Open Energy Efficiency

Consultation Paper: Data Access Models for Energy Data
Comments prepared for
Australian Competition and Consumer Commission

23 Marcus Clarke Street
Canberra, Australian Capital Territory, 2601

22 March 2019

Contact Information
McGee Young
Chief Technology Officer
mcgeeyoung@openee.io
(414) 759-2599
364 Ridgewood Avenue
Mill Valley, CA 94941

Introduction and Response Overview

Open Energy Efficiency (OpenEE) appreciates the opportunity to respond to the Australian Competition and Consumer Commission's (ACCC) Consultation Paper *Consumer Data Right in Energy* (the Paper). OpenEE provides transparent, timely, and secure measurement of load impacts from demand-side energy management programs based on metered data. Our analytics platform utilizes open-source code (the OpenEEmeter), which tracks and normalizes metered load impacts at an individual-building level. Individual projects are aggregated into portfolios that are competitive and reliable as a demand-side resource. By generating near real-time performance insights, OpenEE enables effective management of traditional regulated programs as well as performance-based distributed energy resource (DER) procurement opportunities. OpenEE is both a consumer of energy data for a variety of use cases on behalf of our public and private sector clients, and a provider of secure and privacy protected platforms for the sharing of data.

OpenEE agrees that providing customers safe, secure, and efficient access to their energy data and rating product options is key to fostering informed choice and reducing costs. OpenEE also believes that safe, secure and efficient access to data, that respects customer privacy, for accredited data recipients (ADR) will be essential for the ACCC to foster modern and competitive energy and resource markets. These values, as well as OpenEE's experience working in emerging DER markets, leads us to recommend that the ACCC adopt Model 1 described in the Paper for a secure and centralised data platform managed by the Australian Energy Market Operator (AEMO). The initial investment to establish this model would ensure consistent and safe data management under the Energy Consumer Data Right (Energy CDR) and enable the dedicated implementation of the Energy CDR needed to effectively serve consumers and manage a rapidly evolving grid for years to come.

With adoption of Model 1, the ACCC has an opportunity to both enable competition and customer choice, while establishing a data framework needed for the AEMO to orchestrate an optimal resource mix as increasing penetration of solar and other renewable and distributed energy sources introduce a high degree of variability on the grid.

Below OpenEE addresses the specific questions posed in the Paper.

Question and Response

Question 1: Are there any other assessment criteria or relevant considerations which the ACCC should use to determine a preferred model for consumers to access their energy data under the CDR?

The ACCC should consider the experiences of other jurisdictions that have addressed energy data transfer. Mission:Data, an American nonprofit that advocates for customer-friendly energy data access policies, has tracked the successes, challenges, and lessons learned when regulated entities have been required to implement consumer data sharing requirements. We recommend that ACCC take into consideration the ten points included in *Energy Data: Unlocking Innovation with Smart Policy* as it crafts the Australian Energy CDR framework.

Question 2: Having regard to the assessment criteria, what are the advantages and disadvantages of each of the models?

The Paper lays out six assessment criteria and OpenEE provides considerations around each below.

1. **User functionality**. The Paper states that "any solution needs to reduce friction for consumers."[1] OpenEE believes this can be accomplished best with Model 1. Especially when the CDR framework is built beyond the initial priority data, Model 1 would make possible the "one-click" functionality that an ADR would need to provide a full suite of services to a consumer. In contrast, Model 3 would require a consumer to authorize a number of data flows independently. Given the Paper's observation that, "Research suggests significant numbers of energy consumers display low levels of engagement with the competitive retail market,"[2]

   OpenEE suggests that Model 3 cannot be reconciled with the need to provide often disengaged consumers functional simplicity. It has been our experience working with GreenButton Connect systems in multiple states in the United States, that there are substantial transaction costs associated with third parties to build and maintain API connections with each retailer or utility. This dramatically increases the cost and reduces the usefulness of an "economy-wide" approach, while creating an inconsistent consumer experience.

2. **Cost-effectiveness.** OpenEE acknowledges that Model 3 may lessen data and IT system infrastructure build in the public sector, but is likely to be significantly more expensive for customers as each retailer invests in their own custom approaches. Given the long-term nature and implications of the Energy CDR, cost effectiveness should be considered on a long-term basis including its impact on rates to customers. The initial investment in Model 1 to establish AEMO as a centralized data repository with automated update protocols via secure APIs will permanently streamline the simplicity and usability of the Energy CDR for all parties, which will minimize costs and increase benefits over the lifetime of the implementation.

[1] Consultation Paper: Data Access Models for Energy Data, p. 32
[2] Ibid, p. 32

3. **Interoperability.** Regarding interoperability the Paper states "The model adopted for the energy sector will need to facilitate convergence in data-driven services...explicit consent, authorisation and accreditation will be required for each data set."[3] OpenEE expects that as products and services are integrated between sectors, consumers will expect and demand simplicity. OpenEE believes that requiring consumers to provide explicit approvals associated with multiple, disparate energy-related data sets, as may be required with Model 3, would severely limit the potential for effective, customer-friendly interoperability of the Energy CDR.

4. **Efficiency of relevant markets.** OpenEE believes that any of the models outlined in the Paper could in principle help facilitate market competition. However, we caution that the Model 3 scenario would place the greatest reliance on market actors for whom an effective CDR would increase competition. OpenEE's experiences elsewhere have led us to believe that when data sharing responsibilities conflict with core business models, sub-optimal experiences for customers and authorized third parties often result.

5. **Reliability, security, and privacy.** OpenEE believes that data safety, security and privacy are paramount. Care must be taken upon adoption of any of the models to ensure proper handling of data. Methods are established and are continually being enhanced to ensure that large datasets can be transferred via secure APIs and stored safely.[4]

   In addition to the core use case of customer data access, we believe there are a range of technical approaches that can be used to provide privacy to end customers, while still maximizing the value of energy data for public purposes. OpenEE is leading an open-source effort in the United States, governed by the Linux Foundation and Joint Development Foundation as part of the Energy Markets Methods (EM2) project, with the National Renewable Energy Lab (NREL) and funded by the Department of Energy, that is implementing differential privacy to enable usage of data for tracking of efficiency savings at a portfolio level, carbon accounting, and other use cases. This approach, pioneered by companies like Apple and Google, can enable data sharing that does not compromise customer privacy.

   Given Australia's grid and climate goals, we strongly urge that the ACCC take into consideration use cases that go beyond the direct sharing of customer data, including the need for robust data systems to inform fluid grid operations with a high saturation of variable generation.

6. **Flexibility and extensibility.** The need to handle data beyond the initial NEM priority data sets highlights the benefits of a centralised model. OpenEE believes that as the Energy CDR expands, the Model 1 approach will prove to be clearly superior. Model 1 would allow for the establishment of repeatable and automated data transfer for all potential sources of data. In Model 1, the issue of data extraction and transfer, including scheduling and timing, need only be addressed once. An "on-demand" gateway functionality as in Model 2 is more likely to

---

[3] Ibid, p. 33

[4] OpenEE's cloud based servers adhere to the highest internationally recognized security standards including but not limited to: ISO 27001, SOC 1 and SOC 2/SSAE 16/ISAE 3402 (Previously SAS 70 Type II), PCI Level 1, FISMA Moderate and Sarbanes-Oxley (SOX). Adherence to these protocols and standards will ensure that a central data repository to serve the Energy CDR can be created and managed safely and securely.

encounter errors that would lead to ad hoc functional issues as more data sets and sources are managed. For instance, the ability of the gateway to fulfill an on-demand request could be impacted if/when data holders exercise even routine maintenance on internal data systems. Again, Model 3 is an inferior option for flexibility and extensibility as each new data set or data holder would require additional processes and procedures to manage interactions with an ADR. Regarding flexibility, only Model 1 would enable ADR to routinely plan and assess modern grid planning functions (see response to Question 5 below).

Question 3: What are the likely implementation/compliance costs for market participants (including accredited data recipients) under each of the models, including costs associated with IT system changes or data storage?

OpenEE recognizes the urgent need to provide ADRs data to support market innovation, non-wires alternatives to grid infrastructure, and decarbonization targets. While there are implementation and compliance costs for all three models outlined in the Consultation Paper, Model 3, the economy-wide CDR model, introduces much more uncertainty in keeping implementation and compliance costs as low as possible.

Within the parameters of Model 3, the potential for data and vendor lockout, where stakeholders are priced out of the market, goes uncombated. Having to source information directly from existing data holders provides additional layers of complication. If a retailer controls access to data generated from their devices, will they each charge for access and thus disincentivize data sharing, locking out those who want to innovate? Will a standard protocol for sharing data, such as Green Button Connect (see Question 5 for further detail), be selected to help promote the utilization of consistent data? Will each API be different and have unique integration and security requirements? Issues including cost requirements for data exchanges, regulating the secure transfer, and investigating breaches in sensitive data will all have to be accounted for in Model 3.

Question 4: What additional requirements should the ACCC consider including in the CDR rules for the energy sector if the gateway model is adopted?

OpenEE has no additional requirements to recommend.

Question 5: What emerging technologies do stakeholders believe will have an impact on the energy sector with respect to the CDR?

The rapid expansion of solar generation in Australia has created a mid-day dip in demand from traditional resources along with an evening demand ramp that is becoming consistently steeper. Addressing this growing variability on the grid will require more effective and targeted use of distributed balancing resources. Storage, dispatchable demand response, electric vehicles, and energy efficiency can all play an important role in addressing the “duck curve,” but these DERs need to be managed on a time-and-locational basis. As the operator of the Australian National Electricity Market and the Wholesale Electricity Market, AEMO is in the unique position to orchestrate an optimal integration of DERs with traditional supply-side resources. AEMO will need to address decarbonization goals, identify opportunities for targeted penetration of beneficial DERs, and direct load shaping to keep costs reasonable for consumers. To serve these new and vital functions effectively, AEMO will need consistent, reliable, and up-to-date granular data that aligns with the CDR, as well as assistance from ADRs to provide expertise, specialty services, and innovation.

Through the CDR, Australia has the opportunity to establish the framework and invest in the technology that will help facilitate the secure use of energy data. Specific emerging technologies and approaches are being developed and utilized to ensure security and privacy when storing, handling, and transferring large amounts of sensitive data. These include:

Green Button: The Green Button initiative is an industry-led effort that responds to a 2012 US Presidential call-to-action to provide utility customers with easy and secure access to their energy usage information in a consumer-friendly and computer-friendly format for electricity, natural gas, and water usage.

Green Button Connect My Data: The Green Button Connect My Data (CMD) standard is the energy-industry standard for enabling easy access to, and secure sharing of, utility-customer energy- and water-usage data. Utilities providing standards-based Green Button customer-consumption and billing data can provide customers new data-driven services, programs, and platforms; digitally empowering customers with the ability to securely transfer their data to third-party solution providers who can further assist them in monitoring and managing energy or water usage.

SEAT Differential Privacy: The Secure Algorithm Testbed for Energy Data Fusion project (SEAT) [5] is a collaboration between National Renewable Energy Laboratory (NREL), which is serving as the lead laboratory, and the Lawrence Berkeley National Laboratory (LBNL), San Francisco Department of the Environment, and OpenEE. The SEAT was initiated specifically to utilize new data sources to inform operational energy analyses. Please see the Applicable Research section of Question 8 for further detail.

Building Energy Data Exchange Specification: The Building Energy Data Exchange Specification (BEDES, pronounced "beads" or /bi:ds/) is a dictionary of terms, definitions, and field formats which was created to help facilitate the exchange of information on building characteristics and energy use. It is intended to be used in tools and activities that help stakeholders make energy investment decisions, track building performance, and implement energy efficient policies and programs.

<u>Question 6</u>: What are the cost differences to participants of providing data once a day (to an AEMO repository) or on demand?

With effective APIs there should be very little cost difference between a once-per-day uploading cadence compared to updating on an as-needed basis in reaction to on-demand requests. The former would provide a more structured model that ensures the AEMO repository is consistently updated. It

[5] awarded by the U.S. Department of Energy Energy Efficiency and Renewable Energy (EERE) Building Technologies Office (BTO)

should be straightforward for the AEMO, working with other data holders and accredited data recipients, to establish automated data transfer protocols on a daily basis via secure APIs.

<u>**Question 7**</u>**:** What is the competitive impact, if any, of accessing data through AEMO rather than through a retailer?

AEMO has a unique opportunity to enable Australia's demand-side resource, load balancing, and decarbonization markets. Accurate, granular system and usage data as well as secure access to those data are vital for cost-effective decarbonization and behind-the-meter flexibility. However, consistently accessing data has often posed a major barrier for policymakers, retailers, system operators, local governments, and market participants, who will all need to work together to ensure energy remains affordable as Australia modernizes and decarbonizes the energy sector.

AEMO can securely and efficiently facilitate the transfer of differential privacy protected energy data into the market, allowing market innovators to operate in the most cost-effective manner. The Model 1 approach can enable data access to ADRs who stand to provide innovative solutions at an individual-customer and system level. There are a range of important use cases enabled through Model 1 that would better inform policy and create responsive markets without compromising customer privacy, including:

Energy Savings Calculations

Site-based energy savings calculations using open-source methods like CalTRACK[6] require access to pre- and post-intervention consumption data. The growth of Green Button services in the U.S. has made it easier for individual customers to authorize building-level data sharing, but connecting customers through Green Button Connect remains a substantial barrier. These services are also mostly useful for managing active program participants. However, many energy efficiency implementers lack historical metered performance data, making it difficult to bid competitively or forecast accurately as the focus shifts to time-and-locational grid needs. This lack of access to historical performance data stands in the way of financing and insurance agreements and the efficacy of both traditional energy efficiency and demand response schemes and new pay-for-performance based approaches. With Model 1, AEMO and qualified ADRs can provide the analytics needed for scalable energy efficiency targeted toward grid needs.

Energy Benchmarking

Both public sector policy and market actors in energy efficiency, including aggregators, program implementers and contractors, are consistently frustrated by a lack of energy profiling data that would enable them to make operational decisions about which customers are good candidates for energy efficiency interventions. This data availability becomes especially important in performance-based settings, where customer targeting and potential analysis are essential competitive advantages.

In many cases, these market actors may have access to their customers' energy data, but the full value of this data cannot be realized in the absence of context. One way to provide this context is

[6] www.caltrack.org

through the energy consumption profiles for groups of customers of the same type and in the same geographical location.

These energy profiles can be represented by metrics that are derived from consumption data, adjusting for weather and occupancy. These metrics, usually comprising coefficients of statistical models and uncertainty values, provide a wealth of information about load profiles (e.g. heating load, baseload and cooling load) and energy use predictability (whether the building's energy consumption can be sufficiently represented by a statistical model).

Comparison Group Savings

Where possible, researchers utilize Randomized Control Trials (RCT) to evaluate the impact of interventions, especially where whole populations might be involved in a trial. However, in many cases, implementing an RCT design is not practical. A more realistic research design may involve selecting a similar set of untreated buildings and differencing the consumption between the treated and untreated sets. Moreover, access to data for groups of untreated customers enables longitudinal tracking of energy impacts by aggregators, program implementers and administrators, net of population-level changes. This type of tracking can empower performance-based schemes by enabling risk quantification and mitigation.

The process of establishing a comparison group often requires matching the load shape (consumption) profile of the treated building to a set of untreated buildings within the same geographical area (as well as potentially other characteristics). Generally, a minimum of five similar, untreated buildings are selected for each treated building. This process requires access to individual building consumption data from non-treated buildings. If individual building owners in non-treated buildings were required to provide permission to access their data, this research design approach would be practically infeasible.

Building Model Testing

As building models are increasingly utilized to predict the impact of distributed energy resources on grid resources, gaining insight into the actual load shape of targeted buildings increases accuracy of forecasting and planning. In particular, it is desirable to significantly improve upon the artificial load shape profiles that are currently available for different types of buildings (e.g., the load shape of a typical laundromat).

Building Characteristics

Certain attributes associated with building energy usage can be estimated algorithmically. For instance, algorithms of varying complexity and accuracy can detect solar PV installations, recently purchased electric vehicles, or presence of periods of with uncharacteristically high usage. These attributes could be used to target schemes to customers who can benefit the most from a specific intervention.

Load Shape and Carbon Accounting

A city, local government, or aggregator would like to know the average monthly load shape for a sector or targeted portfolio to understand consumption patterns or estimate carbon footprint. 24-hour load shapes that represent broad patterns over the course of a months or across many buildings can be useful even in the absence of more granular - and more sensitive - full time series of AMI data.

<u>Question 8</u>: Are there any other issues that stakeholders wish to raise?

OpenEE reiterates our recommendation that the ACCC adopt Model 1 described in the Paper for a secure and centralised data platform managed by AEMO. This approach most effectively supports the objectives outlined in the CDR. With the growing demand for data in all forms, AEMO is best suited to serve as the central data holder that can safely and securely share CDR data directly to ADRs. This approach lends itself to simplicity, reducing potential complications experienced by individual consumers.

Additionally, Model 1 has the potential to better position AEMO in remaking Australia's power infrastructure into a modern energy grid. As stated in their 2018 report, *Operational and market challenges to reliability and security in the NEM*, AEMO has a well documented understanding of the rapid transformation currently taking place in Australia's energy markets. A changing mix of energy resources to include renewables like solar and other DERs is already presenting a number of grid balancing issues, exacerbating system ramps to meet peak demand, and increasing the risk of disruption to service. AEMO has stated that data driven processes such as improved forecasting, strategic deployment of DERs, and better systems integration will bolster its ability to properly maintain the National Electricity Market. These proactive options can only work with the proper capacity to use readily available energy data at an increasingly granular level.

Without jeopardizing the security of Australian consumers' data, Model 1 has the potential to provide the most benefits to all stakeholders.

<u>Applicable Research</u>: *Creating a secure testbed for algorithms fusing public and protected data to predict energy use and potential savings*

OpenEE is part of a group currently working on grant-funded research to develop a Secure Energy Algorithm Testbed (SEAT) that allows energy algorithms to run on multiple public or protected data sources to generate useful outputs while protecting secure information.

This multi-year project develops the Secure Energy Algorithm Testbed (SEAT, described in the response to Question 5) that can merge data from multiple sources and allow algorithms, including Open Source CalTRACK / OpenEEmeter based advanced measurement and verification (also known as M&V 2.0) calculations, to run on protected data while ensuring that sensitive information remains secure. This open source effort is part of the Energy Market Methods Consortium under the umbrella of the Linux Foundation Energy, and Joint Development Foundation.

The SEAT system will support the growing number of M&V 2.0 applications evaluating energy savings from installed upgrade measures. Additional applications will be unlocked by giving researchers access to sensitive real-world data in a framework that allows algorithms to transition quickly from benchtop prototypes to production use.

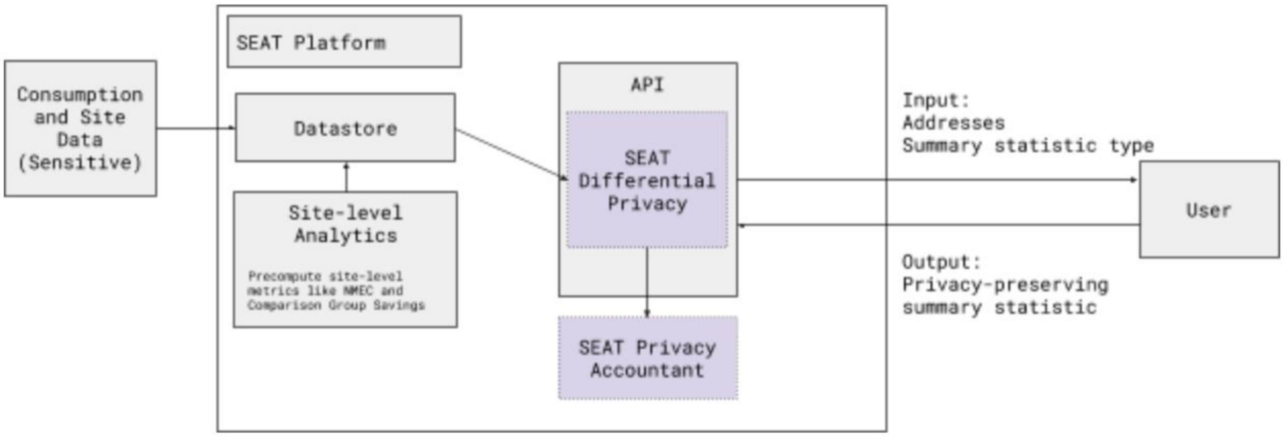

Figure 1: SEAT Architecture Diagram for a typical deployment

Explanation of Privacy Preserving Mechanisms

A design based on Differential Privacy is proposed for SEAT. The mechanism underlying this technology was first described in 2006 by Cynthia Dwork and has received extensive academic treatment since then. There have been a number of industry deployments, notably by Uber and Berkeley in 2018 for running arbitrary SQL queries against rider data [Github], a project that was motivated by GDPR.

The basic workings of Differential Privacy are intuitive to understand. Imagine that you would like to find out the average annual energy consumption for a group of buildings. These consumption values would range from zero to some maximum, C_{max}. One way to protect the privacy of the buildings in this query would be to add noise to the result such that it was not possible to ascertain the contribution of an individual building, even if you knew the consumption of all the other buildings. The maximum amount that a single building can contribute to the average, ΔA_{max}, is

$$\Delta A_{max} = (C_{max} - 0) / N$$

where N is the number of buildings in the query. You would want to add about ΔA_{max} noise to the resulting average to hide the contribution of any building to the overall average.

Differential Privacy provides a framework for generating this noise. An additional parameter, ε is introduced to quantify the privacy guarantees. Lower epsilon values result in higher privacy guarantees at the cost of usefulness of results.

The most common Differential Privacy mechanism is the Laplace Mechanism. In our example, the noise would be drawn from a Laplace distribution with mean zero and parameter *b* equal to

$$b = \Delta A_{max} / \varepsilon$$

The probability distribution function would look something like the following, where many values of noise would come from around zero, but there was a reasonable chance that values further away from zero were chosen.

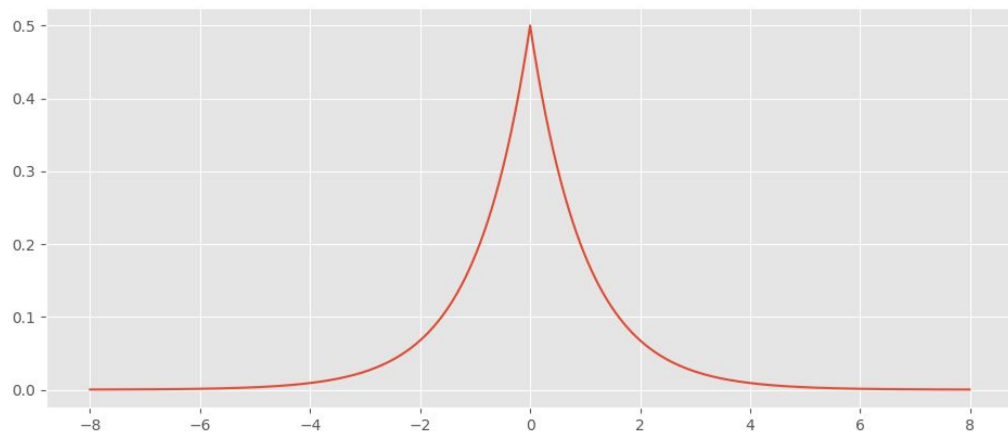

Figure 3: Example Laplace distribution

Finally, to further make the case for the relative simplicity of the Laplace mechanism, here is an example implementation in Python for computing a differentially private mean:

```
import numpy as np

s = np.genfromtxt("consumptions.txt", delimiter=",")

N = len(s)
df = (max(s) - min(s)) / N
e = 1.0
b = df / e

def differentially_private_mean(s):
     return np.mean(s) + np.random.laplace(0, b)
```

For an approachable source for additional details about the Laplace Mechanism see Orazio et al. *Differential Privacy for Social Science Inference*.

Differential Privacy research provides tools for extending this approach to more complex queries and datasets as well as additional guidance about how much noise to add. A number of existing systems exist for Differential Privacy for SQL queries, such as Microsoft's PINQ (Privacy Integrated Queries) and Uber/Berkeley's SQL Differential Privacy Query Rewriter.

Privacy Budget and Risk

A fundamental concept in Differential Privacy is the “privacy budget”. Every anonymization technique, from k-anonymity, to the 15/15 rule, to Differential Privacy, reveals some amount of information about the individuals in the dataset. For example, imagine if one were able to find out the consumption information for 14 of the 15 individuals in a dataset anonymized by the 15/15 rule; it would be possible to then deduce that final participants energy usage. This sort of attack is not only theoretical – Netflix’s de-identified movie review database was re-identified using publically available IMDB data. Even more shocking was the re-identification of Maryland patient data, which was de-identified in accordance with HIPAA standards, but turned out to allow researchers to uncover the health information of the Governor when cross-referenced with public voting records. In contrast to other privacy techniques, Differential Privacy quantifies the risk of each individual contained in a dataset *no matter how much additional data is released*. This risk is quantified as ε, the privacy parameter.

As part of the development of the SEAT platform, OpenEE will evaluate the privacy risks of configuration of each of these modules against real consumption datasets in order to provide guidance to administrators. As confidence and understanding of the privacy approaches used by SEAT grows, additional statistics that require the more advanced Noise Modules and Privacy Accountant can be deployed.

Summary Statistic Design

The basic architecture of the SEAT platform anticipates returning privacy-preserving summary statistics for groups of sites. The following table maps Use Cases to candidate summary statistics.

These candidate statistics will be explored in detail in proof-of-concept and prototype development. In particular, the required accuracy for useful outputs will be characterized. It is possible that not all proposed summary statistics will be possible to implement because of high sensitivity or small sample sizes. However, initial analysis suggests that it should be realistic to implement many of them, and the analysis used to characterize this initial batch of use cases will allow the rapid development of further use cases. Appendix A includes details of our analysis characterizing the application of the Laplace Mechanism to savings and consumption data.

Summary Statistic	**Use Case**
Average Monthly Energy Savings	Energy Savings Calculations Targeting
Average Annual Hourly Traces	Building Model Testing Load Shape / Carbon
Monthly Comparison Group Savings	Comparison Groups
Model Coefficients	Energy Profiling and Benchmarking

Appendix A: Applying Laplace Mechanism for Differential Privacy of Energy Analytics
Here we attempt to characterize the error introduced by the Laplace Mechanism for two representative queries to the SEAT system.

Our analysis considers computing the differentially private mean of two quantities: NMEC savings and average hourly consumption.

The specific algorithm for computing differentially private means is detailed in Orazio et al. *Differential Privacy for Social Science Inference*, with the relevant section copied below: