INVITED PAPER

Review and Perspectives on Data Sharing and Privacy in Expanding Electricity Access

By Jonathan T. Lee, *Graduate Student Member IEEE*, Jordan Freitas, *Graduate Student Member IEEE*, Isa L. Ferrall, *Graduate Student Member IEEE*, Daniel M. Kammen, Eric Brewer, *Member IEEE*, and Duncan S. Callaway, *Member IEEE*

ABSTRACT | Increased sensing and data collection in electric power systems from utility to minigrid to individual household scale are resulting in an explosion of data collection about users and providers of electricity services. In the push to expand energy access for poor communities, the collection, use, and curation of these data have historically taken a back seat to the goal of expanding energy access but are increasingly being recognized as important issues. We review the nascent literature on this topic, characterize current data management practices, and examine how expanding access to data and data sharing are likely to provide value and pose risks to key stakeholders: end users of electricity, microutilities, macroutilities, governments, development institutions, and researchers. We identify the key opportunities and tensions and provide recommendations for the design and implementation of new data-sharing practices and platforms. Our review and analysis suggest that although a common and open platform for sharing technical data can mitigate risks and enable efficiency, fewer benefits are likely to be realized from sharing detailed financial data. We also recommend codesigning practices with each stakeholder group, increasing legal protections for end users of electricity and using deep qualitative data in addition to quantitative metrics.

KEYWORDS | Data privacy; electric utilities; energy management; open systems; power system planning; smart grids; smart meters; social implications of technology; sustainable development.

I. INTRODUCTION

The challenge of expanding access to electricity to all people on earth is being addressed at the same time as the paradigms for electricity service and communication are rapidly changing worldwide. As new technologies and electricity products are being introduced as alternatives to the traditional grid and utility models, sensors, communication, data processing, and automatic control are becoming ubiquitous and essential in infrastructure systems. This is leading to a massive increase in collected information and an associated set of new values and risks to the stakeholders involved. Social concerns about data security, privacy, information asymmetries, and increasing automation are much broader than electricity systems, but debates over automatic metering infrastructure and energy management aspects of "smart homes" are prime examples of these concerns in countries with widespread existing electricity infrastructure. Importantly, in countries with widespread electricity infrastructure, by nature of the dependence of information technology on electricity, these debates around data in electricity systems arise in the context of the broader conversation about the appropriate use of information. In countries without widespread electricity infrastructure, access to electricity is tightly

Manuscript received August 21, 2018; revised February 5, 2019; accepted May 16, 2019. This work was supported by the United States National Science Foundation through the CyberSEES Program under Award 1539585.
(Corresponding author: Jonathan T. Lee.)

J. T. Lee, **I. L. Ferrall**, **D. M. Kammen**, and **D. S. Callaway** are with the Energy and Resources Group, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: jtlee@berkeley.edu; isa.ferrall@berkeley.edu; kammen@berkeley.edu; dcal@berkeley.edu).

J. Freitas and **E. Brewer** are with the Department of Computer Science, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: jordanfreitas@berkeley.edu; brewer@berkeley.edu).

Digital Object Identifier 10.1109/JPROC.2019.2919306

coupled with access to information. Therefore, expanding electricity access is also expanding access to information technology and likely implies a large increase in (in some cases, the first instance of) the harvesting of data through sensors in people's lives who are newly electrified.[1] The methods of collection, use, and curation of these data have critical ethical concerns with significant potential risks and opportunities but are often overlooked as efforts by governments, development institutions, researchers, microutilities, macroutilities, and users of electricity are focused on expanding electricity access with financial and technical sustainability. As such, there is little academic literature addressing the nexus of data management and electricity access. However, the need for guidance and policy on this issue is recognized in industry [5], and the role of data is likely to become increasingly significant as electricity service providers expand their capacity to leverage data for technical efficiency, new electricity markets, cheaper finance, and bundling complementary products and financial services with electricity.

This paper makes the case that there is both value to be gained and associated risk from the use and sharing of different kinds of data associated with electricity service (in the context of expanding access to electricity in primarily less developed regions) and outlines important guiding principles in the use and sharing of these data. We examine these values and risks from the perspectives of different stakeholders, identify opportunities and tensions from the alignment or misalignment of their incentive structures, and review and outline important guiding principles for data-sharing frameworks and practices. As there is yet little academic literature on data sharing and privacy in the context of electricity access, we synthesize literature from these domains and draw from industry and institutional reports, anecdotes, and analogs in fields such as public health to construct a framework for continuing research and developing effective data management policy and practices. We also include primary evidence from an energy access workshop held in July 2018 in Nairobi, Kenya, where participants in a small focus group representing the stakeholders described below shared their perspectives (see Appendix A). Our mapping of types of data, stakeholders, and their incentives provides an overview of the critical issues and technology options and serves as a foundation for further engaging stakeholders in developing standards and framing the design of data management systems.

[1]It is important to note that access to the Internet and digital communication has rapidly outpaced electricity through mobile phones [1], [2, p. 127] and people often have access to phone charging through shared use and charging without direct access to electricity services [3]. Thus, people may already have an online profile, for example, without electricity access. However, access to more electricity typically enables greater access to mobile phones, computers, and other data-collecting appliances and introduces usage metering and payment collection in many service provision models from classical postpay to prepay [2], [4].

We group the data in question into three categories: technical, financial, and demographic. Technical data are physical measurements that can be further divided into end user of electricity data (i.e., energy usage and load profiles) and system performance data (i.e., quality of service metrics such as power quality and reliability, and product longevity). Financial data can also be grouped by the end user of electricity (payment records) and by the service provider (itemized costs of providing service). Demographic data are socioeconomic data about the end users of electricity and/or their households and communities they live in. This can be thought of as data that could appear on a census, data that might be gathered to understand the relationship between electricity usage to household dynamics or local economies and geospatial data. The possibilities and ramifications of how all of these data could be linked or anonymized are within the scope of this paper.

We group the stakeholders into six groups: end user of electricity ("user" for brevity), microutility, macroutility, government, development institution, and researcher. While they are mostly distinct, there can be overlap (e.g., a state-owned utility or a stakeholder, such as a researcher or utility, sponsored by a development institution). We also sometimes jointly refer to microutilities and macroutilities as service providers. There are additional stakeholders and further delineations that could provide important insight in further study, such as investors in service providers, complementary industries, and the differing goals or legal obligations of particular stakeholders. Although there can be exceptions, we assume, for our analysis, the following loose definitions of roles.

1) *User:* An individual, business, or some other social organization who pays for electricity as a service.
2) *Microutility:* A private corporation that generates energy at the household or community scale and sells directly to users, e.g., Pay-As-You-Go (PAYGO) home solar providers, such as M-KOPA, BBoxx, Off-Grid (Zola) Electric, and Mobisol, and minigrid companies, such as Rafiki Power, PowerGen, and Gram Power (see [4], [6], and [7] for review and analysis of these prominent companies and their business models).
3) *Macroutility:* A public or historically regulated utility. If it is restructured, we mostly refer to the distribution company.
4) *Government*: The body having jurisdiction over electricity infrastructure and service providers. This is typically a national government though, in some cases, the relevant government could be multinational (e.g., a power pool of countries that share grid infrastructure) or subnational (e.g., states in India).
5) *Development Institution:* A global-scale institution with the mission of improving human quality of life, e.g., multilateral groups such as the World Bank, the United Nations Development Programme, Sustainable Energy for All, and the International

Renewable Energy Agency, as well as unilateral aid agencies such as United States Agency for International Development (USAID) and GIZ, NGOs, and philanthropic foundations.

6) *Researcher:* An individual or group seeking to provide insight into technical and social aspects of expanding electricity access and development. Researchers could belong to universities or think tanks, for example, with the latter in some cases blurring the line with development institutions. Although their object of study is typically influenced or directed by other stakeholders, their risks and value from data usage are unique and tied to publication and expertise within the field.

We use the concept of sharing as the movement toward more data exchange among stakeholders, more open access (not necessarily fully public), and more standardization, transparency, and guidelines for the content and provenance of the data being shared. We make the case that there is a value to be gained relative to current practices by each of the stakeholders from sharing, but they each also face risks to varying degrees dependent on the category of data and the data sharing architecture. Important emergent themes from mapping stakeholders' incentives are as follows.

1) While reducing information asymmetries between groups of stakeholders tends to provide net benefits and efficiencies to the groups as a whole, there will likely be winners and losers within the groups.
2) Standardization is a key to realizing many potential values but inherently sacrifices nuance and context, raising limitations to the actions that can be justifiably taken and requiring ground-truthing and critical qualitative analysis of the data.
3) Codesign of practices is necessary for usability and risk mitigation.
4) Privacy and anonymity mechanisms can protect stakeholders from some risks, but protective policies, education, and critical review of practices and decision-making are important elements of a mutually beneficial data-sharing environment.

II. OVERVIEW OF CURRENT PRACTICES AND DATA-DRIVEN DECISION-MAKING

A. Data Collection

Currently, on large scales, service providers are the dominant collectors and owners of technical and financial data, along with some demographic data, while governments and institutions are the dominant collectors and owners of demographic data through official records and large-scale surveys. Researchers variously collect all categories of data in their study sites. Service providers employing the state-of-the-art practices collect technical data with smart meters on time scales of typical minutes,[2] and data are transmitted either instantaneously or in batches to remote databases. In some cases, technical data are stored locally at the meter or on a local server and periodically retrieved manually. Service providers collect user financial data by recording payments, often through a mobile money provider. The relationship between service providers and mobile money providers is significant: macroutilities are increasingly using mobile money and microutilities are both some of the largest mobile money recipients in Africa and the reason for many users opening accounts.[3] Service providers also often collect demographic data during the customer registration or design process, which in the case of minigrids can be quite extensive.

The practices and standards of data collection differ by stakeholder. Large-scale demographic surveys conducted by international groups, such as the World Bank,[4] have standard practices for enumeration and are aimed to be uniform; however, government records may differ by jurisdiction, and there is no standardization of demographic data across utilities. Technical data are collected based on the service providers' capacity and business needs; there is no enforced standard that applies to microutilities and macroutilities. The underlying sensor technology is subject to accuracy and communication standards depending on jurisdiction, but the structure of stored data and the computation of aggregate metrics are fragmented at the individual business level (and in some cases by the version of the technology used when an electricity system was built). Many of the microutilities referenced earlier employ proprietary software for data management and analytics described on their websites but, in no cases, are specific details on the implementation made public nor do they mention data privacy or sharing platforms on their websites as of 2018. No standardized system or practice has emerged. Stakeholder interviews are needed to more fully characterize current practices and perceptions, which is an ongoing research effort by the authors. As a contrast to microutilities in developing regions, regulated macroutilities in more developed regions have a longer track record of data collection and use, though still without standardization. Sample deidentified sets of electricity meter data have been made available for researchers by private companies (e.g., Pecan Street, Inc. [11] where access is gained through a request form) or through agreements between researchers and utilities [12]. In another case, employees of Google were recruited to have their consumption metered and shared through Google proprietary software [13], [14]. Some macroutilities share

[2]In some cases, adjustable time scales can be remotely programmed, with data capture on the order of seconds.

[3]The Consultative Group to Assist the Poor (CGAP) estimates that 30%–50% of solar microutility customers outside of Kenya opened a mobile money account to pay for service [8]; solar-home-system company Fenix International was the third largest consumer of mobile payments in Uganda and PEG was the largest in Ghana for telecom MTN in 2016 [9]; macroutility (including Umeme) revenue share from mobile money payments in Uganda increased from 1.1% in 2012 to 27.5% in 2017 [10].

[4]For example, the Microdata Library http://microdata.worldbank.org/index.php/home.

statistics and comparison-based reports with their customers: in [15], the data management company Opower participated with researchers to study the effect of this data sharing on energy consumption.

Demographic and financial data collected by researchers are typically subject to Institutional Review Boards, which provide some protection for users, and the scientific community enforces the standards of data collection and transparency to a degree.

See Appendix B for a summary of industry groups and institutions that maintain databases relevant to this field and to what extent they have data policies that address sharing, collection, and privacy. We find that the industry groups generally do not treat sharing and privacy as explicitly as multilateral institutions.

B. Use of Data in Global Development

The appeal of data-driven monitoring and evaluation approaches to development stems from a desire to iterate interventions and data collection strategies rapidly and then to draw more sound conclusions from the analysis of impact than seems otherwise possible based on experiential evidence and stories. Trends in data may reveal unrealized strengths and weaknesses of interventions or validate known issues and opportunities. There is a wealth of literature on the history and consequences of metrics such as those used to measure progress toward the Sustainable Development Goals [16], [17]. The more intentional data sharing can be in terms of mitigating risks and enabling value, the more conclusions can be collaboratively derived, debated, and used to inform consequential decisions.

C. Data Access and Sharing

We consider "access" in this case as the ability to derive benefits from categories of data, using the definition of Ribot and Peluso [18], which is more general than having legal rights. Important in this theory of access is who has control of the data via technical and social mechanisms. As it is clear from the current collection practices and our categories of data, the data in question are about the users and the service providers, with the latter collecting and maintaining control of technical and financial data. In some cases, the user relinquishes the control of demographic data (along with monetary payments) to service providers in exchange for electricity service or to development institutions or researchers freely or in the hopes of aid. While the user retains control of personal data that are kept private, they are unable to exercise control or restrict access once it is collected by other stakeholders. In contrast, service providers are able to control other stakeholders' access to their and their customers' data. There are exceptions to this pattern, such as technical data on reliability that can be crowd-sourced independently of a utility, for example, the "GridWatch" platform [19] that uses plug-in sensors to track outages or the World Bank Enterprise Surveys on infrastructure that record the experience of businesses and are published online [20]. Additionally, satellite imagery is being used and controlled by private companies and governments to estimate demographic data [21]. Even when those who control the data allow access by other stakeholders, it is not guaranteed that they will be able to derive benefits from it, for reasons that can include lack of appropriate technology, capital, institutional capacity, knowledge, or access to markets [18].

The status quo of access to data is important for understanding how changes will affect the various stakeholders. Although it could be argued that stakeholders should retain control of data that are collected by or about them, we think that it is more useful to study how transferring or expanding access and control results in an exchange of value and the loss or derivation of new value and risks. Through this lens, we can predict how possible changes to the status quo disparately affect stakeholders, and thus, which changes are mutually beneficial, which are likely to produce tension or be blocked by tensions, and which require policy or technological interventions for benefits to be realized. Throughout the scope of this paper, the reader can extend our methods to ask questions about whether different stakeholders are aware of, or consent to, the value that is being exchanged under the current and potential future paradigms and whether such exchanges are ethical.

III. VALUES AND RISKS FROM DATA

The data policies of both the United States Agency for International Development and the United Kingdom's Department for International Development recognize that there can be potential risks alongside the potential values of sharing data. Expanding beyond this recognition, here, we elaborate the key potential values and risks to each stakeholder group for each category of data.[5] We give an overview of the current access and control of data and discuss possible changes and how the values and risks are dependent on data sharing practices, technology, and policy. Certain stakeholders have similar values and risks to each other, for example, microutilities and macroutilities share many in common, and some stakeholders face values and risks that are the same across all data categories. In particular, those issues faced by researchers span categories of data and are summarized here first, outside of the categories of data.

Access to data is fundamental to analysis in research, giving researchers a strong incentive to advocate for increasing access. While a survey of academics showed that most state their research would benefit by having better access to published data and most are willing to allow access, a large number of respondents still do not publish their data for reasons that include lack of standardization in the publication and citation process, lack of resources for management, and ethical and legal concerns around

[5]A more detailed table of the data collection, sharing, and privacy policies of noteworthy stakeholders can be found in Appendix B.

protecting the privacy of subjects [22]. In most cases, researchers gain more recognition from the publication of analysis than that of data, which introduces an incentive to maintain control of data sets in order to generate analysis publications; an incentive that is likely to persist until more recognition and standardization are given to data publication and citation. This problem requires action and change within the research community, but the lack of resources for data management, privacy concerns, and publication standards can be addressed together with other stakeholders and the academic community and open-access journals at large to facilitate sharing. While by-in-large, researchers stand to benefit from having better access to data, they face the risk of working with inaccurate or improperly collected data and potentially lose understanding of the context of data if they are not involved in the collection effort.

A. Technical Data

For technical data, we consider usage data (effectively meter data) and performance data (e.g., power quality, reliability, efficiency, and product lifetime). Usage data are typically recorded on time scales of minutes to hours, but state-of-the-art meters can have their sampling frequency remotely updated to capture high-resolution data on the order of seconds for periods of time. Installing robust sensor networks for collecting these data in energy-poor communities often faces the challenge of lack of wired communication infrastructure. In response, several start-up companies have metering products specifically designed for these contexts.[6] Typically, a local sensor network for usage meters is installed simultaneously with the electricity service using Wi-Fi (IEEE standard 802.11), long-range-radio (LoRa), or meshed networks. Power-line communications have been proposed but are not widely used. A data aggregator typically mediates communication between the local network and remote servers over a backhaul, though in some cases individual devices are equipped with cellular, SMS, or satellite backhauls and the local network is bypassed. Cellular backhauls of different generations (GSM, UMTS, and LTE) are commonly used as cellular coverage is often available. To reduce cellular data transmission costs, point-to-point microwave communications are sometimes used between remote sites and locations where there are wired backhauls. Poor communication reliability and more efficient data transmission are addressed by storing measurement data locally and transmitting batches of data intermittently.

Usage data can be strictly anonymized, but as discussed next, various benefits of these data require some linkage between financial and demographic data, and additional information can be inferred from usage data alone. Performance data can, in some cases, be inferred from usage data (e.g., meters will show power outages) but, in most cases, are measured by the service provider using dedicated sensors and are aggregated by service provider, geography, and so on. Key potential values and risks from technical data to each stakeholder group are shown in Table 1 and further elaborated here.

The user stands to benefit from better access to their usage, as they currently have limited access to information recorded by their meter. While traditional macroutility smart meters display instantaneous metrics, it is generally a burden for the user to constantly observe this. If the detailed historical data are available, it usually requires a computer and a level of mathematical literacy to derive significant benefits. Microutility meters vary in their interfaces, using display screens, SMS, or mobile interfaces that typically provide current information on battery state of charge, account balance, and instantaneous power consumption but do not provide historical data. A study in Nicaragua found that users were willing to sacrifice monthly cash payments for energy consumption information reports that they were able to codesign with researchers. The users then adjusted their usage and reduced consumption to save money accordingly [24]. These energy efficiency benefits can pose a loss-of-revenue risk to utilities using volumetric pricing, but this risk should both be considered small relative to opportunities for increasing the customer base through more resource efficient usage and a challenge that can be addressed by decoupling profits from energy sales. Beyond the direct value of efficiency and controlling usage, participants in the Nairobi focus group indicated education, which we interpret as better energy and technology literacy, as a positive impact from expanding users' access to data.

Service providers can also benefit from pooling usage data with one another to create larger data sets that can be used for operation and planning. This practice is used in North America, where generators and transmission system operators report reliability metrics to the North American Electric Reliability Council (NERC), which publishes data and assessment reports that address reliability, system supply and demand, and distributed energy resources.[7] Combining usage data with performance data (and further with demographic and financial data) can enable deeper understanding of how power quality and reliability affect how people use appliances in different contexts and how system performance metrics impact development indicators and economic growth. Machine-learning approaches for forecasting electricity demand in real time, as well as estimating demand and demand

[6]Examples companies are SparkMeter (https://www.sparkmeter.io), New Sun Road (https://www.newsunroad.com), and SteamaCo (https://steama.co), see [23] for an example open-source design and technology discussion.

[7]See the NERC Reliability Assessment and Performance Analysis Group website at https://www.nerc.com/pa/RAPA/Pages/default.aspx, and the Demand Response, Generation, and Transmission Availability Data Systems (DADS, GADS, and TADS, respectively) for a description of the data and example reports.

Table 1 Values and Risks From the Technical Data

Stakeholder	Values	Risk
User	• Usage management and energy efficiency • Consumer information / product differentiation • Fair pricing • Better quality of service • Education	• Privacy / load identification • Invasive advertising • Exploitative pricing • Focus on most affluent customers
Micro-utility	• Power system planning, design, and operation • Performance assessment and comparison to competitors • Load identification • Streamlined permitting and regulation • Product differentiation	• Competition with other service providers • Cybersecurity • Intellectual property theft • Invitation for regulation • Reduced consumption • Product differentiation • Responsibility for protecting data
Macro-utility	• Power system planning, design, and operation • Load identification • Streamlined permitting and regulation • Product differentiation	• Competition with other service providers • Cybersecurity • Intellectual property theft • Expectation of data utilization • Invitation for regulation • Reduced consumption • Product differentiation • Financial and service barriers to low-income or first-time users • Responsibility for protecting data
Government	• Enforce regulations • Quantifiable, evidence-based, progress tracking • Objectivity • Constituent benefits in quality of service and education	• Pressure to act (potentially based on malaligned indicators) • Expose poor performance or bias • Attacks on physical infrastructure
Development Institution	• Inform investment • Performance assessment • Progress tracking • Increased return on investment in the form of research output	• Pressure to justify priorities • Expose poor performance or bias
Researcher	• Data are fundamental for analysis • Replication of findings • Overcome financial barriers to data access • Streamlined data publication	• Security in IRB approval • Losing first rights to study publication • Responsibility for protecting data • Inaccurate or non-rigorously collected data • Lack of understanding context of data

growth, benefit from larger data sets.[8] Multiple studies suggest that these machine-learning approaches perform better than bottom–up survey models, though they require some demographic data as features [25], [26]. A review of electricity planning models[9] for the developing-world context finds that "a structured data gathering and sharing system can contribute to the enhanced accuracy of the [electricity planning models], as well as the effectiveness of the resulting policies," particularly through improving electricity demand estimates [31]. Broader participation in these planning efforts and access to the underlying data could reduce competition between microutilities and macroutilities by improving assessments of which geographies are more cost-effectively served by different technologies and providers. While service providers would derive benefits from sharing these data with each other, restricting access provides barriers to entry against competitors. Local knowledge and experience with particular communities and regions is critical and takes years to develop. While much of the important knowledge is cultural and social, releasing data about demand and usage would be to give away some of the competitive advantages that service providers hold.

Service providers and users can also gain insight into appliances and usage from fast time scale data using machine-learning techniques known as "nonintrusive load monitoring" or "load disaggregation" [32]. These techniques can be used to improve energy efficiency and assess the capacity for demand response but can also create inferences about users' private behavior. This behavioral knowledge could be used to infer financial and demographic information, for example, the purchase of a new appliance or the change in usage of a key income-generating appliance, which could in turn be used by the service provider to set exploitative pricing or sold for targeted advertising. For example, from the 2018 Lighting Global Off-Grid Solar Market Trends Report:

[8]This is not to suggest that larger data sets are always a substitute for good data provenance. It is plausible that a relatively small, simple random sample would be more useful than a larger one with unclear collection bias.

[9]See [27]–[30] for examples of such models.

> "There exists a potential risk around the all-important data assets that PAYGO companies are building. The level of data collected can reduce a customer's bargaining power and lead to predatory behavior. If data can predict that a wedding is coming up in a village, as noted above, bombarding people with advertisements for a wedding loan (to make the wedding grander) could be the next step. Some investors would see that as a measure of progress, and potentially replacing loan sharks, while others would consider it irresponsible" [5, p. 11].

Information that is politically and/or legally sensitive could also be vulnerable. Religious affiliation could be inferred by appliance usage during holidays and criminal activity could also be detected. Not only can the use of television be detected but also can the channel being watched [33]. Service providers should also have concerns around their intellectual property and system security: proprietary control algorithms and system vulnerabilities could be inferred from usage and performance data, especially if it is geographically identified. Even when the data are not directly linked, relationships can be inferred. For example, if service providers report a power outage at a specific time in a specific region, the outage will be reflected on user meters, allowing the region of the meter data to be inferred, which can be linked with public census data or satellite imagery to build a profile of the user connected with the meter.

While key benefits and risks to sharing and expanding access to usage data are tied to better learning algorithm performance through larger data sets, sharing of skill capacity and structural privacy benefits are also important. By expanding access to other stakeholders, service providers can leverage analysis skills that they may not possess in-house. Focus group participants indicated specifically the potential for researchers to use data to assist in power system design and site selection and to increase the value and accessibility of data. From the authors' experiences as researchers, these collaborations happen but are implemented under individual non-disclosure agreements and use *ad hoc* methods for transferring data that place a burden on service providers' personnel and the researchers and do not necessarily use best practices for security. The connection of usage data to personally identifiable information (PII) should be treated with great caution. Best practices and privacy mechanisms can be more easily implemented and enforced by using a system to analyze data on behalf of individuals, and greater anonymity is possible as data sets increase in size.

The issue of product differentiation touches many stakeholders. Lighting Global finds that low-quality electricity products are a major barrier to the industry at large and to users [5]. Standardized reporting of the technical performance data by product and service provider that is accessible to users can allow them to make informed decisions and reduce information asymmetries that lead to inefficient transactions. Aid agencies and investors could use these data to direct funds to better "bang for their buck" products. Within service providers—especially microutilities in direct competition with each other for customers and investors, those who have better performance metrics will benefit from differentiation, while those who have worse performance will be hurt. Thus, many providers have an incentive to restrict access to technical performance data. Furthermore, even those companies who have better performance may not know where they stand relative to competitors because their competitors' data are private, and status quo bias could dissuade them from expanding data access when the risks and benefits are uncertain. Similar to product differentiation, governments could use standardized reporting of performance metrics to streamline permitting and regulation, something that would also benefit some service providers, but potentially invite more regulation or hurt those incumbents who are able to take advantage of the current permitting and regulatory system.

Transparency and standardization in reporting practices are critical and currently lacking in this space, resulting in huge discrepancies in reported metrics and mistrust among stakeholders. For example, researchers found that in 109 primarily low- and middle-income countries utilities report average customer outage time to be only 15% of what customers report[10] [34], a discrepancy that could result from flawed incentives, but could be resolved through data transparency. Ground truthing of reported connections in Kenya calls into question the official statistics on electrification rates [35]. Participants in the Nairobi workshop indicated the improved quality of services as a potential benefit to users, which we infer to come in part from greater accountability. While users, development institutions, some service providers, and governments should benefit from better reporting of these metrics, there is the potential for some service providers to lose credibility and also for the exposure of corruption or ineffectiveness in some governments and development institutions. This can lead to opposition to transparent policies and, in some cases, can cause significant harm to individuals or institutions [36]. Conversely, increased transparency and open-access data allow governments and development institutions to better defend their objectivity.

B. Financial Data

For financial data, we consider specifically records of transactions for electricity and the costs of providing service. Key values and risks to different stakeholders are shown in Table 2.

The tracking of payment for electricity systems, especially those sold on credit, is often the first banking relationship and form of credit history for many users. This can be of great value to users and can open up new access to

[10] 15% is the slope of the best fit line relating utility reported to customer reported System Average Interruption Duration Index (SAIDI).

Table 2 Values and Risks From the Financial Data

Stakeholder	Values	Risk
User	• Access to financial services • Access to better rates with good payment history	• Exploitative pricing • Poor credit score • Disqualification from service • Service discrimination across socioeconomic spectrum
Micro-utility	• Assess customer financial risk • Understand customer willingness-to-pay • Sell customer data • Planning models	• Competition with other service providers • Losing revenue from selling financial data
Macro-utility	• Assess customer financial risk • Understand customer willingness-to-pay • Planning models	• Competition with other service providers • Customer defection
Government	[Same as technical data]	[Same as technical data]
Development Institution	• Inform investment • Performance assessment • Progress tracking	• Expose poor performance or bias
Researcher	[Same as technical data]	[Same as technical data]

financial services or lower the cost of PAYGO systems for users with strong payment histories; however, similar to how different service providers will be helped or harmed by reporting performance metrics, different users could be helped or harmed depending on their financial performance. It cannot be assumed that users are fully aware of the implications of making late payments or defaulting that go beyond just having their electricity turned off. Though users will likely have participated in informal banking and credit systems, those systems tend to be tightly intertwined with social networks; the notion of a credit history or credit score and how those numbers are used may be entirely new to users.

Perhaps even more important than to users, is the value that these data provide to service providers as those who currently control access to it, both through reselling the data and quantifying and reducing the risk of their financial portfolio:

> "What different PAYGO companies can do with the customer data they collect through OGS use and payments will be key for extracting long-term value from an otherwise investment-heavy model. Strong capabilities to mine this data will support both product and service sales while helping reduce portfolio risks.
>
> For instance, continuous repayment data allows companies to refine their credit assessment models and build a consumer profile, allowing for the segmentation of consumers by their credit risk. This, in turn, allows companies to offer upgrades and additional services to low-risk customers. As the database of customers grows and is further refined, it turns into an asset that could provide avenues for partnership with other product manufacturers and asset-financing companies, and could also potentially be sold to third-parties. . . due attention will need to be paid to customer privacy as well as transparency in this regard " [5, p. 193].

We can see from the above-mentioned statement that some of these benefits, namely, assessing risks, can be enhanced by sharing and expanding customer databases. However, revenues from partnerships with other manufacturers and the sale of data could be compromised if the data are more open. If the data are shared only without PII, the service provider could retain control of much of the upselling opportunity, yet service providers are likely to want to restrict access. We also emphasize again that though reducing financial risk is likely to lower costs and help the industry as a whole, some potential users will find themselves disqualified, perhaps unjustly, as the presence of human bias in machine learning and algorithm errors is becoming recognized as a widespread social issue [37].

The ability to use technical data combined with financial and/or demographic data to estimate willingness-to-pay and for service providers to charge exploitative pricing is discussed next. We note here that financial data alone are sufficient to do this to some degree, even when data are not shared beyond an individual service provider. As such, there is a risk of exploitative pricing to users even without movement to more open access, and oversight is likely required. In general, the service provider benefits from economic growth and increased demand in the communities they serve, so they have this incentive to keep prices low, but they can also be expected to maximize profit rather than community benefit or utilization of electricity. Conversely to the risk of exploitation, expanding access to transaction information to stakeholders besides just other service providers could increase the fairness of pricing if users and governments are aware of market averages and can use that information to negotiate or regulate pricing. Indeed, standardized reporting of tariff and payment information to governments would be fundamental in expanding regulation to microutilities and are widely used by regulated utilities, but greater transparency to other stakeholders can guard against corruption or misreported data. Metrics on tariffs, payment frequency, and default rates need only be shared aggregated by region or by service provider to realize these benefits; individual records are not necessary. It is important that these aggregate metrics include contextual and qualitative information about the type of system and its intended use case, i.e., the cost per kilowatt hour for a solar lantern may seem exorbitant, but the cost for the service of lighting can at the same time be affordable and competitive to alternatives [38]. In addition, access to financial information can lead to innovation in pricing models that can jointly benefit service providers and users. For example, PAYGO models have revolutionized the industry in response to high connection costs preventing users from accessing service [39]; access to financial records by service providers, researchers, and development institutions would likely lead to further innovation.

Service providers and their investors also stand to benefit from better knowledge of industry-wide financial data so that they can better assess their performance against industry averages and leaders. Although organizations, such as Lighting Global and IRENA, have published reports on costs in the sector, access to any of the underlying data is restricted and little aggregated data on costs for systems larger than home solar systems are available [5], [40]. These cost data and the ability to run customized queries are critical for electricity infrastructure planning, research, and forecasting prices in the industry. Service providers naturally face competitive risks in sharing this data, and the geographic segmentation of the industry makes it likely that any geographical information would enable identification of the company; however, there are many opportunities to make anonymized and aggregated financial data more widely available. Development institutions, governments, and researchers already expend resources to collect this information in a limited fashion and would benefit substantially from larger, standardized, data sets and reduced collection effort.

C. Demographic Data

Demographic data are about users and their communities and can include income, occupation, household information, address or location, age, cultural affiliation, and so on. These data are not specific to electricity systems but can be collected by service providers in order to estimate demand, assess financial risk and willingness-to-pay, and track the impact of their service. Similar to how electricity service can be the first banking and formal credit relationship of a user, census data are often lacking in energy-poor communities and surveys by the service provider can be valuable records that increase visibility into communities for health, education, sanitation, and so on [41]. If these data are to be used, guidelines and standardization for administering surveys, ensuring consistency, and maintaining provenance are necessary. Beyond this quality assurance issue, these data must be used carefully as their purpose is related to energy service and would likely obscure other factors that are important on the ground. Nevertheless, this increased visibility has the potential to pressure governments, provide useful information to development institutions, and validate and augment existing census data (see Table 3).

Demographic data are important for tracking the impact of electrification and thus guiding investment priorities by governments and development institutions. However, tracking this impact is difficult because of the complexity and interdependences of the many ingredients for economic development and changes within communities (see [42] for a review of socioeconomic dynamics and causality issues). As a result, purely quantitative methods do not consistently find evidence that electrification has substantive impacts on development indices such as income, literacy, and health [39], [43], [44], but case studies and qualitative analysis do consistently find benefits from electrification (see [38] and [45]). Disaggregating energy access and use data alongside demographic variables can help in modeling and better understanding these complexities, as well as measuring nuances in access [41], [42]. While these data should be reported and analyzed with evolving state-of-the-art methods, the limitations of quantitative analysis need to be considered and compared with deep qualitative knowledge when using them to guide action.

Finally, we reiterate and expand how demographic data can be used in applications such as estimating electricity demand and assessing risk. Demographic features are statistically associated with demand characteristics, risks, and so on, and then, these features can in turn be used to predict these characteristics for new users or to forecast behavior. Once the association is developed, the prediction can happen in reverse, and demographic characteristics can be estimated from technical or financial data, potentially compromising anonymity efforts. Demographic data are beneficial, but not critical, for predicting future behavior of a particular user (estimates can be made strictly from past technical or financial data but are improved by comparison to similar users), but they are critical for predicting the behavior of new users, which is an essential part of infrastructure planning and design. Inaccuracies in these predictions can lead to improperly sized electricity systems that initiate vicious cycles of poor performance and financial losses that negatively impact all stakeholders and add risk to the industry as a whole [46].

IV. PRIVACY STRATEGIES

Depending on the category of data, stakeholders will have different motivations for protecting data related to their risks outlined earlier. One reason to implement privacy

Table 3 Values and Risks From the Demographic Data

Stakeholder	Values	Risk
User	• Raise awareness about socioeconomic conditions • Demonstrate impacts of electrification	• Over-emphasis on measurable factors • Demonstrate lack of impacts of electrification • Exploitative pricing • Becoming a statistic
Micro-utility	• Demand estimates • Assess financial risks • Understand user willingness-to-pay • Sell user data	• Competition with other service providers • Losing revenue from selling user data
Macro-utility	• Assess customer financial risk • Understand customer willingness-to-pay • Planning models	• Competition with other service providers
Government	• Progress tracking • Augment census data • Constituent benefits	[Same as technical data]
Development Institution	[Same as technical data]	[Same as technical data]
Researcher	[Same as technical data]	[Same as technical data]

techniques would be to ensure that data about individuals are not revealed in accordance with privacy policies or another reason might be to limit dissemination of proprietary knowledge. In the former case, thoughtful filtering techniques prevent query results from revealing information about specific individuals. In the case of proprietary knowledge, even broad statistics, such as average cost of a solar-home-system, ought to be guarded, although this does not preclude selective collaborations or publication of industry-wide statistics. In many scenarios, both strategic data filtering and selective sharing can and should be implemented.

A. Strategic Data Filtering

The tension between openness and privacy is reflected in the spectrum of privacy techniques with the essential tradeoff of different methods being how much information can be learned about groups of interest versus how well privacy of individuals in the data set is protected. On one end of the spectrum is not sharing data at all, the only perfectly private strategy. The opposite is publicly disseminating raw data, which maximizes the potential reuse and secondary analysis as well as the potential to learn personal information about individuals in the data. Everything in between is an attempt at balancing privacy and information.

Deidentifying data typically refers to removing entire columns containing obvious PII such as name, age, social security number, contact information, and location data. Often, this is considered good enough, and yet, the differential privacy literature [47] includes an impossibility result, demonstrating that once a data set has been removed of columns to be sufficiently private, there would not be enough information left to be considered data, i.e., fully deidentified data are not useful. For example, age representation in a population and proximity to roads or the grid could be the invaluable context that would be removed as PII, while "type of roof" or electricity usage characteristics may be enough to identify an individual when combined with auxiliary information or other seemingly non-PII in the data set. Location data could be removed from a data set and still estimated from combining renewable energy production with public weather data. Alternatives to simple deidentification are different forms of aggregation and adding noise.

Differential privacy is a theoretical framework involving a measured addition of noise to query results to optimize accuracy under the restraint that the differentially private query results are skewed enough that they cannot be used to determine whether or not an individual exists in the data set; this is considered the gold standard definition of privacy protection. The amount of noise turns out to be very small when asking broad questions about a large data set, while a significant amount of noise is added to specific queries about a small population. Because differential privacy involves adding noise, it is not a good fit for every application. Population statistics are good enough for many scenarios and fundamentally not enough information for others. The latter category includes combining data with other dependent data sets and building models from data, including machine learning. These likely need to be run as workflows on the full data set.

Data aggregation techniques fall somewhere in between deidentification and differential privacy in terms of privacy protection and interesting information retained. Aggregation reduces the granularity of data by summarizing the data in statistics or replacing precise values with ranges. To what extent aggregation preserves enough information and protects privacy depends on the strategy.

B. Role-Based Access Control

The intention is to learn as much as possible about issues and populations while mitigating risks to stakeholders. When multiple organizations independently release anonymized data about overlapping populations, all of the data are vulnerable to composition attacks [48]. As such, it is best to limit new access to data to only what is necessary to accomplish reasonable data analysis goals. Since reasonable intentions depend on the role of the stakeholder and the application of the data analysis, we recommend fine-grained access control based on user roles and corresponding intentions and privileges, described originally and in more detail in an earlier paper [49]. Role-based access control can be used to implement the concept of sharing data on a need-to-know basis even when the need is to learn as much as possible or to know a precise measurement of the population in the data set.

Each collaborator can be assigned one or more user role based on intentions and minimum necessary access privileges, which may correspond to different data filtering strategies, to satisfy those intentions. Collaborators from the same organization can be assigned to a user group to avoid potential work-around schemes of combining individuals' limited access to learn more than intended. For example, the concept of a privacy budget [47] can be shared by group members. The mechanism for measuring aggregation would compare a request with information that has already been granted to the individual as well as the group. Access based on the sensitivity of the data, i.e., potential risk to individuals, should be implemented in addition to permissions. For example, the Open Science Platform vision includes six privacy levels based on risk and associated security mechanism such as user authentication, password and two-factor authentication, and data use agreements [50]. DataTags is a model for automating the determination of these risks and policy suggestions that comply with legal and technical standards even in the absence of security and privacy expertise [51].

C. Using a System to Run Analysis on Behalf of Collaborators

Interfaces to data sets that help project owners navigate privacy risks while allowing just enough access to potential collaborators to validate or expand on original research would be a major improvement over personally selective sharing of static views of the data. A system can accept queries and analysis code, access the original database on behalf of the collaborator, and calculate how to return output based on aspects of what the query is requesting, who the collaborator is along with his or her relationship to the data set as determined by his or her user role, a log of previous requests, and metadata about uniqueness of certain values and sensitivity of certain features in the data. FLEX is a system that accepts an SQL query as an input, analyzes the query against precomputed database metrics, and processes the original query results to return differentially private results [52]. Ideally, access to data via a system, such as FLEX, would be set up for certain user roles, while other data privacy techniques might be more appropriate for other roles.

Comprehensive data management platforms have also been proposed [49], [50] for shared access to data, tools, and workflows in secure ways. These systems offer provenance that enables reproducibility—a key benefit of using systems that necessarily keep track of users, requests, and ideally versioning of data and processing code too. Privacy-preserving data analysis components of such systems can be validated for a variety of circumstances and become familiar to institutional review boards and policy writers, encouraging higher standards and more consistency. Section I describing the proposed Mezuri platform [49] includes an overview of common data management systems in several fields, noting that often these systems are specialized for the field, not the type of data. A distinguishing feature is whether or not the system includes built-in data processing capacity or simply a repository to deposit and download data in an organized although harder to track way. The systems with support for end-to-end workflows are those enabling more robust approaches to privacy protection as described next.

We hope that realizing the value of sharing data and workflows described in this paper is not dependent on such systems becoming popular and instead might be accomplished incrementally as key features are implemented into workflows. That said, until comprehensive and flexible data management platforms are easily adoptable by technical and nontechnical practitioners, much of the work of setting up information infrastructure is redundant across similar projects.

An aspect of shared data management infrastructure worth reflecting on in any conversation of stakeholders is to what extent it is centralized, the definition and implications of which depend on context [53]. Certain features of database systems make it easier to thoughtfully open up access to data and workflows and generate a record of who is involved in which ways. Centralized systems also preserve an existing aspect of data management in which control remains with original data owners. For example, it would be more feasible for participants in a study or customers of a service to be given access to data collected about themselves and permission to compare it with data set population statistics. However, this is optional and does not automatically empower other stakeholders to influence decision-making processes. Decentralized systems are not inherently more accessible by a greater number of or more diverse stakeholders. However, once control of a decentralized system is divided among disparate parties, it would be hard or impossible to reverse this division of power, unlike in a centralized system where access privileges can simply be revoked.

A security feature of current interest associated with certain peer-to-peer decentralized systems is immutability as a result of built-in consensus mechanisms

Table 4 Aggregated Responses From the Strathmore University Energy Access Workshop

Stakeholder	Potential Positive Impacts	Potential Negative Impacts
User	• Education/impact • Habit changes • More rapid deployment of technologies • Improvement in the quality of services • Energy efficiency	• Risk to privacy of consumption patterns or personal identifiable information • Inaccurate representation of challenges / becoming a statistic
Micro-utility	• Better understanding of: current and future customers, competitors, market insights, and technical performance • Better customer support • Ability to lower costs • Larger datasets for machine learning and data science	• Privacy competition with other energy sources • The spread of sensitive info like tariffs and revenues • Market leaders lose a competitive advantage
Macro-utility	• Data analysis and feedback from many organizations (academia, independents) • Better response rate, increased number of findings and identification of issues • Better understanding of customers • Better performance of system	• Increased competition with smaller generating sites • Sensitivity on the use of public funds (i.e. cost of data) • Contradictions could be revealed
Government	• Better reporting capabilities • Increase ability of citizens to understand and engage with their electricity access • Offers capacity development opportunities for data science students	• More responsibility for regulation • Encroaching consumer rights • Concerns with quality of data
Development Institution	• Better analytics + impact reporting + performance • Visibility into impact evaluation • Able to select projects with better rates of success • Evidence-based fund-raising with donors	• Bad performance will be easy to see
Researcher	• Efficient way to collect tons of on-the-ground data • Facilitates publication of papers • Education + learning with impact • Assist with detailed, optimized mini-grid site selection for most impact, community-tailored design • Researchers can: add value to raw data, make findings more accessible to wider range of users, make better predictions with bigger data-sets, better understand the current and future status of the energy field	• Security and intellectual property issues in regards to data access, exchange, ownership and sharing • The risk of inaccurate or non-rigorously collected data (sub-academic methodology) • A lack of understanding of the context of the data
Other		• People dont know what to use the data for

(e.g., blockchains [54], [55]). For the data we describe, privacy would need to be protected by either encrypting the data or limiting participation to known, trusted parties. Decentralized consensus mechanisms can also be expensive and energy intensive [56], which begs the questions who can afford to participate in validation, what would their incentives be to do so, and whether immutability of the data is worth the cost of this approach to it. If participation in data sharing is limited to trusted parties, there would be a low risk of tampering. If a distributed repository of encrypted data is publicly accessible, control over who gains access is determined by who holds the decryption keys, likely the original data owners.

V. CONCLUSION

A. Opportunities and Tensions

Key opportunities from expanding access to information are efficiency in the planning, regulation, design, operation, and use of electricity systems, facilitation of innovation, privacy protection by virtue of aggregation of large data sets, transparency and visibility, and thoughtful balancing of risk and capacity through a shared data management platform. Efficiency in the use of electricity systems requires better access to technical data (including education) by end users of electricity, while efficiency in design and operation is enabled by technical data sharing among service providers and efficiency in planning and regulation is enabled by technical data sharing among service providers, governments, and development institutions. Some efficiency opportunities can be gained by having data shared privately between individual service providers and users, but the value in forecasting, planning, and risk assessment requires linkages among technical, financial, and demographic data. These linkages can compromise privacy, which can expose users to legal, political, and financial risk. Broader access by all stakeholders, especially including researchers and service providers, to all categories of information would facilitate

Table 5 Data Collection, Sharing, and Privacy Policies of Noteworthy Stakeholders

Organization Name	Policy online? Location?	Link	Refers to sharing policy?	Refers to collection practices?	Refers to privacy protection?	Notes
IRENA	yes, under "Data Methodology"	https://irena.masdar.ac.ae/gallery/#gallery, http://dashboard.irena.org/download/Methodology.pdf	yes	yes	"Do not disaggregate"	
The World Bank	yes, under "Legal, terms of use"	http://www.worldbank.org/en/about/legal/terms-of-use-for-datasets, https://openknowledge.worldbank.org/terms-of-use	yes	yes	In separate policy: https://policies.worldbank.org/sites/ppf3/PPFDocuments/Forms/DispPage.aspx?docid=e569ea42-d6ca-4004-ab36-6f48457471a7	
GOGLA	yes	n/a	no	yes		
Lighting Global	no	n/a	no	yes	no	
EIA	yes	https://www.eia.gov/about/copyrights_reuse.php	yes	yes		Refers to Freedom of Information Act (FOA)
SEIA	no					
USAID	yes	https://www.usaid.gov/sites/default/files/documents/1868/USAID_OpenGovPlan2016.pdf	yes	yes	yes	
NIH	yes	https://grants.nih.gov/grants/policy/data_sharing/	yes	yes	yes	
DFID	was withdrawn	https://www.gov.uk/government/publications/information-strategy [Withdrawn]	no	no	no	Refers to International Aid Transparency Initiative (IATI)

innovation in technology, business models, and development practices. Furthermore, there are potential benefits from increasing the energy literacy of society at large and providing capacity development opportunities to data science students and emerging technology sectors that should be in the interest of government and development institutions. Although expanding data access can threaten privacy, having more users and service providers participating in a shared platform has anonymity benefits through numbers; for example, as more users with similar demographic information in a region have their information recorded, it becomes more difficult to identify them from technical, financial, or demographic data. Adoption of a shared set of practices that are approved by governments, and even a common data management platform using such practices, is a way for the current curators of data (mainly service providers) to absolve themselves of some responsibility. A common platform can also allow capacity pooling in data analysis and security or leveraging the analytical capacity of researchers, development institutions, and other consultants. Furthermore, standard practices or a common platform can facilitate better transparency, which is viewed as a necessity by microutilities in order to realize values and avoid future consequences:

> "PAYGO companies need to get ahead of the data and privacy issue by either individually, or preferably as an industry, signing up to transparency and outlining guidelines for the handling of customer data. The World Bank's focus group study with retailers and distributor networks showed universal concern over private consumer data being shared externally by the providers, while recognizing the potential for benefits to consumers from prudent and confidential uses of their data" [5, p. 11].

Though there are promising opportunities, several tensions need to be recognized and addressed in order to minimize the resistance to changing practices and to evaluate disparate impacts among stakeholders. In general, there will be winners and losers within stakeholder groups from reducing information asymmetries. These are likely to be service providers who are disparately affected by transparency in their product performance and pricing, users who are granted or denied access to financial services, and governments and development institutions who will each face in some cases pressure to act or validation from increased transparency. Between stakeholders, there are cases where values and risks are in direct opposition, such as the potential for exploitative pricing or the sale of data to benefit service providers and harm users. Expanding access to all categories of data is likely to erode barriers to entry in providing service, which could benefit users and align with development institutions and governments who are not also service providers but lessens the competitive advantage held by current service providers, an advantage that in some cases is key to their viability. In evaluating these tensions, it will be necessary to carefully consider the current power dynamics between stakeholders and their complexities, how data will be used under these dynamics, and how data will alter them. Data can be used not only to increase transparency and expand access but also to reinforce the power dynamics underlying development between rich and poor countries or between users and service providers.

We also emphasize the importance of the implied responsibility from expanding access to data. This responsibility can present itself as a liability, where those controlling data access are likely to be assumed responsible by the public and/or legally responsible for protecting data; consider, for example, the high-profile case of Facebook selling user data to political strategy firms [57]. Open data policies often use language deflecting responsibility for privacy risks to individual projects and cannot be relied upon for legal protection.[11] In this sense, more open access under standardized practices could release individual stakeholders from their responsibility as curators, but it can also introduce additional responsibility for stakeholders who are gaining access to act accordingly, sometimes explicitly through data access agreements. For example, governments may face additional pressure to enforce regulation,[12] service providers to improve their quality of service or users to change their behavior. While these changes could align with stakeholders' missions and be mutually beneficial, the right incentives to realize them are not necessarily in place, and status quo bias could further impede change.

B. Recommendations

Codesign of data sharing practices and platforms among all stakeholders is critical to expanding access, realizing benefits, and mitigating risks. Users must be represented through focus groups and workshops and advocated for by representatives who have limited conflicting interests with stakeholder groups, with particular attention to data accessibility and privacy concerns. Education in interpreting technical data and understanding the implications of financial data collection is necessary, and development institutions and governments should provide resources and oversight to assist service providers in this effort. Service providers need to take a leading role in recommending data sharing practices to mitigate their competitive risks and protect value streams. Governments, development institutions, and researchers need to provide guidance on practices for collecting demographic data that can be useful for broader development goals and visibility. Again, codesign is critical; specifically, governments and communities with ongoing relationships to development programs need to be included in designing indicators and deciding what kinds of data are relevant.[13] These discussions need to be held in the context of a changing regulatory and macroutility business model landscape, such that data sharing practices enable forward-looking goals and are not restricted by the needs of the current paradigm.

We find that few benefits are likely to be realized from expanding access to detailed payment information; aggregate metrics are sufficient to inform users and governments, and there is a substantial risk to users and lost value to service providers from exposing these data, even anonymously. As such, we do not recommend expanding access to these data, though we do recommend guidelines and legal standards for the appropriate handling of payment data by service providers. Furthermore, legal standards for protection against exploitative pricing need to be established as service providers continue to collect these data. These regulations need to allow for business model innovation and the diverse contexts of electrification projects; price ceilings on the cost per unit energy are commonly used but overly limiting restrictions, while special attention should be paid to variable electricity rates and predatory lending.

We recommend increasing access to anonymized and standardized technical and demographic data through a common platform. Not only will this facilitate efficiency gains and technological innovation in the planning and operation of power systems, and better tracking of electrification impacts and energy use, but it will allow state-of-the-art privacy-preserving techniques such as differential privacy to be employed to prevent reidentification of sensitive data and will reduce the responsibility of curation that is currently borne by individual service providers. Development institutions are in a position to provide resources to support the codesign and implementation of such a platform. We also recommend that service providers grant access to aggregated data on the cost of providing service to improve infrastructure planning efforts and set industry benchmarks.

Finally, we recommend that caution be taken by all stakeholders, especially development institutions and governments, in taking action based on data and statistical knowledge. Overemphasis of measurable impacts oversimplifies complexity, neglects nuanced knowledge, and has the potential to exacerbate inequalities and asymmetric power relationships. Qualitative and deep local knowledge are indispensable in meeting global development challenges in electricity access and beyond. ■

APPENDIX A
ENERGY ACCESS WORKSHOP RESULTS

In July of 2018, Strathmore University's Energy Research Centre (SERC) in partnership with the Energy and Resources Group (ERG) at the University of California at

[11]For example, USAID's open data policy—Automated Directives System 579 section 579.3.2.3: *Redacting Data and Exceptions to the Open Data Mandate*, and the *Final NIH statement on sharing research data*. Interestingly, the *DFID Digital Strategy 2018 to 2020: doing development in a digital world*, simply references *The Principles for Digital Development* which briefly names data security as an issue without going into detail or context.

[12]This pressure could come from the public, political groups, or also from rich countries to poorer.

[13]For a detailed discussion of the history and processes around development indicators, see [16].

Berkeley hosted a one-day workshop on Decentralized Energy Solutions for East Africa and the Role of Research. The approximately 60 attendees were invited to voluntarily and anonymously complete and return a blank table in the model of Tables 1–3. The handout was introduced to the attendees with the goal of understanding the perspectives in the room in answering the question, "What are the possible impacts to each stakeholder group from increased data sharing in the context of energy access work?" A synthesis of the ten of the returned forms where respondents explicitly granted consent for distribution is shown in Table 4. Respondents were not asked to differentiate between technical, financial, or demographic data.

APPENDIX B
DATA COLLECTION, SHARING, AND PRIVACY POLICIES

A summary of institutional and industry group data policies that are published online is given in Table 5.

Acknowledgments

The authors would like to thank the Strathmore University's Energy Research Centre and Dr. Rebekah Shirley of Power for All for their work in bringing together stakeholders at July, 2018 Decentralized Energy Solutions Workshop, and the stakeholders who generously shared their perspectives with them.

REFERENCES

[1] J. C. Aker and I. M. Mbiti, "Mobile phones and economic development in Africa," *J. Econ. Perspect.*, vol. 24, no. 3, pp. 207–232, 2010. [Online]. Available: https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.24.3.207

[2] D. Chaplin *et al.*, "Grid electricity expansion in tanzania by MCC: Findings from a rigorous impact evaluation," Mathematica Policy Res., Washington, DC, USA, Tech. Rep., 2017. [Online]. Available: https://www.edi-global.com/wp-content/uploads/2017/03/TZ-Electricity-Impact-Report.pdf

[3] J. Burrell, "Evaluating shared access: Social equality and the circulation of mobile phones in rural Uganda," *J. Comput.-Mediated Commun.*, vol. 15, no. 2, pp. 230–250, Jan. 2010. [Online]. Available: https://academic.oup.com/jcmc/article/15/2/230-250/4064811

[4] P. Yadav, A. P. Heynen, and D. Palit, "Pay-as-you-go financing: A model for viable and widespread deployment of solar home systems in rural India," *Energy Sustain. Develop.*, vol. 48, pp. 139–153, Feb. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0973082618309839#bb0015

[5] International Finance Corporation. (2018). *Off-Grid Solar Market Trends Report 2018*. [Online]. Available: https://www.lightingafrica.org/wp-content/uploads/2018/02/2018_Off_Grid_Solar_Market_Trends_Report_Full.pdf

[6] B. Sergi, M. Babcock, N. J. Williams, J. Thornburg, A. Loew, and R. E. Ciez, "Institutional influence on power sector investments: A case study of on- and off-grid energy in Kenya and Tanzania," *Energy Res. Social Sci.*, vol. 41, pp. 59–70, Jul. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214629618303566

[7] H. Louie, E. O'Grady, V. Van Acker, S. Szablya, N. P. Kumar, and R. Podmore, "Rural off-grid electricity service in sub-saharan Africa [technology leaders]," *IEEE Electrific. Mag.*, vol. 3, no. 1, pp. 7–15, Mar. 2015.

[8] J. Winiecki. (May 2015). *Four Ways Energy Access Can Propel Financial Inclusion*. [Online]. Available: http://www.cgap.org/blog/four-ways-energy-access-can-propel-financial-inclusion

[9] GSMA. (2017). *Mobile for Development Utilities*. [Online]. Available: https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2017/01/Lessons-from-the-use-of-mobile-in-utility-pay-as-you-go-models.pdf

[10] V. Mutyaba. (Apr. 2018). *Mobile Money Transforming Electricity Payment in Uganda*. [Online]. Available: https://www.esi-africa.com/mobile-money-transforming-electricity-payment/

[11] K. Nagasawa, C. R. Upshaw, J. D. Rhodes, C. L. Holcomb, D. A. Walling, and M. E. Webber, "Data management for a large-scale smart grid demonstration project in Austin, Texas," in *Proc. ASME 6th Int. Conf. Energy Sustainability*, Jul. 2012, pp. 1027–1031. [Online]. Available: http://proceedings.asmedigitalcollection.asme.org/proceeding.aspx?doi=10.1115/ES2012-91198

[12] M. E. H. Dyson, S. D. Borgeson, M. D. Tabone, and D. S. Callaway, "Using smart meter data to estimate demand response potential, with application to solar energy integration," *Energy Policy*, vol. 73, pp. 607–619, Oct. 2014. [Online]. Available: https://www-sciencedirect-com.libproxy.berkeley.edu/science/article/pii/S0301421514003681#fn1

[13] A. Albert and R. Rajagopal, "Smart meter driven segmentation: What your consumption says about you," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4019–4030, Nov. 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6545387/

[14] S. Houde, A. Todd, A. Sudarshan, J. A. Flora, and K. C. Armel, "Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence," *Energy J.*, 2013.

[15] H. Allcott and T. Rogers, "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *Amer. Econ. Rev.*, vol. 104, no. 10, pp. 3003–3037, Oct. 2014. [Online]. Available: http://pubs.aeaweb.org/doi/10.1257/aer.104.10.3003

[16] S. E. Merry, *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago, IL, USA: Univ. Chicago Press, 2016.

[17] M. Chan, "From new estimates to better data," *Lancet*, vol. 380, no. 9859, p. 2054, Dec. 2012. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/23245596

[18] J. C. Ribot and N. L. Peluso, "A theory of access," *Rural Sociol.*, vol. 68, no. 2, pp. 153–181, Jun. 2003. [Online]. Available: http://doi.wiley.com/10.1111/j.1549-0831.2003.tb00133.x

[19] N. Klugman, J. Rosa, P. Pannuto, M. Podolsky, W. Huang, and P. Dutta, "Grid watch: Mapping blackouts with smart phones," in *Proc. 15th Workshop Mobile Comput. Syst. Appl.*, 2014, Art. no. 1.

[20] (2017). *Enterprise Surveys*. [Online]. Available: http://www.enterprisesurveys.org/data/exploretopics/infrastructure#sub-saharan-africa

[21] T. Monroe, S. Debere, K. M. Gaba, D. Newhouse, and T. Killic, "Putting big data innovation into action for development," in *Earth Observation Open Science and Innovation*, P.-P. Mathieu and C. Aubrecht, Eds. Cham, Switzerland: Springer, 2018, pp. 285–291. [Online]. Available: http://link.springer.com/10.1007/978-3-319-65633-5_15

[22] P. Wouters and W. Haak, "Open data: The researcher perspective," Tech. Rep., 2017.

[23] J. Rosa, P. A. Madduri, and D. Soto, "Efficient microgrid management system for electricity distribution in emerging regions," in *Proc. IEEE Global Humanitarian Technol. Conf.*, Oct. 2012, pp. 23–26. [Online]. Available: http://ieeexplore.ieee.org/document/6395730/

[24] D. Ponce de Leon Barido, S. Suffian, D. M. Kammen, and D. Callaway, "Opportunities for behavioral energy efficiency and flexible demand in data-limited low-carbon resource constrained environments," *Appl. Energy*, vol. 228, pp. 512–523, Oct. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261918309863

[25] C. Blodgett, P. Dauenhauer, H. Louie, and L. Kickham, "Accuracy of energy-use surveys in predicting rural mini-grid user consumption," *Energy Sustain. Develop.*, vol. 41, pp. 88–105, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0973082617304350

[26] D. H. Fabini, D. Ponce de León Baridó, A. Omu, and J. Taneja, "Mapping induced residential demand for electricity in Kenya," in *Proc. 5th ACM Symp. Comput. Develop.*, vol. 5, New York, NY, USA, 2014, pp. 43–52. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2674377.2674390

[27] A. Miketa and N. Saadi, "Africa power sector: Planning and prospects for renewable energy," Int. Renew. Energy Agency, Tech. Rep., 2015. [Online]. Available: http://www.irena.org/DocumentDownloads/Publications/IRENA_Africa_Power_Sector_synthesis_2015.pdf

[28] F. Kemausuor, E. Adkins, I. Adu-Poku, A. Brew-Hammond, and V. Modi, "Electrification planning using network planner tool: The case of Ghana," *Energy Sustain. Develop.*, vol. 19, no. 1, pp. 92–101, Apr. 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S097308261300121X

[29] U. Deichmann, C. Meisner, S. Murray, and D. Wheeler, "The economics of renewable energy expansion in rural Sub-Saharan Africa," *Energy Policy*, vol. 39, no. 1, pp. 215–227, 2011. [Online]. Available: http://ac.els-cdn.com/S0301421510007202/1-s2.0-S0301421510007202-main.pdf?_tid=48f50c60-5c3d-11e7-95e0-00000aab0f6b&acdnat=1498680574_392c16ce7145d43a8782453e365fa191

[30] S. Szabó, K. Bódis, T. Huld, and M. Moner-Girona, "Sustainable energy planning: Leapfrogging the energy poverty gap in Africa," *Renew. Sustain. Energy Rev.*, vol. 28, pp. 500–509, Dec. 2013.

[31] K. B. Debnath and M. Mourshed, "Challenges and gaps for energy planning models in the developing-world context," *Nature Energy*, vol. 3, no. 3, pp. 172–184, Mar. 2018. [Online]. Available: http://www.nature.com/articles/s41560-018-0095-2

[32] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive Load Monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.

[33] U. Greveler, B. Justus, and D. Loehr, "Multimedia content identification through smart meter power usage profiles," *Comput., Privacy Data Protection*, vol. 1, no. 10, 2012.

[34] J. Taneja. (2017). *Measuring Electricity Reliability in Kenya*. [Online]. Available: http://blogs.umass.edu/jtaneja/files/2017/05/outages.pdf

[35] K. Lee *et al.*, "Electrification for 'Under Grid' households in Rural Kenya," *Develop. Eng.*, vol. 1, pp. 26–35, Jun. 2016. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S235272851530035X

[36] Y. Anokwa *et al.*, "Stories from the field: Reflections on HCI4D Experiences," *Inf. Technol. Int. Develop.*, vol. 5, no. 4, pp. 101–116, 2009. [Online]. Available: http://www.itidjournal.org/index.php/itid/article/view/427/195

[37] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, 2016.

[38] P. Alstone, D. Gershenson, and D. M. Kammen, "Decentralized energy systems for clean electricity access," *Nature Climate Change*, vol. 5, no. 4, pp. 305–314, 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/nclimate2512

[39] K. Lee, E. Miguel, and C. Wolfram, "Appliance ownership and aspirations among electric grid and home solar households in rural Kenya," *Amer. Econ. Rev.*, vol. 106, no. 5, pp. 89–94, May 2016. [Online]. Available: http://pubs.aeaweb.org/doi/10.1257/aer.p20161097

[40] International Renewable Energy Agency. (2016). *Solar PV in Africa: Costs and Markets*. [Online]. Available: https://www.irena.org/DocumentDownloads/Publications/IRENA_Solar_PV_Costs_Africa_2016.pdf

[41] V. C. Broto *et al.*, "A research agenda for a people-centred approach to energy access in the urbanizing global south," *Nature Energy*, vol. 2, pp. 776–779, Oct. 2017.

[42] F. Riva, H. Ahlborg, E. Hartvigsson, S. Pachauri, and E. Colombo, "Electricity access and rural development: Review of complex socio-economic dynamics and causal diagrams for more appropriate energy modelling," *Energy Sustain. Develop.*, vol. 43, pp. 203–223, Apr. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S097308261731390X

[43] K. Lee, E. Miguel, and C. Wolfram. (2018). *Experimental Evidence on the Economics of Rural Electrification*. [Online]. Available: http://www.catherine-wolfram.com/uploads/8/2/2/7/82274768/repp-jpe_2018-01-31-final.pdf

[44] F. Burlig and L. Preonas. (2016). *Out of the Darkness and Into the Light? Development Effects of Rural Electrification*. [Online]. Available: https://ei.haas.berkeley.edu/research/papers/WP268.pdf

[45] C. Kirubi, A. Jacobson, D. M. Kammen, and A. Mills, "Community-based electric micro-grids can contribute to rural development: Evidence from Kenya," *World Develop.*, vol. 37, no. 7, pp. 1208–1221, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0305750X08003288#!

[46] D. Schnitzer, D. S. Lounsbury, J. P. Carvallo, R. Deshmukh, J. Apt, and D. M. Kammen. (2014). Microgrids for rural electrification: A critical review of best practices based on seven case studies. United Nations Foundation. [Online]. Available: https://rael.berkeley.edu/wp-content/uploads/2015/04/MicrogridsReportEDS.pdf

[47] C. Dwork, Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.*, 2008, pp. 1–19.

[48] S. R. Ganta, S. P. Kasiviswanathan, and A. Smith, "Composition attacks and auxiliary information in data privacy," in *Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2008, pp. 265–273.

[49] A. Kipf *et al.*, "A proposed integrated data collection, analysis and sharing platform for impact evaluation," *Develop. Eng.*, vol. 1, pp. 36–44, Jun. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352728515300014

[50] L. Sweeney and M. Crosas, "An open science platform for the next generation of data," *Electron. J.*, Apr. 2013. [Online]. Available: http://www.ssrn.com/abstract=2302084

[51] M. Bar-Sinai, L. Sweeney, and M. Crosas, "Datatags, data handling policy spaces and the tags language," in *Proc. IEEE Symp. Secur. Privacy Workshops (SPW)*, May 2016, pp. 1–8.

[52] N. Johnson, J. P. Near, and D. Song, "Towards practical differential privacy for SQL queries," *Proc. VLDB Endowment*, vol. 11, no. 5, pp. 526–539, 2018. [Online]. Available: https://dl.acm.org/citation.cfm?id=3177733

[53] L. Winner, *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago, IL, USA: Univ. Chicago Press, 1986.

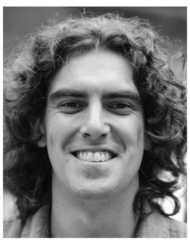
[54] S. Nakamoto. *Bitcoin: A Peer-to-Peer Electronic Cash System*. [Online]. Available: www.bitcoin.org

[55] G. Zyskind, O. Nathan, and A. S. Pentland, "Decentralizing privacy: Using blockchain to protect personal data," in *Proc. IEEE Secur. Privacy Workshops*, May 2015, pp. 180–184. [Online]. Available: https://ieeexplore.ieee.org/document/7163223/

[56] D. Malone and K. J. O'Dwyer, "Bitcoin mining and its energy footprint," in *Proc. 25th IET Irish Signals Syst. Conf. China-Ireland Int. Conf. Inf. Communities Technol. (ISSC/CIICT)*. Limerick, Ireland: IET, Jun. 2013, pp. 280–285. [Online]. Available: https://digital-library.theiet.org/content/conferences/10.1049/cp.2014.0699

[57] M. Rosenberg, N. Confessore, and C. Cadwalladr. (Mar. 2018). *How Trump Consultants Exploited the Facebook Data of Millions*. [Online]. Available: https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

ABOUT THE AUTHORS



Jonathan T. Lee (Graduate Student Member, IEEE) received the B.S. degree in electrical engineering and computer science, the B.S. degree in environmental economics and policy, and the M.S. degree in energy and resources from the University of California at Berkeley, Berkeley, CA, USA, in 2013, 2013, and 2018, respectively, where he is currently working toward the Ph.D. degree.

His current research interests include design and control of decentralized solar photovoltaics and battery systems, and expanding access to electricity to communities in poverty.

Isa L. Ferrall (Graduate Student Member, IEEE) received the B.S.E. degree *(magna cum laude)* in mechanical engineering and the Certificate in energy and environment from Duke University, Durham, NC, USA, in 2015, and the M.S. degree in energy and resources from the University of California at Berkeley, Berkeley, CA, USA, in 2018, where she is currently working toward the Ph.D. degree in energy and resources.

Her research uses data-driven approaches to investigate the impact of decentralized solar on rural electrification and energy transitions in East Africa.

Jordan Freitas (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from Creighton University, Omaha, NE, USA, in 2013. She is currently working toward the Ph.D. degree in computer science at the University of California at Berkeley, Berkeley, CA, USA.

She is currently an Adjunct Faculty Member with the Masters of Science in Health Informatics Program, University of San Francisco, San Francisco, CA, USA. Her current research interests include collaborative data management tools and workflows as a means of more accurate exploration of social issues, particularly interested in applications related to sustainable energy and global health.

Daniel M. Kammen received the B.A. degree in physics from Cornell University, Ithaca, NY, USA, and the M.A. and Ph.D. degrees from Harvard University, Cambridge, MA, USA, in 1986 and 1988, respectively.

He was the Director of the Transportation Sustainability Research Center, University of California at Berkeley, Berkeley, CA, USA,3 from 2007 to 2015. He held postdoctoral positions at the California Institute of Technology, Pasadena, CA, USA, and Harvard University. He was an Assistant Professor and the Chair of the Science, Technology and Environmental Policy Program, Woodrow

Wilson School, Princeton University, Princeton, NJ, USA. He is currently a Professor with the University of California at Berkeley, where he is also the Chair of the Energy and Resources Group and directs the Center for Environmental Policy and the Department of Nuclear Engineering, Goldman School of Public Policy. He is also the Founding Director of the Renewable and Appropriate Energy Laboratory (RAEL), University of California at Berkeley.

Dr. Kammen was appointed by Secretary of State Hillary Clinton in 2010 as the first Energy Fellow of the Environment and Climate Partnership for the Americas (ECPA) initiative. He began his service as Science Envoy for U.S. Secretary of State John Kerry in 2016, but resigned over President Trump's policies in 2017. He has served the State of California and the U.S. Federal Government in expert and advisory capacities, including at the U.S. Environmental Protection Agency, the U.S. Department of Energy, the Agency for International Development, and the Office of Science and Technology Policy. He has served as a contributing or coordinating lead author on various reports of the Intergovernmental Panel on Climate Change since 1999. The IPCC shared the 2007 Nobel Peace Prize.

Eric Brewer (Member, IEEE) is currently a Professor Emeritus of computer science with the University of California at Berkeley, Berkeley, CA, USA, and the Vice President of Infrastructure at Google, Mountain View, CA, USA. He has pioneered the use of clusters of commodity servers for Internet services based on his research at Berkeley. His CAP Theorem covers basic tradeoffs required in the design of distributed systems and followed from his work on a wide variety of systems, from live services, to caching and distribution services, to sensor networks.

Dr. Brewer is a member of the National Academy of Engineering and the AAAS. He was a recipient of the ACM Computing Prize.

Duncan S. Callaway (Member, IEEE) received the B.S. degree in mechanical engineering from the University of Rochester, Rochester, NY, USA, in 1995, and the Ph.D. degree in theoretical and applied mechanics from Cornell University, Ithaca, NY, USA.

He is currently an Associate Professor of energy and resources with the University of California at Berkeley, Berkeley, CA, USA. His current research interests include control strategies for demand response, electric vehicles and energy storage, distribution network management, and pathways to electrify and decarbonize the world's low-income regions.