processing utility readings should be mindful that combined readings from all utilities may be available at some point, providing a multi-dimensional view into household habits.

²⁷ Guidelines for Smart Grid Cyber Security: Vol. 2, Privacy and the Smart Grid. National Institute of Standards and Technology. NISTIR 7628., August 2010.

²⁸ Smart metering implementation programme data access and privacy consultation document. United Kingdom Department of Energy and Climate Change, Consultation Document, April 2012.

Readings and load profiles have direct and indirect uses. They are used directly by the energy industry to monitor and balance production / consumption, forecasting energy needs in the short and long term data, plan for future distribution capacity, and bill customers at a coarse or fine granularity. Where the energy sector is private and competitive, meter readings are also used to settle contracts in the energy market. Billing customers according to the time they consume electricity is particularly promising to provide incentives to reduce consumption at peak time, and is generally called time-of-use tariffs.

Indirect uses are also foreseen for detailed readings for both research and operations: they can be used for monitoring and providing advice on energy efficiency of homes and devices, understand penetration of smart vehicles in different areas, insurance, marketing of renewables, risk management of credit, etc. These are indirect uses since they are not vital for the day to day operation of electricity provision, and may not be performed by the traditional players in the energy industry. In fact, indirect uses are of great interest since they may create new services, or optimize and economically “disrupt” existing ones. Research is a particularly important area that requires data, and by its very exploratory nature, it might require more access than an operational system.

The focus of this document is to provide an overview of technical and other options that support processing of the meter readings to support both direct and indirect uses, and their benefits, while minimizing the exposure of the readings and providing protection of the privacy of households, businesses and government agencies making use of modern grid technologies.

OVERVIEW OF THREATS

Fine grained meter readings recorded by smart meters from households are widely recognized as privacy sensitive. NIST²⁹, in the US, recommends they are processed as PII (Private Identifiable Information) and jurisdictions with horizontal data protection regimes (Canada and the EU) consider that load profiles fall under their provisions³⁰. Substantively, detailed smart meter reading provide a record of activity from within a household that might otherwise be difficult to infer. This activity might be sensitive for occupants. We outline here a number of possible privacy and security threats resulting from the collection and mining of readings:

- Meter readings at the granularity of 15-30 minutes can be used to infer the occupancy of a home, since aggregate half-hourly consumption goes when one is at

²⁹ Guidelines for Smart Grid Cyber Security: Vol. 2, Privacy and the Smart Grid. National Institute of Standards and Technology. NISTIR 7628., August 2010.

³⁰ Opinion 12/2011 on smart metering. Article 29 Decision, April 4 2011.

home. They leak information about when occupants may be away on holiday, at work or not. As a result compromised readings contain information that could be used to target homes for burglary when they are empty. Interestingly, one of the earliest cases of widespread indirect use of meter readings involved inferring occupancy to detect safe houses of German terrorists³¹. This particular practice was later deemed unconstitutional by German courts.

- Similarly, granular readings can be used to estimate the number of inhabitants at a particular time. Third parties also profile inhabitants in relation to their family situation: for example to discover whether a spouse is working or not. Houses shared by multiple unrelated occupants also exhibit a different pattern of electricity consumption than houses inhabited by a single family.
- Detailed smart meter readings contain information about the sleeping patterns of inhabitants, which can be surprisingly intrusive. Sleeping patterns are associated with specific religious groups: comparatively early morning activity in the months of Ramadan is a sign of a practicing Muslim household. Erratic patterns of sleeping are also indicative of poor health: irregular use of electricity at night may be indicative of early stages of prostate cancer. A change in the use of electricity (for frequent washes) as well as night time patterns of use may indicate to a third party a household with a young child.
- Non-intrusive appliance monitoring³² techniques detect which appliances are in a home, and when they are used, from fine grained readings of a whole household. While the frequency of readings in current smart-metering deployments is too coarse for a direct application of those techniques, it is clear that some information on appliances, such as the presence of an electric vehicle, a fridge, air-conditioning, or an electric oven can be inferred. It is noteworthy that modern smart meters can be configured, even remotely and without the knowledge of the household, to take readings at a finer granularity. More recent studies have demonstrated under laboratory conditions that electricity consumption can even leak information about which TV channel is being watched³³.
- Even more intrusive information can be inferred when combining electricity with other utility readings, for example water and gas readings. Such combined readings can be used to detect different patterns of cooking in a household, since cooking activity exhibits correlated uses of electricity, gas and water. Similarly, the frequency of use of a dishwasher or washing machine can be inferred. Finally, the combined use of large volumes of water along with either gas or electricity can be

³¹ B. S. Amador. The federal republic of Germany and left wing terrorism. Master's thesis, Naval Postgraduate School, Monterey, CA, December 2003.

³² G. W. Hart. Residential energy monitoring and computerized surveillance via utility power flows. IEEE Technology and Society Magazine, June 1989.

³³ M. Enev, S. Gupta, T. Kohno, and S. Patel. Televisions, video privacy, and powerline electromagnetic interference. In Proceedings of the 18th ACM conference on Computer and communications security, pages 537–550. ACM, 2011.

used to infer how often members of the household have showers. Electricity and water provides information about night time patterns of sanitation, and even how often and when inhabitants use the toilet overnight.

Besides the above sample privacy threats, the rationale for storing and processing of meter readings is the extraction of some level of information about a consumer. As such any argument about the value of meter readings at the granularity of a household becomes an argument about potential privacy invasion, as the information originates from, and characterizes, a household. In line with fair information practices³⁴ this information should only be used with the knowledge and consent of the household, to ensure their best interests are at the heart of any indirect processing.

Besides legal or substantive privacy concerns, smart meter deployments have been jeopardised partly through the poor handling of customer privacy and protection concerns. For example, the smart meter deployment in the Netherlands³⁵ had to be put on hold due to consumer revolt.

As a result of the above we consider there are serious risks associated with the bulk storage, processing and availability of detailed utility meter readings. First of all, organizations holding such data can be compromised, or lose the data due to mishandling. This is a serious threat to consumers, and the reputation of the entity that that is a victim of a cyber-attack or a mistake. Organizations holding data may also be compelled to reveal the readings they hold, though the legal process of countries they operate in. In some jurisdictions even divorce or private dispute cases can lead to organizations being compelled to reveal information about their customers. Finally, organizations themselves may be tempted to process the readings to gain an unfair advantage in their commercial dealings with customers.

PARTIAL SOLUTIONS AND CAVEATS

A number of solutions are popular to mitigate the perceived risks of handling and processing detailed meter readings. In particular opt-in/opt-out mechanisms, anonymization, and naïve aggregation rules are popular due to their conceptual ease, and relative low cost of implementation. Despite being valuable parts of a larger strategy, in themselves, these mechanisms cannot guarantee the level of protection one would hope for the privacy of readings and households.

³⁴ FTC Fair information practices (<http://www.ftc.gov/reports/privacy3/fairinfo.shtm>)

³⁵ Cuijpers, Colette and Koops, Bert-Jaap, Smart Metering and Privacy in Europe: Lessons from the Dutch Case (February 15, 2013). In: S. Gutwirth et al. (eds), *European Data Protection: Coming of Age*, Dordrecht: Springer, pp. 269-293 (2012).

OPT-IN/OPT-OUT

Both guidelines for processing PII in the US (fair information processing practices) and data protection regimes consider that, where possible, the informed consent of the data subjects should be sought for any otherwise non-necessary processing. The UK regulator DECC³⁶ has proposed a gradual system of consent to enable processing of increasingly invasive data: the provision of one reading a month per household is absolutely necessary and therefore obligatory; the provision of a reading per day is subject to customer opt-out, but in its absence collection and processing can go ahead; finally any finer grained processing (as for computing time-of-use tariffs) requires an explicit opt-in from the customer.

The requirement to obtain consent for collection and processing is in itself positive, particularly for indirect uses of readings, where a customer may not have reasonably foreseen it. Yet, it does not alleviate all risks: despite consent to collect and process, readings are still sensitive, and could still be lost or compromised. Therefore some technical protection is still necessary to ensure this sensitive information is stored and processed to minimize its exposure to external or internal risks. Furthermore once bulk readings are available in clear it is difficult to audit what they are used for, to ensure that only authorised processing is taking place.

Finally, a key limitation of solely relying on opt-in as a privacy protection is purely economic. In case time-of-use tariffs become the norm, and added value services relying on energy readings are commonplace, households opting out will find themselves marginalized or possibly unable to benefit from the best prices for the goods and services they receive. Therefore they will be faced with a harsh choice of either opting into a system with poor privacy or being charged a premium for opting out. For this reason it is important to consider additional technical privacy protections even for customers opting in advanced services.

ANONYMIZATION

One option for minimizing the danger to households, from the processing of any private information is to first anonymize it. Anonymization³⁷ removes any personal identifiers from the data in an attempt to make it difficult to link it back to a specific individual or household. Anonymization is an extremely flexible mechanism: full load profiles over time are available to researchers and any function can be computed on them. Sadly, robust

³⁶ Smart metering implementation programme data access and privacy consultation document. United Kingdom Department of Energy and Climate Change, Consultation Document, April 2012.

³⁷ C. Eftymiou and G. Kalogridis. "Smart grid privacy via anonymization of smart metering data." 2010 First IEEE International Conference on Smart Grid Communications, pages 238–243, 2010.

anonymization of load profiles is extremely difficult due to this abundance of data on one side, and the abundance of side information on the other.

Firstly, household energy consumption is rather regular over time. This means that the availability of a short period of non anonymized data can be used to link anonymized load profiles back to the household³⁸. Concretely this means that an entity that has a short period of readings from a household, for example a month, can use those readings to pick a longer anonymized load profile related to the same household. To do this, a number of markers would have to be extracted from the raw identified load profile, such as the presence of certain household devices, number of occupants, typical patterns of occupancy related to the schedule of inhabitant's work, school or recurrent appointments. Then the anonymized profiles can be sieved according to the same markers, looking for a match. Different households may be susceptible to this matching to different degrees but some, with very stable unique markers, will be trivially re-identifiable.

Secondly, detailed load profiles are correlated with activities in the home that may be known, public or discoverable by others. Thus markers can be constructed to match other activities linked with specific individuals with anonymized load profiles. Any side-information associated with occupancy can be used³⁹: public traffic schedules, a short period of direct physical observation of the home, mobile phone location records or internet access records can be used to construct markers. Thus anyone in the possession of such data sets can create an approximation of a load profile over time, and then attempt to match it with the database of anonymized load profiles. This technique is likely to be much more successful than the previous one, since it does not rely on regularity of habits over time.

For the sake of clarity we present a concrete de-anonymization attack using side-information:

- Consider an on-line web service, like webmail, on which a known target user has an account and checks periodically both from home and outside the home.
- The service logs contain a time series of accesses, and the network address (IP address) of these accesses. The network address leaks whether the user is at home or outside the home, through differentiating between a home internet service provider and a mobile or business internet service provider. Using a different computer at home than at work, can also be leveraged to mount the re-identification attack.

³⁸ M. Jawurek, M. Johns, and K. Rieck. "Smart metering de-pseudonymization." In ACSAC, pages 227–236, 2011

³⁹ A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin. "Private memoirs of a smart meter." In Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, BuildSys '10, New York, NY, USA, 2010. ACM.

- The service is then provided with a large number of anonymized electricity load profiles, and wishes to re-identify a target user. To achieve this, the service makes the reasonable assumption that a user at home consumes more electricity than a user outside the home.
- For each anonymized trace the services computes this simple statistic: it adds the readings corresponding to times the target user was observed at home, and subtracts the readings when the target user was observed outside the home.
- The anonymous trace corresponding to the target user should achieve a high value of this statistic – ultimately the highest value.

This is the result of the actual trace matching perfectly the observations of occupancy, while other traces being partially independent of it. The more side-information the service has about the user, meaning more accesses to the on-line service, the better the estimation of the statistic and the more confident it can be the de-anonymization attack will be successful. This example illustrates that mounting a de-anonymization attack against an anonymized load profile is computationally cheap, and the side information required only needs to be vaguely related to occupancy – and as such is plentiful and in the hands of many third parties.

De-anonymization techniques may be new in the field of smart-grids, but general techniques are already very mature in related fields of statistical databases privacy or social network privacy. Recently, researchers have demonstrated the inherent dangers of publishing rich anonymized datasets: they managed to de-anonymize a number of users from a dataset of movie preferences published by the Netflix Company using side information from other public sources⁴⁰. In that work they used particular combinations of movie preferences attached to known persons as “markers”, and then detected those markers in the anonymized data set to link it to individuals.

Thus, anonymization through the mere removal of obvious identifiers is now recognized as a very weak privacy protection mechanism⁴¹. It could be used to protect load profiles from mistakes or accidental disclosure, but it is fundamentally a mechanism to keep honest people honest. It cannot protect against a malicious entity that, for example compromised the dataset and is trying to identify specific households.

NAÏVE AGGREGATION RULES

In terms of flexibility another option, besides anonymization, involves providing an “aggregation service” that computes aggregate statistics on specific data items on request,

⁴⁰ Narayanan, Arvind, and Vitaly Shmatikov. “How to break anonymity of the netflix prize dataset.” arXiv preprint cs/0610105 (2006).

⁴¹ Ohm, Paul. “Broken promises of privacy: Responding to the surprising failure of anonymization.” *UCLA Law Review* 57 (2010): 1701.

and returns only the aggregate results. The hope is that aggregation obscures information about individual households, alleviating privacy concerns. Rules are put in place to ensure each datum is computed on the basis of many households and rounding or suppression can be used to obscure items that do not conform to the rule. One such example is the so-called “15/15 Guideline” that stipulates that at least 15 households are involved in any aggregate.⁴²

Sadly there is an extremely mature⁴³ and rich⁴⁴ literature outlining generic attacks against systems that provide the facility to query datasets and return statistics in a naïve manner, despite complex sanitization rules. It has been shown that special queries (called “Trackers”) can be crafted, each conforming to the rules, but jointly leaking private information.

Building a tracker for the 15/15 rule is simple. The rule stipulates that a query can only be performed if it concerns a certain minimum number of households: an analyst can submit a query that concerns a large number of specific households (say 1000); then a second query over the same households plus an additional one (namely 1001 records) is performed. The result of the two queries jointly leaks all information about the record that was included in the second query, despite the fact that the queries are compliant with the 15/15 rule. Furthermore, one can show that it is very expensive to audit for sets of queries that are crafted to leak information about single records: one would have to consider the potential leakage of all subsets of queries – and the number of these subsets is very large indeed.

Thus, while allowing querying of a database of records provides flexibility, it has to be supported with great care to ensure no information about individual households is leaked. Positive guarantees of security and privacy must be proven for any sanitization rule to ensure that tracking queries cannot be crafted to extract information.

ROBUST PRIVACY TECHNOLOGY OPTIONS

Privacy protection through procedures or technology is an exercise in risk management that has to balance the benefit of processing the data and the potential privacy risk to households. It is important to note that the benefits of indirect processing may in fact not directly benefit households. Therefore regulators must be very cautious to ensure those benefiting from the processing do not choose alone what constitutes an acceptable risk. In many cases, technology can help to minimize risks, while also maximizing benefits, and thus privacy does not have to be a zero-sum game. A privacy-by-design methodology can

⁴² Audrey Lee, Marzia Zafar. “Energy Data Center”. Briefing paper. September 2012.

⁴³ Denning, Dorothy E., Peter J. Denning, and Mayer D. Schwartz. “The tracker: A threat to statistical database security.” *ACM Transactions on Database Systems (TODS)* 4.1 (1979): 76-96.

⁴⁴ Adam, Nabil R., and John C. Worthmann. “Security-control methods for statistical databases: a comparative study.” *ACM Computing Surveys (CSUR)* 21.4 (1989): 515-556.

be applied to identify the privacy issues throughout the development of a smart-metering system⁴⁵, and appropriate privacy technologies can be deployed to support privacy policies⁴⁶.

SAMPLING LOAD PROFILES, ANONYMIZING & LICENCING

The first, mostly procedural, option for processing detailed readings is to establish a scheme to provide sampled anonymized load profiles to clearly identified, authorized and overseen researchers for pre-determined uses. In that case anonymization is used to ensure that data leaks do not happen accidentally. A high sampling rate, of say one household in 100-1000 could be used to ensure that any compromise would not leak a very large volume of information, and that any specific target household for which there might be a lot of information is not likely to be in the set of load profiles available for analysis.

Yet, providing anonymized data under a licence or an NDA is not a perfect protection, and some household may have valid reasons to object to taking this risk. It is worthwhile considering explicit opt-in from households for use of load profiles in indirect processing for research through such a scheme. To be fully honest consent should be obtained under the assumption the sharing of the data is not fully anonymized, and possibly financial incentives should be provided to participating households.

On the technical side, getting data under licence should be accompanied with a robust audit of an organizational operations and technical procedures to ensure the security of that data. This should include secure authentication, storage, transport, audit, deletion mechanisms and an ownership structure that ensures the data will be processed according to the licence.

This mechanism is ideally suited for advanced R&D that requires access to full load profiles for exploration. It might also be used to perform computations as part of operations, when complex calculations need to be performed on full load profiles.

AGGREGATION & QUERY PRIVACY

The workhorse of most processing is likely to be access to aggregates and statistics based on a number of load profiles. For example, it is legitimate to monitor the aggregate consumption per region, changes over time, or even extract “average” load profiles for researching tariff structures or to train forecasting models. All those uses require readings only as a means to aggregating them into statistics, and not to make decisions on individual

⁴⁵ “Operationalizing Privacy by Design: The Ontario Smart Grid Case Study.” Information & Privacy Commissioner, Ontario, Canada. February 2011.

⁴⁶ “Smart Meters in Europe: Privacy by Design at its Best.” Ann Cavoukian, Ph.D. Information and Privacy Commissioner, Ontario, Canada. April 2012.

households. A number of privacy technologies allow access to those aggregates without making available detailed readings.

To compare to the naïve aggregation rule architectures, architectures that allow secure privacy friendly aggregation rely on a centralized party (or parties) holding the readings, and accepting queries to be performed on the data. Once the query is performed the answer is returned, possibly with some slight modification to ensure that information is not leaked. Queries can be pre-registered and data streams for each query can be produced ahead of time and made available to third parties in real-time.

For simple aggregation, involving sums and weighted sums, a very high degree of privacy can be provided through the use of appropriate encryption technologies^{47 48}. Meter readings can be stored encrypted, thus preventing even the storage service from accessing them in detail. Queries are performed on the encrypted readings, for example to compute encrypted sums over time or space, and returned to the relying services. Special encryption techniques can be used that “unlock” the results of queries to uncover the results, without giving access to any individual readings, with the help of a set of authorities overseeing the privacy policy. This architecture ensures that only the final aggregate result is available to anyone processing the readings. No one has access to raw readings, neither the storage service, nor the authorities nor the party receiving the result. Queries can be overseen by authorities for compliance to any policy, or to ensure they are appropriately rate limited to avoid exposing too much information to the any single entity.

Some aggregation is more complex than simple weighted sums. For example non-linear operations might have to be performed on readings before they are aggregated. In those cases the storage service needs to keep the readings in clear and process them to get the results. As we discussed, it is important to ensure no information can leak from specific or repeated tracker queries. One principled framework for achieving this is to ensure that statistics computed are differentially private⁴⁹, namely they are not overly influenced by the existence or absence of any single record, irrespective of the others (to protect against side information attacks).

We describe here two example mechanisms for ensuring an arbitrary statistic is differentially private:

⁴⁷ Klaus Kursawe, George Danezis, Markulf Kohlweiss: “Privacy-Friendly Aggregation for the Smart-Grid.” PETS 2011: 175-191

⁴⁸ Marek Jawurek, Florian Kerschbaum: Fault-Tolerant Privacy-Preserving Statistics. Privacy Enhancing Technologies 2012:221-238

⁴⁹ Cynthia Dwork: A firm foundation for private data analysis. Commun. ACM 54(1): 86-95 (2011)

- The first differentially private mechanism is called “the *Laplacian* mechanism”⁵⁰. One first computes the sensitivity of the statistic, as the maximum difference the inclusion or exclusion of any single item could make to the result of a query. Then some random noise is added to the result, drawn from a specific noise distribution, to mask any specific item, while providing information about the aggregate.
- The second mechanism is called “the *Subsample and Aggregate* mechanism”⁵¹. It is based on splitting a data set into smaller sub-sets; computing the statistic on each set; and then aggregating the result with some noise. Despite the fact the results are noisy, the average magnitude of the noise added is constant, therefore not overly influencing or biasing the result of queries on larger datasets.

The architecture of submitting queries to a service and getting back results, instead of processing load profiles locally, might be a departure from the habits of some researchers. In case few load profiles are processed a scheme based on licencing a sample of them may be preferable. Yet, in case large volumes of readings have to be processed, centralized processing in a data centre or private cloud may be the best option irrespective of privacy concerns. In that case the privacy-friendly architecture, that requires submitting queries to a service, aligns perfectly with the remote processing that would have to take place anyways, and is easy to add to existing computational models such as map-reduce⁵². Query based privacy mechanisms are highly scalable, and provide the ability to audit activity, and very flexible processing. There is no impediment to registering queries ahead of time, and receiving results in real time.

Privacy-friendly query systems can be made very privacy friendly. For simple statistics, they ensure that no single entity can ever get access to raw readings while providing real time access to aggregates and statistics. More complex computations require a storage service to store and process data in clear, but differential privacy mechanism ensure that the results cannot be used to infer much about any single household. They are also very efficient and scale to very large datasets.

USER AUTHORIZATION & DATA EXPORT

Ultimately some who would make indirect uses of meter readings may prefer per-household detailed load profiles. In those cases none of the previous privacy technologies are applicable, since they rely on sampling or aggregation. In such cases the reading storage service can still incentivise a privacy friendly use of the data by third parties by managing user authorization of processing.

⁵⁰ Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith: Calibrating Noise to Sensitivity in Private Data Analysis. TCC 2006: 265-284

⁵¹ Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In STOC, pages 75–84. ACM, 2007.

⁵² Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51.1 (2008): 107-113.

Conceptually, the storage service can manage the authentication of households to whom the data belongs, as well as services that wish to use the data. The storage service then ensures that permissions to access customer information have been granted by customers for each service. This is not dissimilar to the permission model used by modern mobile platforms (such as *Android* or *Windows Phone*) when an application wishes to access personal data from users. Social network platforms such as *Flickr* or *Facebook*, implement a similar authorization service for third party applications to access user feeds. Google dashboard also provides a model of an interface where a customer can go to manage their authorizations to applications, view and delete the results of computations. Providing such authorization and transparency mechanisms in one central place is highly advised.

Besides providing a well-defined API that allows third party services to access the data, after proper authorization and authentication from customers, the reading storage service can also provide to authenticated users their own household readings to use as they wish. In fact, one of the gravest challenges to privacy – in its information self-determination sense – is that a plethora of services may have access to customer information, when the customer does not. Besides providing access to raw readings, special cryptographic techniques can be used to ensure customer applications can process the data and compute results that can be used with third party services in a privacy friendly manner -- even without leaking the raw readings⁵³. These facilities can be used, for example, to produce verifiable time-of-use bills on customer devices, without leaking the raw readings. Any central store of information has a key role to play when it comes to facilitating and enabling a privacy friendly eco-system of applications. If it does not support core privacy services like private aggregation and queries, rich interfaces for authentication, authorization and data export it might block valuable applications due to privacy concerns, or force privacy invasive practices as the only option.

DESIGN FOR PRIVACY

The generic privacy protections presented are quite flexible, but specific applications using electricity readings may have features that make them amenable to other mechanisms for protecting privacy. It is therefore important to include in any R&D program a component that looks at the most privacy friendly way to gain value out of data, and provide rich services.

Unlimited and full access to vast amounts of data and all load profiles in R&D is detrimental to the development of privacy friendly solutions in the long term. The assumption of unlimited availability of data leads to lazy design, where such access becomes a necessity.

⁵³ Rial, Alfredo, and George Danezis. "Privacy-preserving smart metering." Proceedings of the 10th annual ACM workshop on Privacy in the electronic society. ACM, 2011.

Limiting access of researchers to only small sample rich datasets for exploration, and then services for privacy friendly processing of bulk data, incentivises the design of both privacy friendly research methods but also privacy friendly final products, business models, and long term operations.

We have seen that for small focused exploratory research projects, mechanisms based on anonymization, sampling load profiles and opt-in can be used to provide researchers with high quality datasets. For the provision of statistics, privacy friendly query services can provide aggregates or results of arbitrary computations on very large datasets without leaking information about any household. Finally, a proper framework for authorization, authentication and data access by users can enable an ecosystem of privacy friendly third party applications. These facilitate competition, can enable privacy friendly alternatives, and allow the user to have control over who is processing their data as they do in other on-line services.

SHORT BIO

George Danezis is a researcher at Microsoft Research on the topic of computer security and privacy. Before joining Microsoft in 2007 he was a visiting scholar at KU Leuven and a research associate at the University of Cambridge where he completed this PhD in 2004. George Danezis has been the program chair of the Privacy Enhancing Technologies Symposium (PETS) in 2005 and 2006, the conference on Financial Cryptography and Data Security in 2011, and the ACM conference on computer and communications security (CCS) in 2011 and 2012. He has published over 50 peer-reviewed scientific articles on the topics of privacy and security in international conferences and journals, and serves on the board of ACM CCS, PETS and ACM Information Hiding and Multimedia Security.