

RESEARCH ARTICLE

Data Platform Guidelines and Prototype for Microgrids and Energy Access: Matching Demand Profiles and Socio-Economic Data to Foster Project Development

DAVIDE FIORITI¹, (Member, IEEE), NICOLÓ STEVANATO², PIETRO DUCANGE³,
FRANCESCO MARCELLONI³, EMANUELA COLOMBO²,
AND DAVIDE POLI¹, (Member, IEEE)

¹DESTEC, University of Pisa, 56122 Pisa, Italy

²Department of Energy, Politecnico di Milano, 20156 Milan, Italy

³Department of Information Engineering, University of Pisa, 56126 Pisa, Italy

Corresponding author: Davide Fioriti (davide.fioriti@unipi.it)

This work was supported in part by the European Union's Horizon 2020 Research and Innovation Program through the SETADISMA project that is part of the LEAP-RE ("Long-Term Joint European Union - African Union Research and Innovation Partnership on Renewable Energy") programme under Grant 963530; in part by the PNRR-M4C2-Investimento 1.3, Partenariato Esteso PE00000013-FAIR—Future Artificial Intelligence Research-Spoke 1 Human-Centered AI; in part by the Italian Ministry of University and Research (MUR) in the Framework of the FoReLab and CrossLab projects (Departments of Excellence); and in part by the University of Pisa through the CRUI-CARE Agreement.

ABSTRACT Energy access is a key need for socio-economic growth. Proven to be a key enabler of development and progress, access to electricity has been prioritized by governments using grid extension actions and off-grid solutions, namely microgrids and home systems technologies, fed by renewable sources. However, achieving universal access to energy remains a huge challenge given the lack of resources and the large population currently unserved. The lack of adequate socio-economic data at granular scale and of a good understanding of demand uptake led by economic growth is a barrier for efficient energy planning. Access to conjoint demand and socio-economic data at local level is crucial, yet hard to obtain: often such data are unavailable or very difficult to collect, and current data platforms often lack the ability to conjointly store variegated socio-economic and time series data. For these reasons, in this paper, we present a comprehensive methodology that, based on an extensive literature review, draws guidelines for developing data-sharing platforms in energy access, develops a proposed architecture to support the data collection of conjoint socio-economic and time-series data, and proposes a prototype of the final application. The methodology leverages on a novel extensive literature review to identify the major determinants of demand uptake and the corresponding consuming entities: villages, households, and appliances. The proposed architecture is able to capture numeric, categorical, and time series information for all consuming entities, based on state-of-the-art NoSQL databases. Finally, a prototype implementation with a web-based interface developed with Angular and Spring is proposed and discussed.

INDEX TERMS Access to electricity, load estimation, open data, NoSQL database, software platform, energy access.

I. INTRODUCTION

A. MOTIVATION

Ending poverty and hunger, providing access to clean water, and ensuring universal access to affordable, reliable, and

The associate editor coordinating the review of this manuscript and approving it for publication was Daniela Cristina Momete¹.

sustainable energy worldwide are of crucial importance, as stressed in the Agenda2030 in 2015 by the United Nations [1]. Energy is a major enabler for growth [2] as it activates the use of modern devices and supports Productive Uses of Energy, among which commercial and industrial, that generate income. However, the nexus among socio-economic growth, energy, local context and appropriate

policy-education measures is still a complex issue which calls for further investigation [3], [4], as the energy-development nexus is also significantly conditioned by several socio-economic, infrastructural and cultural factors and pre-conditions [5], [6], which are often hard to monitor and quantify.

Data for energy applications and, in particular, electricity are of utmost importance in the appropriate planning of energy access plans [7]. Recognizing this challenge, many institutions have proposed data platforms, such as the Energy Access Explored by the World Energy Institute [8] and the Global Electrification Platform by ESMAP [9]. The databases of these platforms, however, rarely contain granular data at household level including flexibility in the quantity and type of data fields, granular appliance information, and/or time series of energy consumption, that are highly needed for high-resolution energy estimates. While demand for data and willingness to release open-data are raising, we believe that it is timely to investigate the development of guidelines and architectures for open data platforms for energy access applications.

B. LITERATURE ANALYSIS

Electricity access is of utmost priority, as it enables supplying most modern devices that support growth and reduces reliance on energy sources like biomass or kerosene, which are often used inefficiently. However, in 2022 still about 800 million people lack access, most of them living in Sub-Saharan Africa, often in fragile and conflict-affected situations and/or rural areas [7], where the Covid-19 pandemic has even further reduced the affordability of such services [10]. However, when considering the expected population and economic growth, the International Energy Agency (IEA) forecasts that total electricity will increase four times between 2018 (200 TWh) and 2040 (800 TWh), but universal access to electricity would require double that amount [11]. When matching these needs with climate change mitigation needs, which is already affecting Africa [12], and the need to reduce the human impact on nature, emerges the need for appropriate energy planning, alongside capillary data to feed models able to capture local dynamics.

Traditional approaches to support electricity access by extending the national infrastructure with new large power plants to reach also remote areas have often led to prohibitive costs [13]. On the other hand, decentralized and distributed energy sources have appeared as a promising economic solution to provide services in both urban and rural areas [14], as they can provide stability services to the grid and even backup service in case of outages. In rural areas, off-grid systems, in the form of microgrids or home systems, have been proposed as a cost-effective measure to speed up the electrification process [15] and even beyond, in some countries, such technologies have been explicitly accounted for in the electrification masterplans [16].

Decentralized solutions, microgrids in particular, have appeared as a suitable technology [17], [18]; nevertheless,

their technical design is the result of sizing procedures that are highly dependent on the local situation and there is no “one-size fits all” solution, as stated by Sustainable Energy for All (SE4ALL) [19]. In particular, the understanding and proper estimation of the load demand that the system is supposed to meet is of crucial relevance for the proper sizing of the system itself [20]. However, the procedure to estimate the future energy demand of non-electrified areas is based on collection of data that can be expensive and complex to carry out [21], as it requires surveys collected by local personnel that shall visit the sites for relatively long periods, without the certainty that the investment will turn economically viable. Once data have been collected and used for a specific purpose, they are usually discarded and even not persistently stored. However, similar operations may be often performed in parallel by different institutions, organizations, and companies – not necessarily in mutual competition – for different purposes, e.g. load assessment [22], population census [23], food or water access [24], among others. This leads to significant waste of resources, loss of efficiency and delays that developing nations cannot afford. On the contrary, provided that appropriate data sharing, privacy and security issues compliance are guaranteed, such data may be shared among different organizations so to leverage on the mutual synergies and foster developments in the electricity access field.

Data needed for estimating load demand at local scale are required to be very capillary, thus the efforts in collecting them, processing and storing cannot be overlooked. For this reason, models have been proposed for estimating energy demand based on proxy data, that can require only generic information at village level [25], while more detailed approaches need for appliance-level details of each household [3], [26] or their aggregates [21]. However, the overlap of these informations with standard surveys for other scopes [23], [24] is quite astonishing, which suggests that there are strong synergies that can be exploited with an appropriate data platform. Given the diversified type of data, that platform should be flexible to accommodate different type of information.

Recognizing such challenges, some institutions have attempted to share data to support electricity access initiatives. The World Bank has not only provided an open platform to share data on power networks in Africa, namely the Africa Electricity Grids Explorer (AEGE) [27], but has also promoted initiatives such as the Global Electrification Platform (GEP) to support investment scenarios to achieve universal electricity access [9]; scenarios are built using credible assumptions and available data. The World Resources Institute has promoted the Energy Access Explorer (EAE) to visualize the state of energy access using credible public data [8]. These tools, which are compared in Table 1, are intended as a support for decision-making and contain aggregated data, often based on assumptions and models. However, they lack the ability to (i) enable users to upload their own data and (ii) to store mixed georeferenced,

TABLE 1. Comparison of current data platforms.

	Stored data types				Resolution of data		Users contributions
	Socio-economic	Appliance	Aggregated Demand	Time-series	Granularity	Consuming unit	
AEGE [27]	no	no	no	no	Geometries	no	no
GEP [9]	yes	no	yes	yes	Raster (>0.008 deg)	no	no
EAE [8]	yes	yes	yes	yes	Raster (various)	no	yes
This study	yes	yes	yes	yes	Geometries	Village, Household, Appliance	yes

socio-economic, and load demand data at local scale, which are important for in-depth analyses as needed for system planning. Table 1, highlights how the existing platforms lack in granularity of information, missing data regarding load curves as time-series, appliances adoption and use at household level, and present their information only at aggregated level, distributed over geospatial rasters. As outlined in [28], the collection of data at different granularity is quite complex, and requires exploring different sources, often not harmonized, as confirmed by the about 20 references needed to collect 61 load profiles of rural microgrids. Data on social behaviors and local economics are sometimes available as results of field surveying campaigns, conducted by International Bodies, such as the Multi-Tier framework by ESMAP [29] or National Statics Bodies [30]. However, they are often distributed across multiple independent databases, with different formats and nonuniform energy consumption data.

As a consequence, any institution, business, or organism in need of energy demand data at local scale finds a limited number of different data sources, that contain data with different formatting styles, and inconsistent sets of information. In addition to a very complex data availability landscape, data sharing is a nonstandard practice for organizations performing data collection themselves and willing to share them for lack of incentive, know-how, or excessive constraints to be respected. This suggests that a well-conceived and well-maintained platform with the purpose of data aggregation and sharing could be very helpful in overcoming these barriers.

C. CONTRIBUTIONS

In this paper, we propose a comprehensive procedure that identifies guidelines for the development of a data-sharing platform for energy access, based on an extensive literature analysis, develops the design of the architecture of the platform, and proposes a prototype implementation. The specific novelties of the paper are:

- 1) Definition of guidelines for the development of data-sharing platforms for energy access
- 2) Extensive literature review to characterize the determinants of demand and requirements of the platforms
- 3) Design of the architecture of a data-sharing platform, based on a NoSQL database
- 4) Development of a working prototype of the data-sharing platform using a web-based interface developed with Angular.JS and Spring

The comprehensive procedure to define the requirements for data sharing platforms in microgrids and energy access, to design the architecture and to develop the prototype is an absolute novelty and is the major contribution of this paper.

D. ORGANIZATION OF THE PAPER

In Section II, a literature analysis of the major players and needs in energy access is discussed in order to introduce the detailed discussion on the key determinants for load assessment in Section III. Section IV summarizes the proposed guidelines that guide the functional design of the platform in Section V and its implementation in Section VI. Then Section VII describes the prototype developed in this study. Finally, some conclusions are drawn in Section VIII.

II. CHALLENGES IN ENERGY ACCESS

A. NEED FOR DATA

Energy access interventions may involve several different activities, such as improving electricity access through grid extension or off-grid solutions [31], improving cooking technologies [32], supporting the use of alternative fuel sources [33], among others. Most large projects aim to address large-scale issues with regional scope, e.g. electrification masterplans [16], but their appropriate planning and implementation need accurate and detailed data, also for realistic cost estimation [34], [35]. Moreover, even in large scale modelling where approximations are often used to make the problem treatable, careful analyses shall be executed to properly tailor the approximations and minimize the errors with respect to the complete model.

In the case of small-scale projects, these problems are exacerbated. While good estimates of renewable energy production are generally available worldwide [36], local-scale demand data are rarely available in developing countries, often not even in developed countries. However, demand is a very critical and uncertain parameter for most projects with uncertainties beyond $\pm 300 - 500\%$ the demand of the first year [17], which leads to high risks for investors. In fact, system sizing is intertwined with demand assessment, which shall characterize energy needs throughout the project lifetime, spanning several years [18] and has traditionally been performed through field pre-electrification interviews. This method, however, suffers from high levels of inaccuracy [15], thus risking to expose projects to the threats explained above. On the other hand, improper sizing can enhance financial or social problems: undersizing the system easily leads to overusing

the components [35], hence compromising their lifetime or increased fuel consumption, while oversizing puts at risk the financial sustainability, since revenues could not cover the costs [34]. Models can play a crucial role in obtaining adequate designs [37], therefore appropriate load assessment techniques are needed. Accordingly, a growing body of literature has emerged on the topic of models for energy planning [17], [18], [38] and load prediction [39], especially for areas with nearly no electricity demand at all. In such conditions, energy estimation can be very difficult because current demand is not measured and must be estimated using other methods, often using public datasets and surveys, yet the corresponding data collection must be designed properly to effectively represent the local conditions [21]. Notably, bottom-up models, which estimate the energy demand based on the analysis of the appliances owned by the households and their typical usage, can guarantee high-accuracy assessments, thanks to their data-driven approach which allows better capturing local dynamics [21]. However, since the analyses are case-dependent and very detailed, this can be a limiting factor for their wide use and accuracy, as data may often be missing [38].

These hurdles, anyway, are not electricity-specific and extend to the whole broad context of energy access, where the availability of electricity is a big part of. Actions for clean cooking facilities must account for local energy sources, and the acceptance and price of technologies, among others [40], [41], [42]. Even in that case, the most technologies to be used, the actions to be carried out to stimulate growth and the geographic locations to prioritize are critical aspects that require detailed Geographical Information Systems (GIS) data to successfully capture the nexus between energy demand and the characteristics of the site [43].

For all these reasons, local detailed data are needed for both large and small-scale projects, including the GIS referencing.

In the context of sizing off-grid energy systems for providing access to energy, proper estimation of load demand is a key aspect for successful planning, but data for estimating load demand of unserved areas results, at the very least, problematic. The main challenges identified in the specific context of data availability and data sharing are related to (a) the presence of different stakeholders with different expectations, access to data and policies for sharing, (b) lack of consistent data structure of collected data across the few existing sources and (c) poor data quality when data is available, requiring data analysis efforts to obtain meaningful demand data for planning purposes.

B. DIFFERENT STAKEHOLDERS INVOLVED

The field of energy access is characterized by the presence of various types of players with different interests [39], that can be reconnected to the quadruple helix theory [44] and summarized in the following. Different stakeholders are characterized by the need for different types of data, according to their nature and interests, they have ownership

of different categories of data and are often characterized by different levels of propensity to share them. This is why it is relevant to have clear in mind when planning a data sharing platform, the different stakeholders potentially involved and their characteristics to maximize its usability.

1) PUBLIC ORGANIZATIONS

This category includes the whole public sector, hence national and local governments, public national agencies, and public utilities, when state-controlled, which is quite common in most developing countries. Their role is to provide regulation, taxation, and funding to steer initiatives towards the citizen prosperity.

Supranational institutions often provide funding and/or know-how. Accordingly, several organizations, such as WorldBank, World Resources Institute, and GiZ, among others, have actually supported and promoted large-scale projects [8], [45], [46].

2) CIVIL SOCIETY ORGANIZATIONS

Civil society means organizations of citizens, such as foundations, Non-Governmental Organizations (NGOs), among others, that provide support usually with a humanitarian or education scope. Their intervention logic is often characterized by a focus on the specific local context, using a participatory approach aimed at achieving people's empowerment through direct involvement [47], [48], [49].

3) ACADEMIA

Universities have historically had as core mission education and research to advance the state-of-the-art, but recently have increasingly participated in field projects alongside industry, government, and NGOs to promote synergies [49]. That's why it is not uncommon for universities to be involved in energy access projects [50], [51].

4) PRIVATE PLAYERS

The private sector has traditionally taken part in all practical project deployment, including the electrification challenges, with particular attention to those with relative high economic return and low risks. Areas less attractive for business activities, such as remote regions, have often been underserved and historically most of the actions were deployed by private players with support of donations, charity branches of multinational companies, or public or supranational grants. The risk of decentralized electrification has been for long time too high to allow for profitable investments or the creation of private players in the field without public intervention, yet this trend is recently changing [52].

C. INCONSISTENCY OF DATA STRUCTURES

Although institutions are progressively opening their data, several hurdles still remain. Unlike the case of weather data which are widely available, and in harmonized form, thanks to satellite data provided by national agencies [53]

or, in selected locations, weather stations, for what concerns demand data, few datasets are available, and when available, collect information in disharmonious forms.

Moreover, data at country scale are rarely publicly available worldwide, and even IEA lacks some data for some countries [54]. Final energy uses and data at local scale are even more scarce [28]. A recent paper [28] proposed a first classification of load demand for about 60 isolated microgrids in developing countries and publicly released the data. However, even in that case, load information was often limited to few representative daily curves, and rarely multi-year dynamics, which are critical for investments, were found in the literature. As observed in the paper, sources of data are often scattered and hence difficult to use and compare in researches that shall first collect them, activity that is very time-consuming. Some consumption data and information are sometimes reported as secondary output of national reports [55], organizations reports [56], general dissemination by private companies [57], or scientific outputs [28], [58]; however, they are often scattered and partially incomplete. On the other hand, the availability of socio-economic information is often limited by public census information [30], information released by NGOs, scientific studies [59], [60], or country-wise statistics provided by super national institutions [61], [62], [63]. However, their format is largely different, as they are collected from different sources tailored each time to their scopes, information is not stored consistently in a unique platform, and data are sparse and not complete, never matched with consumption data, hence limiting the possible impact. One of the most relevant campaigns to assess the energy consumption and living standards of rural populations, in terms of socio-economic parameters, is represented by the ESMAP Energy Access Diagnostic Reports based on the Multi-Tier Framework [64]. This data collection effort culminated in the production of reports and related databases (e.g.: [61], [62], [63]). The issue related with the produced material is that each single national survey collects different information, and the same categories of information are categorized differently across them. ESMAP itself conducted a second set of surveys, in the framework of the Living Standard Measurement Study Plus (LSMS+) [65], that collected different information, and when the same, in a different format. An example of this inconsistency can be found in how the information “Education Level of Household Head” is reported across all different questionnaires: 65 different observations can be found that do not match the 11 categories of the International Standard Classification of Education ISCED of 2011 [66].

All these issues hamper the ability to further investigate the nexus between demand and socio-economic characteristics that can improve efficiency in investment and speed up the deployment of the analysis. Recent national and supranational efforts have provided tools for simplifying energy planning, such as the Global Energy Platform [9] or the Web app for electrification for Nigeria [67], which are good instruments to support decision-making, but their data

assessment is not based on granular information, and they often base their analysis on the Multi-Tier framework [64].

For these reasons, in this study we propose a standardized methodology that is able to consistently store both demand and socio-economic information with GIS location, so to ease the data analysis, planning, and policy tailoring.

D. POOR QUALITY TO FEED DEMAND ESTIMATION MODELS

Energy demand is a critical input data for energy models. However, in the context of rural electrification, measured electricity demand data are rarely available. Therefore, accurate assessment techniques have been developed to account for the estimation of both the load profile, which is especially critical for sizing off-grid systems, and its growth over time [17], [18], [38] conditioned to the local socio-economic conditions that have a paramount effect [21], [68]. In the literature, various techniques have been used [22], [59], [60], [69]. Louw et al. [59] estimated the average demand of two rural villages using a log-linear regression model based on survey data: demand has been found to be inelastic with respect to price. Hence, they suggested that sizing methodologies shall be cost-based. Regression has also been used in [60], and compared to Artificial Neural Networks (ANNs), for an Iranian case study; ANNs turned out to increase estimation accuracy. Dominguez et al. [69] studied drivers for electricity consumption in rural Kenya, accounting for access technology (no access, home systems and grid connection) and transition probabilities. Other studies [15], [22] performed survey-based methodologies, yet authors in [22] proposed a simplified approach to reduce uncertainties in the results by survey methodologies and [15] focused on estimating long-term dynamics in the demand. Notably, the Remote-Areas Multi-energy systems load Profiles (RAMP) model by Lombardi et al. [26], based on LoadProGen [70], provides a probabilistic methodology to estimate demand by employing a bottom-up approach: by using survey data to tailor appliance adoption and usage, a probabilistic technique aims to identify and aggregate the overall load profile for each community.

Overall, most of the methodologies relied on socio-economic information at granular level, which are not easily available [3], [28], [71]. Furthermore, when input parameters are lacking or of poor quality, the entire energy modeling is compromised, regardless of the quality of the demand assessment tool [21]. Low data quality will propagate throughout the energy model down to the results, thus hindering any policy or technical decision.

Therefore, data platforms for energy access shall be able to properly store these types of information in good quality alongside measurements of existing energy demand, when available, to properly support energy studies, planning, and policy analyses. Furthermore, the same information could also support socio-econometric studies beyond energy access itself. The platform shall be openly accessible to support

collaboration and avoid synergies across the data collectors, hence supporting efficiency and quality.

In Section III, a detailed review of energy assessment studies for planning is performed to identify the main determinants to estimate energy demand, which should be prioritized to be included in the data platform approach.

III. DETERMINANTS OF DEMAND

In order to properly identify the types of information needed for the proposed data platform, a deep literature analysis has been performed to identify the main determinants of energy consumption at household level, with the goal of drawing guidelines for data-sharing platforms. This information is then categorized and prioritized.

A. LITERATURE ANALYSIS PROCEDURE

Given the strong need for demand estimation in energy access applications, a wide literature analysis was carried out with the goal of identifying the major relevant drivers that shall be included in the proposed data platform. Drivers will then be prioritized and a selection of them will be considered as key inputs for the platform, as detailed in Section V.

Papers related to drivers for household energy consumption, appliance adoption, load profile classification and modeling, energy planning, and system dynamics for rural areas have been reviewed with the goal of understanding the major inputs needed for energy assessments. Scopus engine [72] has been used while focusing the research on journals and conferences related to energy and sustainable development.

B. THE CLASSES OF INFORMATION

Several reviews focusing on social dynamics and longitudinal socio-econometric studies highlighted that the major classes of information for demand assessment can be grouped into the following major categories [3], [28], [39], [71], [73]:

- 1) Socio-economic data, such as population information, job type, income, culture
- 2) Dwelling factors, such as the type and quality of the dwelling
- 3) Appliance, such as ownership, usage pattern
- 4) Geographical information, such as location, proximity to major point of interest
- 5) Supply data, such as type of connection, tariffs
- 6) Alternative energy sources
- 7) Past demand

Kuster et al. [39] reviewed beyond 100 forecasting models for electricity demand, also highlighting the issue of the many data required for bottom-up approaches devoted to long-term prediction of electric demand. The paper also classified papers based on socio-economic, weather, building and occupancy, and past demand information. Similar considerations have been obtained in the review by Jones et al. [74], yet including appliance factors, which are widely adopted in some bottom-up forecasting models [26], [70].

Accordingly, the proposed data platform for energy access, which is the objective of this study, shall account for the above-mentioned classes. In particular, a detailed literature analysis will be performed to identify the specific drivers in these classes that are more relevant for estimating demand for energy access applications.

C. DRIVERS

A total of 34 papers have been analyzed. The main determinants of energy demand are reported in Table 2, and classified according to the information classes defined in Section III-B. In particular, every driver shall describe the characteristics of the village (V), the household or consumer (C) or an appliance (A), as reported in the table. This information is critical for the proper definition of the data platform and hence reported.

The proposed literature analysis has highlighted 70 main drivers, some of which group multiple similar questions that have been used by scholars. The largest fraction of drivers is indeed socio-economic and supply-related. In fact, as discussed in the literature, the behavior of the community, the financing options and local availability of capital play an important role in the appliance adoption and the productive activities, which support socio-economic growth, but, usually, higher electrical consumption needs shall be met. On the other hand, to properly investigate and develop energy studies to model the energy nexus, past historical data shall be stored as well. In particular, to properly foresee the energy demand, time series of the consumption at village, household, or appliance level would be a nice-have, so that better studies could be performed, provided that enough data are gathered. Moreover, the platform shall be flexible enough to accommodate various types of information and to ease the addition of different kind of data, e.g. different new drivers, in case of need. This confirms that data platforms for energy access must be able to store various not-predefined socio-economic information along with energy demand information.

Furthermore, it turns out that a large variety of such data are obtained by local surveys, although some entries (e.g. urban/rural location, distance with respect to infrastructures, among others) could be obtained by post-processing different inputs and in particular the GPS location of the survey. Accordingly, the GPS location is an important feature that databases for demand assessment shall be able to store.

IV. GUIDELINES FOR DEVELOPING DATA SHARING PLATFORMS IN ENERGY ACCESS

According to the proposed literature analysis in Section VI, we summarize below the major recommendations that data platform shall meet:

- 1) multiple stakeholders shall have access to data;
- 2) the data have different types, including numeric, string, list of elements, etc.;

TABLE 2. List of drivers for estimating electricity demand. For each driver, we report the class, the element which is related to (V: Village, C: Consumer and A: Appliance), its inclusion status in the data platform (Y: Yes, YM: Yes using a proxy, N*: not explicitly, but the platform can handle it) and the references.

Id	Driver name	Driver class	Entity	Included	References
1	Connection type or fee	Supply	V	Y	[3], [43], [68], [75]–[78]
2	Electricity price	Supply	C	Y	[60], [68], [69], [75], [79]
3	Installed power capacity of the system	Supply	V, C	Y	[28]
4	Hours of available electricity	Supply	C	Y	[3], [68], [75], [76], [80]
5	Pre-paid/post-paid tariff (tariff scheme)	Supply	V, C	Y	[28], [79]
6	Tariff type, method, or frequency	Supply	V, C	Y	[28], [81]
7	Time passed since the electrification	Supply	V, C	Y	[58], [59], [71], [82]
8	System management (business model)	Supply	V	N*	[28]
9	Number of connections of the system	Supply	V	N*	[28]
10	Prospect of grid access in the next year	Supply	V, C	N*	[83]
11	Share of households by energy access (grid, SHS, etc.)	Supply	V	N*	[69]
12	Capacity building implemented	Supply	V, C	N*	[3], [68], [75]
13	Relative wealth index (GPS location proxy)	Socio-economic	V	YM	This Paper
14	Urban/rural location (GPS location proxy)	Socio-economic	V	YM	[43], [71], [80]
15	Recently affected by a natural disaster (GPS location proxy)	Socio-economic	V	YM	[83]
16	Employment (Working hours proxy)	Socio-economic	C	YM	[43], [84]
17	Nighttime lights from satellite imagery (GPS location proxy)	Socio-economic	V	YM	[43], [76]
18	Presence of a household business (Working hours proxy)	Socio-economic	C	YM	[3], [68], [69], [75]
19	Size of business activity in terms of employees	Socio-economic	C	Y	[83]
20	Age of household head	Socio-economic	C	Y	[69], [76], [78], [80], [83]
21	Education level of household head	Socio-economic	C	Y	[76], [78], [80], [83]
22	Income of the household	Socio-economic	C	Y	[3], [43], [59], [78]–[80], [83], [85]
23	Number of people/adults/youngsters/elderly in the household	Socio-economic	H	Y	[59], [69], [78], [80], [82], [83], [86]
24	Number of working people within the household	Socio-economic	C	Y	[78], [84]
25	Size of business activity in terms of revenues	Socio-economic	C	Y	[83]
26	Age of business activity	Socio-economic	C	N*	[20]
27	Number of years the household lived in the community	Socio-economic	C	N*	[78]
28	Population density	Socio-economic	V	N*	[76], [87]
29	Ratio of females within the household	Socio-economic	C	N*	[83]
30	Ratio of household members with high education level	Socio-economic	C	N*	[83]
31	Business hours of micro-enterprise	Socio-economic	C	N*	[83]
32	Cooking fuel usages	Socio-economic	A	N*	[78]
33	Information on the owner of micro-enterprise	Socio-economic	C	N*	[83]
34	Monthly energy expenditure of household	Socio-economic	C	N*	[78]
35	Ownership of a SHS before grid connection	Socio-economic	C	N*	[69]
36	Access to credit in the past	Socio-economic	C	N*	[3], [59], [68], [75], [83]
37	Dimension of land owned	Socio-economic	V	N*	[43], [83]
38	Monthly household expenditure	Socio-economic	C	N*	[3], [68], [75], [82], [87], [88]
39	Ownership of large livestock	Socio-economic	C	N*	[69], [76]
40	Ownership of motorized vehicle	Socio-economic	C	N*	[69], [76], [78], [80]
41	Ownership of small livestock	Socio-economic	C	N*	[69], [76]
42	Presence of street lighting in the neighborhood	Socio-economic	V, C	N*	[76]
43	Respondent category (Household, business, etc.)	Socio-economic	C	N*	[14], [26], [28], [81], [89]
44	Seasonality of business activity	Socio-economic	C	N*	[89]
45	Sex of household's head	Socio-economic	C	N*	[69], [76], [78], [80], [82], [83]
46	Size of public service in terms of area occupied	Socio-economic	C	N*	[90]
47	Socio-economic status of household head	Socio-economic	C	N*	[69], [76], [78], [83]
48	Type of business activity	Socio-economic	C	N*	[14], [26], [78], [89]
49	Type of public service	Socio-economic	C	N*	[14], [26], [43]
50	Monthly electricity consumption	Past-demand data	V, C, A	Y	[91]
51	Monthly electricity expenditure	Past-demand data	C	N*	[88]
52	Climate zone	Geographical	V	Y	[28]
53	Distance from the nearest city	Geographical	V	Y	[43], [82]
54	Altitude	Geographical	V	Y	This paper
55	Agriculture-related information	Geographical	C	N*	[43]
56	Floor space of the dwelling	Dwelling	C	N*	[86]–[88]
57	Number of rooms of the dwelling	Dwelling	C	N*	[59], [69], [78], [80], [83], [86], [87]
58	Ownership status of the dwelling	Dwelling	C	N*	[78], [80]
59	Quality of the dwelling (roof/wall material, etc.)	Dwelling	C	N*	[69], [76], [78], [80], [84]
60	Marital status of household head	Cultural	C	Y	[78], [83]
61	Religion or Race	Cultural	V, C	Y	[80], [82]
62	Number of hours studied per night by school children	Cultural	C	N*	[78]
63	Bedtime and wake time	Cultural	C	N*	[78]
64	Presence of a smoking person within the household	Cultural	C	N*	[83]
65	Price of appliance	Appliance	A	Y	[60], [80]
66	Affordability of appliance	Appliance	V, C	N*	[80]
67	Planned appliances	Appliance	C	N*	[78]
68	Price of energy sources	Alternative sources	A	Y	[3], [59], [69], [75], [78], [83], [84]
69	Main lighting fuel	Alternative sources	C	N*	[84]
70	Type of main cooking fuel	Alternative sources	C	N*	[69], [76], [78]

N*: not explicitly included, yet the platform shall be flexible to accommodate it

- 3) the key determinants of energy demand relate to consumers, including households, the village characteristics and the appliances owned by the consumers;
- 4) data shall be easy to access, download, and upload with limited barriers: a web-based interface may be preferred;
- 5) data shall be accessible by multiple users at the same time;
- 6) the storing platform shall be able to accommodate custom data fields beyond those in the initial design;
- 7) data shall be of good quality.

These considerations emerge as guidelines for the data platform design and implementation, described in the following sections. The proposed guidelines, the procedure used to derive design requirements of the data platform, and the proposed prototype described in the following, are novel contributions that can drive the development of novel platforms with respect to the state-of-the-art highlighted in Table 1.

V. PLATFORM DESIGN

This section shows details on the design cycle of the proposed data management platform. First, the scope and the requirement analysis are introduced. Second, the major actors and their level of authorization are discussed. Third, a brief description of the data that will be managed by the platform is provided.

A. SCOPE AND REQUIREMENT ANALYSIS

The initial stage of the platform design is the detailed analysis of the requirements that the software must satisfy.

The main goal of the platform is to support users, depending on their level of authorization, in sharing and visualizing data of variegated social, economic and technical information and usage pattern of resources for energy access purposes, at different level of disaggregation: Village-, Consumer- and Appliance-level. Examples of the desired data include time series of energy consumption (for villages, single consumers or appliances), geo-referenced information of population, appliance preferences; more details can be found in Section III.

The platform has to allow users to share and exploit data stored in its database. Specifically, it has to support users in executing the following operations, depending on their level of authorization:

- 1) To download and upload data in the form of both description of the different involved data entities and of time series of energy consumption;
- 2) To enable data review and validation of new incoming data;
- 3) To provide tools to support the search for major relevant information in the database;
- 4) To analyze aggregated information on the usage pattern of villages, consumers, and appliances;
- 5) To manage users and different levels of authorization;

TABLE 3. Rights by user.

Type of user	General	Download	Upload	Verify	Manag.
Unregistered	✓	×	×	×	×
Basic	✓	✓ ¹	×	×	×
Authorized	✓	✓	✓	×	×
Moderators	✓	✓	✓	✓	✓ ²
Administrators	✓	✓	✓	✓	✓

¹May be limited.

²Limited: no changes to *Moderator* and *Administrator* levels.

- 6) To visualize statistics on the quantity of data stored in the database and the major contributors;
- 7) To provide a reward point-based mechanism to grant access to the records of the database.

As regards the uploading features, they may be limited to a two-fold approval by moderators: moderators enable only a subset of users to provide data and moderators must approve the new entries. This is aimed at increasing the data quality of the stored information and providing an expert filtering on the input data.

To encourage the upload of data, visibility shall be given to contributors according to their data license. Thus, we expect that: (a) the platform shall clearly state the contributors, (b) the data licensing by users must be respected, (c) the visibility and recognition to contributions must be provided when such data are downloaded, and (d) incentive measures to support data upload may be suggested. To adhere to (d), in this study, a reward point-based approach is proposed: to get access to extra data, such as more recently uploaded data, users shall use reward points; reward points can be earned by uploading new data that must be validated by moderators.

Moreover, the application shall be highly available, have fast response, and be easy to use by the general audience target of the application.

B. THE USERS OF THE PLATFORM AND THEIR NEEDS

The users of the proposed platform may be divided in five main categories with different roles, namely: *Administrators*, *Moderators*, *Authorized Users*, *Basic Users* and *Unregistered Users*. In Table 3, we show the rights of each user. The first column represents the user type; the other columns show the different rights. “General” stands for the ability to read general statistics of the data; “Download” and “Upload” rights aim to enable downloading or uploading of selected data from the platform. The “Verify” right has the scope to enable verification tools of new data: if a user uploads new data, they will be actually considered available for other users only if a moderator will check and approve them. Finally, the “Management” right gives the authority to change the rights allocated to the users.

Any new subscription of a *Basic User* shall be authorized by an *Administrator* or a *Moderator*. *Basic Users* can only access and download the data of the platform, but they cannot upload data, which can be done instead by *Authorized Users*, *Administrators* and *Moderators*. *Administrators* and *Moderators* have the rights to change the user status, such

as upgrading or downgrading *Authorized Users* and *Basic Users*, as well as banning them. Only *Administrators* can modify (upgrade or downgrade) the rights of the other types of users. *Unregistered Users* can only access public statistics data in the homepage.

C. THE DATA TO STORE

The application shall manage the data presented in Table 2, which represents the major drivers for energy demand assessment. It is worth noticing that villages, consumers, and appliances represent the major subjects of attention for energy demand studies, each of them being characterized by multiple types of information. Indeed, in this study villages, consumers and appliances are also referred to as Energy Consume Entities (ECEs).

Each ECE is characterized by a number of generic attributes and time series representing usage patterns of resources. Attributes may include numeric information (e.g. number of people, rooms, etc.), categorical data (e.g. type of business, quality of the dwelling, etc.), or general text. Electrical tariffs for villages and households shall also be characterized, as well as any existing generation devices. Time series shall characterize measured data, such as metered electrical consumption, or usage habits that describe usage pattern of appliances, for example. The number of attributes and time series can be arbitrary for each ECE. To properly characterize the data, a consumer shall belong to a village and, similarly, an appliance shall belong to a consumer.

Since data are shared by users, the data license for each entry shall also be stored, so that when data are downloaded the references are provided as well.

VI. PLATFORM ARCHITECTURE AND IMPLEMENTATION

In the following, we discuss the architecture of the proposed platform, including the major details of its implementation as a web application.

A. PLATFORM ARCHITECTURE

Fig. 1 shows the architecture of the proposed platform. It is a classical Model-View-Control (MVC) architecture [92] in which the user interact with the platform by using a client-side web application developed.

The web interface has been developed using the AngularJS framework¹ by implementing the appropriate Angular components, including views and services for data handling. View templates represent a blueprint for the web interface used by the user, and they have been used to fast-track the application development. Finally, appropriate services have the goal to interface the web interface and the back-end developed using MongoDB and Spring, as later discussed.

Interactions between the web application and the data stored in the database are managed by a back-end software layer implemented using the Spring and Spring Boot

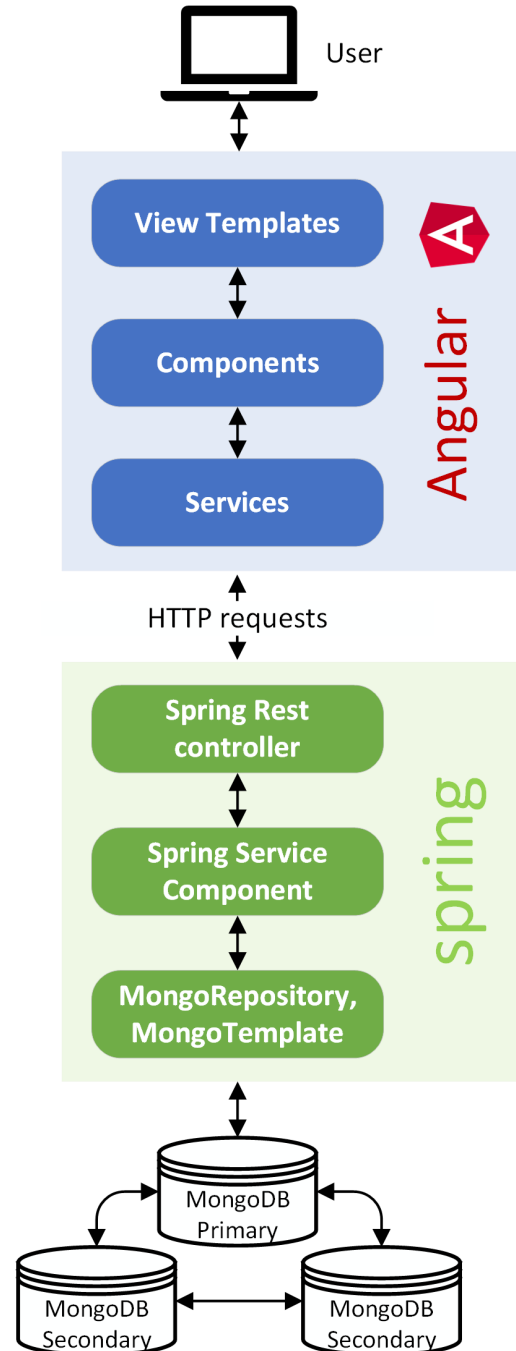


FIGURE 1. The software architecture of the proposed platform.

frameworks.² Thanks to these two frameworks, both the JAVA classes and the web server can be easily created, thus facilitating the development, the update, and the deployment of the application. The communication between the web application and the back-end layer are based on REST services [93]. The main characteristics of this type of communication are:

- 1) A REST service to provide resources but no methods
- 2) Data are in JSON format

¹<https://angular.io>

²<https://spring.io>

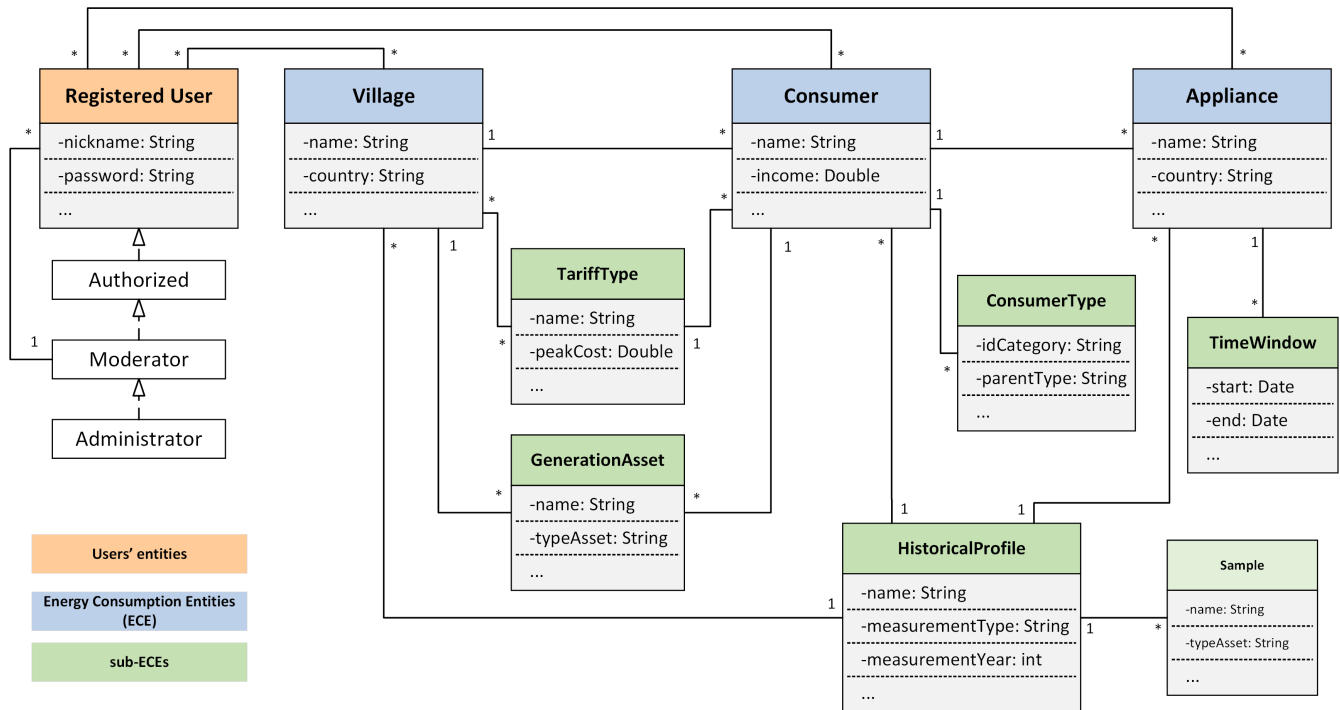


FIGURE 2. Class diagram of the entities; orange classes are related to the management of the users, blue classes refer to the ECEs and the green ones are the sub-ECEs to populate specific informations of the ECEs.

3) REST model is developed using the HTTP protocol using a Client-Server type architecture.

Finally, data are stored in a distributed NoSQL database, namely MongoDB³ which supports data stored in documents, saved as Json files. Unlike relational databases [94], MongoDB ensures flexibility of the data structures (no preliminary schemes must be defined) and provides great support to data replication ensuring high level of scalability, service availability, and reliability. These features are specifically relevant for the proposed platform, in which data can be characterized by unstructured and even missing information. Moreover, since we aim to offer a high quality of experience to the user when interacting with the platform, low latency and high availability of the service have to be guaranteed. Finally, as we expect a global distribution of the platform, the scalability feature plays a fundamental role.

B. THE UML CLASS DIAGRAM

Fig. 2 shows the main UML class diagram developed according to the requirement analysis described in Section V and stakeholders identification in Section V-B. The figure aims to represent the major relationships between the entities involved in the developed platform, using associations in UML format [95]. Each entity denotes a major cluster of information with several attributes; however, for the sake of brevity the details of the specific attributes for each class have been omitted, but the complete list is reported in Appendix.

A *User* of the platform can be specialized in *Basic User*, *Authorized User*, *Moderator* and *Administrator*, with increasing rights, as described in Table 3. A *Moderator* can *Approve* upgrades of the datasets provided by an *Authorized user*, and the same *Moderator* can *Approve* multiple upgrades (this justifies the multiplicity *one-to-many* (1-*) in the relation between entities *Moderator* and *User* Fig. 2). Furthermore, a *User* initially starts with access to no ECE, but he/she can *Get Access* to ECEs by spending *Reward Points*. Multiple users can *Get Access* to multiple ECEs and one ECE may be accessible by many users (the multiplicity *many-to-many* (*-*) describes the relation between each ECE and *User*). A *Basic User* cannot earn *Reward Points* beyond the amount provided at registration. An *Authorized User* and a more privileged user can earn additional *Reward Points* by proposing new data to be added to the dataset: after validation by one *Moderator*, the data are accepted and the *Authorized Users* earn additional *Reward Points*.

Each *Village* can be composed by a set of *Consumers*, who in their turn can possess a set of *Appliances*. Each *Appliance* is associated with a list of time windows (*TimeWindows*), which specify the typical pattern of usage of a specific appliance. Examples of *TimeWindows* for a lighting device may be during the early morning and/or during the evening.

In order to properly describe any existing power generation asset, *Villages* and *Consumers* can be characterized by a list of generation asset entities (*GenerationAsset*); each of them, if any, describes a generation component, its type, and any information the user may be willing to add. Examples of *GenerationAssets* can be PV systems, inverters, diesel

³<https://www.mongodb.com/>

generators, among others. Similarly, to properly capture and store variegated tariff structures, each *Consumer* can be associated with a tariff type (*TariffType* entity), whereas each *Village* can be characterized by a list of values, if any. Each *Consumer* can have associated a specific type, (*ConsumerType* entity), similarly to an *Appliance* whose type is specified in the *ApplianceType* field.

Given the strong focus of the proposed activity in supporting the nexus between socio-economic and geographical data with time series of power and energy uses, each ECE may *Store HistoricalProfiles*. A *HistoricalProfile* refers to an arbitrary time series of resources associated with the corresponding ECE. As an example, a *HistoricalProfile* can consist in a time series spanning several years of power consumption data for a household or a village. The entity *HistoricalProfile* enables storing general information on the time series that are stored under the form of a list of *Samples*, each of them being characterized by the timestamp and the numeric value of the measurement.

C. FUNCTIONAL WEB PAGES OF THE PLATFORM

In the following, we briefly describe the main web pages that allow the user to interact with the developed prototype of the platform. In particular, the following web pages have been implemented:

- 1) Homepage
- 2) General information: web page where both registered and unregistered users can have access to get general statistics on the stored data
- 3) Sign up page: page where unregistered users can get access to the platform
- 4) Sign in form: page where users can get access to the platform
- 5) Add/Modify tool: pages where data can be added or modified by *Authorized Users*
- 6) Personal data: page where users can get access and modify their own personal information
- 7) Settings section: page where administrators and moderators can modify the user access rights
- 8) Contributors and users: page where statistics on the contributions by the users to the platform are shown
- 9) Search tool: page where users can search information and get access to data; selected data may be also downloaded
- 10) Review page: web page where *Moderators* can accept or reject new entries proposed by users
- 11) Activity page: visualization page in which data aggregation and summaries are reported.

In the following section, with the aim of showing examples of usage of the developed prototype of the platform, we show the snapshots of some web pages discussed above.

VII. PROTOTYPE PRESENTATION AND RESULTS

The proposed prototype has been implemented and tested on a case study populated with open data for energy access, based on the literature review described in Section III.

TABLE 4. Data by ECE contained in the dataset.

	Village	Consumer	Appliance	HistoricalProfiles
Quantity	52	64	632	47
Reference	[28]	[26]	[26]	[28]



FIGURE 3. Home page.

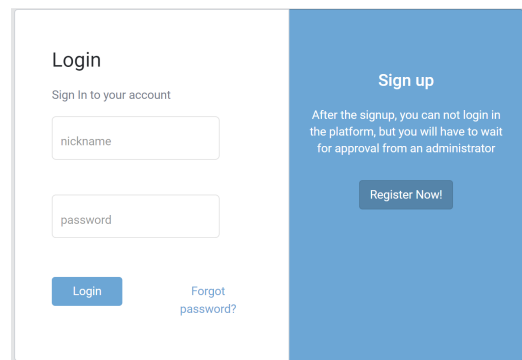


FIGURE 4. Login and sign-up page.

The dataset has been populated with open data of villages and consumers to map diversified but complete information including electric profiles for the communities. In summary, the data reported in Table 4 have been used for testing the prototype.

Figures 4-7 depict selected relevant web pages of the developed prototype. Specifically, in the figures we show, respectively, the homepage, the login and sign-up page, the general statistics page, the interface to add new data and an example of data visualization.

In the homepage in Fig. 3, any user can access the login and sign-up page, as shown in Fig. 4, as well as the general statistics page of the platform, shown in Fig. 5. It is worth noticing that in the latter form, users can be given credit for their contributions: the amount of contributions provided by the most contributing users are highlighted. Once a user is registered and has been promoted to *Authorized User*, it can access the add form shown in Fig. 6, where additional entries can be added to the dataset. New entries to the database refer to the ECE mentioned in the previous subsections, populated with arbitrary attributes. Time series profiles can be also added, as shown in the web page titled “Add Historical Profile”; uploading data from files is also supported. Such information can be then validated by a *Moderator* using the Review tool, in the specific page.

The performances of the prototype have been tested on a normal desktop computer and the responsiveness has

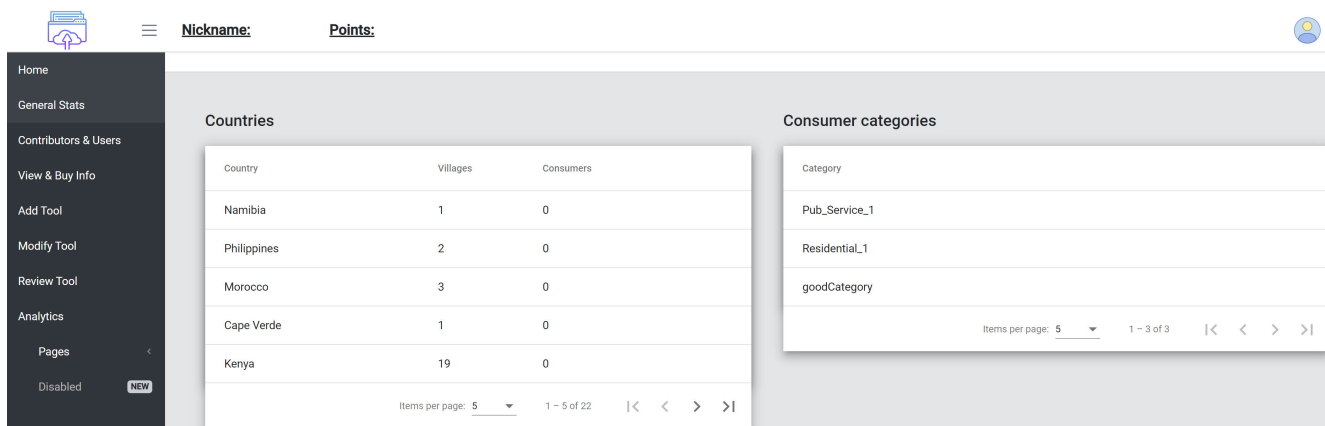


FIGURE 5. General statistics of the application.

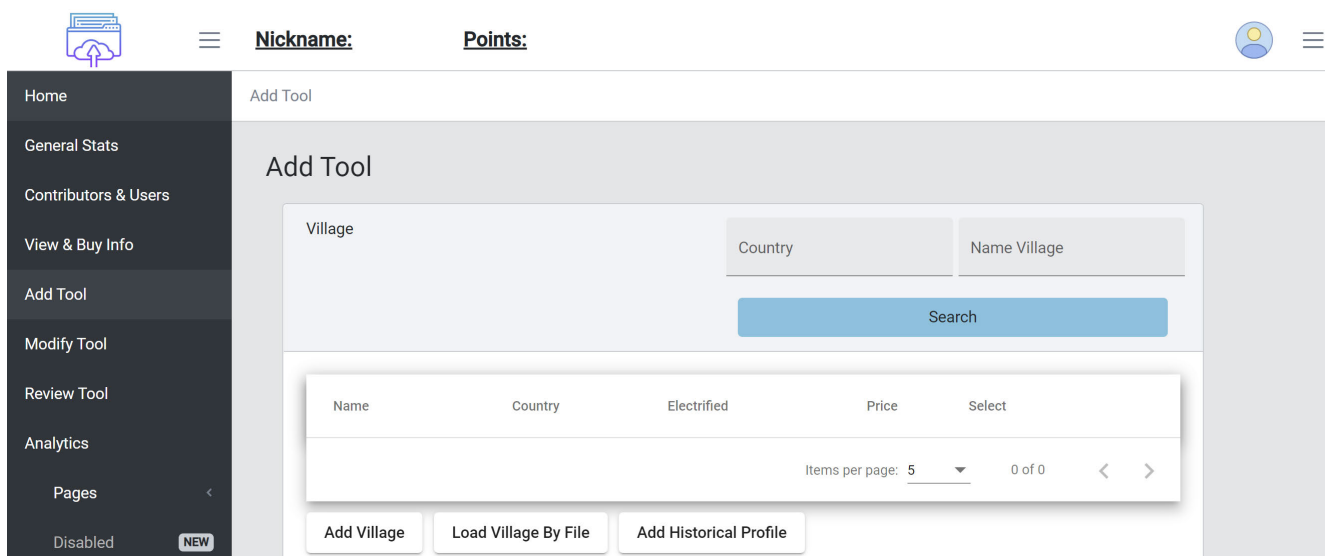


FIGURE 6. Web page to add new data to the tool.

successfully met the desired quality. The back-end and the database have been deployed on a small cluster composed by three workstations. To ease the visualization of the data contained in the database, in the prototype of the web application we have implemented a collection of analytics and statistics such as:

- 1) To show the Average load profiles by ECE at different aggregation level (country, time, etc.);
- 2) To highlight the distribution of the most used tariffs by village;
- 3) To highlight the most frequent time windows where appliances are used.

Fig. 7 highlights an example of visualization for the analytics to show the usage pattern of the data for the Village “El Sena”. For example, the image highlights that the usage pattern of the fridge (in red) is constant across the day, as it is always connected; conversely outdoor lights are generally used only during the night. These type of visualization tools are useful for users of the platform to preliminarily investigate the data and easily extract information.

TABLE 5. List of fields for User.

Field	Description
nickname	Nickname of the user chosen during registration
password	Password chosen by the user (encrypted)
name	User name
surname	User surname
country	User country
email	Private email of the user
public email	Public email of the user shown publicly
institution type	Type of the institution where the user works
institution name	Name of the institution where the user works
contributions	Number of entries added by the user
joining date	Date when the user has been approved
approved by	User that approved the registration of the user
public visibility	Flag to show if the users authorize personal data to be public
banned	Flag to specify if the user is banned
current points	Number of points of the user
authority level	Privileges of the user

VIII. CONCLUSION

The paper proposes guidelines, a comprehensive design procedure, and a prototype implementation of a software data

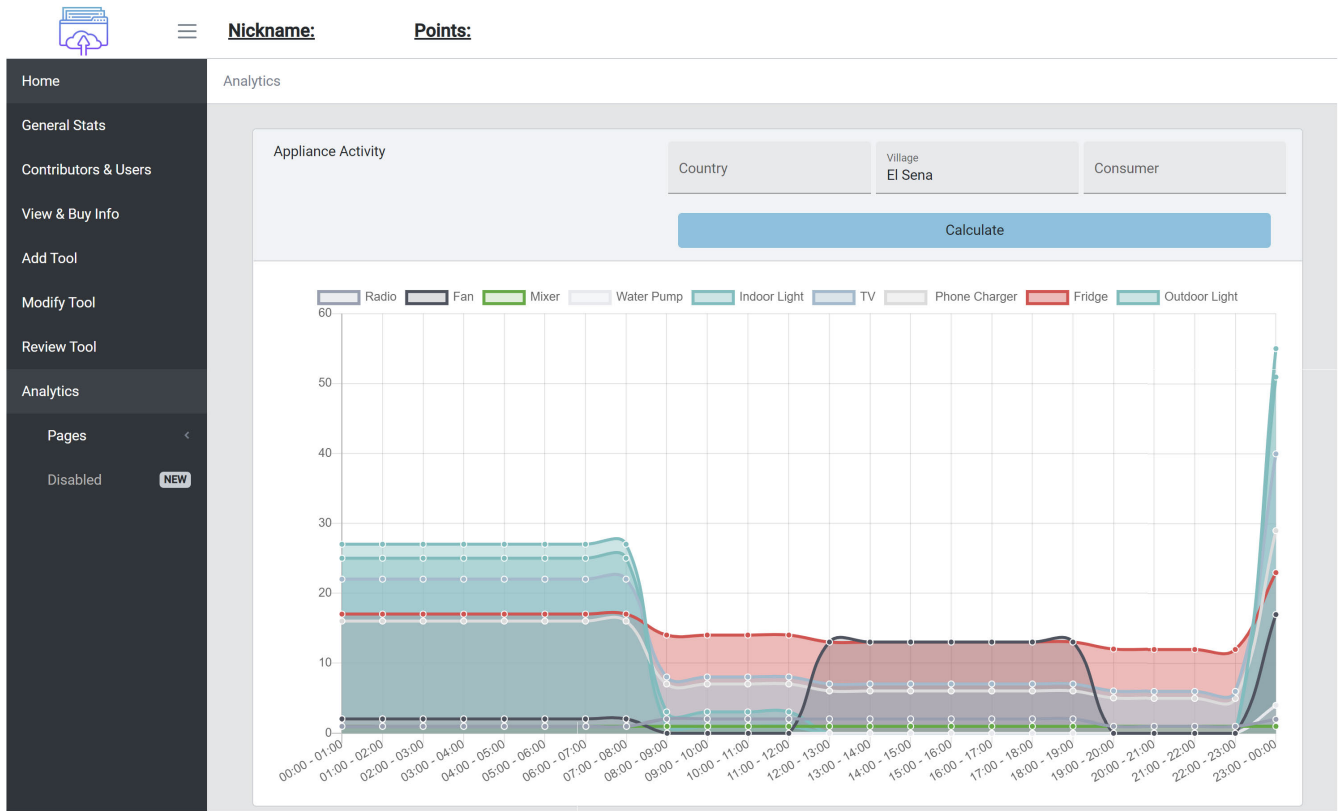


FIGURE 7. Example of data visualization for appliances: the plot highlights the usage pattern of different appliances for a microgrid.

TABLE 6. List of fields for village entity.

Field	Description
entry date	Date when the Village data have been approved
last change	Date when the data have been last modified
first add user	Nickname of the user who first added the entry
first add mod.	Nickname of the user who approved the entry
last change user	Nickname of the user who last proposed any change
last change mod.	Nickname of the user who approved last changes
name	Village name
country	Village country
region	Village region
latitude	Village latitude
longitude	Village longitude
MASL	Meters above sea level of the village
climate	Acronym of the village (Koppen classification)
population	Number of people in the village
surface	Village surface
ethnic	Ethnic group of the village
connection fee	Cost to connect a user to the public grid
hours electricity	Number of hours of electricity service
currency	Village currency
price	Points needed to buy the village data
distance near city	Distance in km of the nearest city beyond 50k population
year access elec.	Year in which the village got electricity access
AC/DC	whether the electrification is in AC or DC
voltage level	Voltage level in the village
frequency	Frequency of the network
references	References used to populate the fields
notes	Additional notes

TABLE 7. List of fields for generation asset.

name	Generation asset name
asset type	Type of the asset
macro type	General classification of generation type
capacity	Capacity of the generation asset
AC/DC	Type of supply
number of units	Number of generation units
yearly production	Yearly production of the asset
references	References used to populate the fields
notes	Notes and description of the Generation Asset

TABLE 8. List of fields for tariff type.

name	Name of the tariff
peak cost	Peak tariff cost
energy cost	Energy tariff cost
fixed cost	Fixed tariff cost
description	Description of the tariff
notes	Notes and description of the tariff type

platform to jointly capture the link between socio-economic, geographical and energy perspectives for energy access

purposes. After a detailed review of the major stakeholders and determinants for demand growth, the paper identifies the consuming entities Village, Households and Appliances as the major classes of data for energy and socio-economic studies, each of them characterized by specific numeric, categorical and time-series information. Correspondingly, a novel data architecture, based on the state-of-the-art NoSQL structure, is proposed to flexibly capture numeric, categorical and time-series data for any of the major consuming entities here identified. The prototype web-based implementation,

TABLE 9. List of fields for consumer.

Field	Description
entry date	Date when the Appliance data have been approved
last change	Date when the data have been last modified
first add user	Nickname of the user who first added the entry
first add mod.	Nickname of the user who approved the entry
last change user	Nickname of the user who last proposed any change
last change mod.	Nickname of the user who approved last changes
id consumer	Consumer name
income	Consumer income
scholarization	Educational qualification of the head of the household
people household	Number of people in the household
hours electricity	Number of hours of electricity service
climate	Acronym of the village (Koppen classification)
working hours	Number of hours the head of the household works
surface	Village surface
price	Points needed to buy the village data
nAdults	Number of adults
nChildren	Number of children
nElderly	Number of elderly
nPpl	Number of people
nWorkers	Number of workers
nEmployees	Number of employees
nAdults	Number of adults
references	References used to populate the fields
notes	Additional notes

TABLE 10. List of fields for consumer type.

Field	Description
id consumer cat.	Id of each category entry
parent type	Parent consumer category
notes	Notes and description of the consumer category

TABLE 11. List of fields for appliance entity.

Field	Description
entry date	Date when the Consumer data have been approved
last change	Date when the data have been last modified
first add user	Nickname of the user who first added the entry
first add mod.	Nickname of the user who approved the entry
last change user	Nickname of the user who last proposed any change
last change mod.	Nickname of the user who approved last changes
name	Name of the appliance
price	Number of points to get access to the entry
affordability	Ratio of appliance price over capital annual expenditure
type	Appliance type
currency	Currency
AC/DC	Type of electricity supply
nominal power	Nominal power
frequency	Network frequency
yearly freq. use	Use of the appliance during a year
efficiency	Efficiency of the appliance
season	Season in which the appliance is used
weekend/weekday	When the appliance is used
frequency use	Frequency usage of the device
average daily use	Average daily use
var. daily use	Variability of daily usage
min. period use	Minimum time in which the appliance is used
references	References used to populate the fields
notes	Additional notes

developed using Angular and Spring, has successfully implemented the proposed architecture, thus confirming the

TABLE 12. List of fields for time window.

Field	Description
start datetime	Start of the usage window
end datetime	End of the usage window
variability start	Variability of the beginning of the window
variability end	Variability of the end of the window
variability length	Variability of the duration of the window

TABLE 13. List of fields for historical profile.

Field	Description
entry date	Date when the Consumer data have been approved
last change	Date when the data have been last modified
first add user	Nickname of the user who first added the entry
first add mod.	Nickname of the user who approved the entry
last change user	Nickname of the user who last proposed any change
last change mod.	Nickname of the user who approved last changes
measur. type	Type of energy that is monitored
measur. unit	Unit of each time-step of the profile
measur. year	Year of the measurement
time resolution	Time resolution of the measurement in hours
season	Season at which the profile refers to
month	Months at which the profile refers to
weekend/weekday	Period of the week the profiles refer to
references	References used to populate the fields
notes	Additional notes

TABLE 14. List of fields for sample.

Field	Description
date	Date of the sample
time	Time of the sample
power	Measurement of the sample (e.g. power)

usability and flexibility of the proposed platform, as also shown with the views of the application. Performances are in accordance with the state-of-the-art and they prove the possible use of the proposed architecture for energy access.

This activity lays the foundation for the development of data platforms for energy sector that can be easily employed in other energy applications, e.g. degradation dynamics of batteries, social behavior in the use of electric vehicles, but also beyond the energy sector, including – but not limited to – health, poverty statistics, water access, consumers behavior among others. Further studies shall finalize the model, potentially including artificial intelligence features to leverage on the stored data, and may investigate the intertwining of the prototype with tools to generate synthetic demand profiles.

APPENDIX COMPLETE LIST OF FIELDS

In Tables 5-14 the detailed list of fields of the entities described in Fig. 2 is detailed. However, given the proposed NoSQL database structure, it is worth noticing that additional entries can be easily added.

ACKNOWLEDGMENT

The authors would like to thank Leo Maltese and Francesco Marabotto, students of the Artificial Intelligence and Data Engineering Master Program of the University of Pisa, for

their support to the implementation of the first prototype of the data platform and Luca Belloni and Tommaso Ferrucci for the preliminary analysis on the drivers.

REFERENCES

- [1] *The Millennium Development Goals Report*, United Nations, New York, NY, USA, 2015.
- [2] *Energy Access Outlook 2017: From Poverty to Prosperity*, IEA, Paris, France, 2017.
- [3] F. Riva, H. Ahlberg, 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.
- [4] P. Casati, M. Moner-Girona, S. I. Khaleel, S. Szabo, and G. Nhamo, "Clean energy access as an enabler for social development: A multidimensional analysis for Sub-Saharan Africa," *Energy Sustain. Develop.*, vol. 72, pp. 114–126, Feb. 2023.
- [5] A. B. Bahaj, L. Blunden, C. Kanani, P. James, I. Kiva, Z. Matthews, H. Price, H. Essendi, J. Falkingham, and G. George, "The impact of an electrical mini-grid on the development of a rural community in Kenya," *Energies*, vol. 12, no. 5, pp. 1–21, 2019.
- [6] J. Falk, M. Angelmahr, W. Schade, and H. Schenk-Mathes, "Socio-economic impacts and challenges associated with the electrification of a remote area in rural Tanzania through a mini-grid system," *Energy, Ecol. Environ.*, vol. 6, no. 6, pp. 513–530, Dec. 2021.
- [7] *World Energy Outlook Special Report Africa Energy Outlook 2022*, Int. Energy Agency, Paris, France, 2022.
- [8] *Energy Access Explorer*, World Res. Inst., Washington, DC, USA, 2022.
- [9] *Global Electrification Platform*, ESMAP, WorldBank, Washington, DC, USA, 2022.
- [10] *World Energy Outlook 2021*, IEA, Paris, France. [Online]. Available: <https://www.iea.org/reports/world-energy-outlook-2021>
- [11] *Africa Energy Outlook 2019*, IEA, Paris, France, 2019.
- [12] I. D. Silva, D. Bricca, A. Micangeli, D. Fioriti, and P. Cherubini, "Triple helix as a strategic tool to fast-track climate change adaptation in rural Kenya: Case study of Marsabit County," in *African Handbook of Climate Change Adaptation*. New York, NY, USA: Springer, 2020.
- [13] M. Moner-Girona, M. Solano-Peralta, M. Lazopoulou, E. K. Ackom, X. Vallve, and S. Szabó, "Electrification of Sub-Saharan Africa through PV/hybrid mini-grids: Reducing the gap between current business models and on-site experience," *Renew. Sustain. Energy Rev.*, vol. 91, pp. 1148–1161, Aug. 2018.
- [14] S. Mandelli, J. Barbieri, R. Mereu, and E. Colombo, "Off-grid systems for rural electrification in developing countries: Definitions, classification and a comprehensive literature review," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 1621–1646, May 2016.
- [15] E. Hartvigsson and E. O. Ahlgren, "Comparison of load profiles in a mini-grid: Assessment of performance metrics using measured and interview-based data," *Energy Sustain. Develop.*, vol. 43, pp. 186–195, Apr. 2018.
- [16] P. Ciller, D. Ellman, C. Vergara, A. González-García, S. J. Lee, C. Drouin, M. Brunanhan, Y. Borofsky, C. Mateo, R. Amatya, R. Palacios, R. Stoner, F. D. Cuadra, and I. Pérez-Arriaga, "Optimal electrification planning incorporating on- and off-grid technologies: The reference electrification model (REM)," *Proc. IEEE*, vol. 107, no. 9, pp. 1872–1905, Sep. 2019.
- [17] D. Fioriti, D. Poli, P. Duenas-Martinez, and I. Perez-Arriaga, "Multi-year stochastic planning of off-grid microgrids subject to significant load growth uncertainty: Overcoming single-year methodologies," *Electr. Power Syst. Res.*, vol. 194, May 2021, Art. no. 107053.
- [18] N. Stevanato, F. Lombardi, G. Guidicini, L. Rinaldi, S. L. Balderrama, M. Pavičević, S. Quoilin, and E. Colombo, "Long-term sizing of rural microgrids: Accounting for load evolution through multi-step investment plan and stochastic optimization," *Energy Sustain. Develop.*, vol. 58, pp. 16–29, Oct. 2020.
- [19] M. Franz, N. Peterschmidt, M. Rohrer, and B. Kondev, "Mini-grid policy toolkit," EUEI-PDF, ARE, REN21, Eschborn, Germany, Tech. Rep., 2014.
- [20] N. Stevanato, F. Lombardi, E. Colombo, S. Balderrama, and S. Quoilin, "Two-stage stochastic sizing of a rural micro-grid based on stochastic load generation," in *Proc. IEEE Milan PowerTech*, Milan, Italy, Jun. 2019, pp. 1–6. Accessed: Apr. 4, 2022.
- [21] V. Gambino, R. D. Citto, P. Cherubini, C. Tacconelli, A. Micangeli, and R. Giglioli, "Methodology for the energy need assessment to effectively design and deploy mini-grids for rural electrification," *Energies*, vol. 12, no. 3, p. 574, Feb. 2019.
- [22] 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.
- [23] *National Bureau of Statistics—National Panel Survey*, Nat. Bur. Statist., Dar es Salaam, Tanzania. Accessed: Mar. 24, 2022.
- [24] S. Fraval et al., "Food access deficiencies in Sub-Saharan Africa: Prevalence and implications for agricultural interventions," *Frontiers Sustain. Food Syst.*, vol. 3, p. 104, Nov. 2019.
- [25] P. Boait, R. Gammon, V. Advani, N. Wade, D. Greenwood, and P. Davison, "ESCoBox: A set of tools for mini-grid sustainability in the developing world," *Sustainability*, vol. 9, no. 5, p. 738, May 2017.
- [26] F. Lombardi, S. Balderrama, S. Quoilin, and E. Colombo, "Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model," *Energy*, vol. 177, pp. 433–444, Jun. 2019.
- [27] *Africa Electricity Grids Explorer*, World Bank, Washington, DC, USA, 2022. Accessed: Mar. 24, 2022.
- [28] L. Lorenzoni, P. Cherubini, D. Fioriti, D. Poli, A. Micangeli, and R. Giglioli, "Classification and modeling of load profiles of isolated mini-grids in developing countries: A data-driven approach," *Energy Sustain. Develop.*, vol. 59, pp. 208–225, Dec. 2020.
- [29] M. Bhatia and N. Angelou, "Beyond connections: Energy access redefined," World Bank Group, Washington, DC, USA, Tech. Rep. 008/15, Jul. 2015. Accessed: May 31, 2016.
- [30] *Kenya Population and Housing Census*, Nat. Bur. Statist., Malindi, Kenya, 2019.
- [31] "The world bank annual report 2021: From crisis to green, resilient, and inclusive recovery (English)," World Bank Group, Washington, DC, USA, Tech. Rep. 164644. [Online]. Available: <http://documents.worldbank.org/curated/en/120541633011500775/The-World-Bank-Annual-Report-2021-From-Crisis-to-Green-Resilient-and-Inclusive-Recovery>
- [32] *Clean Cooking*, SE4All, Vienna, Austria, 2022.
- [33] *LPG and the Global Energy Transition*, World LPG Assoc., Neuilly-sur-Seine, France, 2015, p. 36.
- [34] E. Hartvigsson, E. Ahlgren, J. Ehnberg, and S. Molander, "Rural electrification through minigrids in developing countries: Initial generation capacity effect on cost-recovery," in *Proc. Int. Conf. Syst. Dyn. Soc.*, 2015, pp. 1–12. Accessed: Mar. 31, 2022.
- [35] M. Toman et al., "Africa's pulse," World Bank Group, Washington, DC, USA, Tech. Rep. 17, Apr. 2018, doi: 10.1596/978-1-4648-1291-0.
- [36] S. Pfenninger and I. Staffell, "Renewables.ninja platform," Tech. Rep., 2020. Accessed: 2022. [Online]. Available: <https://www.renewables.ninja/>
- [37] H. Louie and P. Dauenhauer, "Effects of load estimation error on small-scale off-grid photovoltaic system design, cost and reliability," *Energy Sustain. Develop.*, vol. 34, pp. 30–43, Oct. 2016.
- [38] F. Riva, A. Tognollo, F. Gardumi, and E. Colombo, "Long-term energy planning and demand forecast in remote areas of developing countries: Classification of case studies and insights from a modelling perspective," *Energy Strategy Rev.*, vol. 20, pp. 71–89, Apr. 2018.
- [39] C. Kuster, Y. Rezgui, and M. Mourshed, "Electrical load forecasting models: A critical systematic review," *Sustain. Cities Soc.*, vol. 35, pp. 257–270, Nov. 2017.
- [40] A. V. Shankar, A. K. Quinn, K. L. Dickinson, K. N. Williams, O. Masera, D. Charron, D. Jack, J. Hyman, A. Pillarisetti, R. Bailis, P. Kumar, I. Ruiz-Mercado, and J. P. Rosenthal, "Everybody stacks: Lessons from household energy case studies to inform design principles for clean energy transitions," *Energy Policy*, vol. 141, Jun. 2020, Art. no. 111468.
- [41] S. A. Mehetre, N. L. Panwar, D. Sharma, and H. Kumar, "Improved biomass cookstoves for sustainable development: A review," *Renew. Sustain. Energy Rev.*, vol. 73, pp. 672–687, Jun. 2017.
- [42] S. C. Bhattacharyya, "Energy access programmes and sustainable development: A critical review and analysis," *Energy Sustain. Develop.*, vol. 16, no. 3, pp. 260–271, Sep. 2012.
- [43] G. Falchetta, N. Stevanato, M. Moner-Girona, D. Mazzoni, E. Colombo, and M. Hafner, "The M-LED platform: Advancing electricity demand assessment for communities living in energy poverty," *Environ. Res. Lett.*, vol. 16, no. 7, Jul. 2021, Art. no. 074038.
- [44] E. G. Carayannis and D. F. J. Campbell, "'Mode 3' and 'quadruple helix': Toward a 21st century fractal innovation ecosystem," *Int. J. Technol. Manage.*, vol. 46, no. 3/4, p. 201, 2009.
- [45] K. Wiese, "Energy 4 all? Investigating gendered energy justice implications of community-based micro-hydropower cooperatives in Ethiopia," *Innov. Eur. J. Social Sci. Res.*, vol. 33, no. 2, pp. 194–217, Apr. 2020.

- [46] *Mini Grids for Half a Billion People: Market Outlook and Handbook for Decision Makers*, Energy Sector Manag. Assistance Program (ESMAP), World Bank, Washington, DC, USA, 2019.
- [47] E. Colombo, S. Bologna, and D. Masera, *Renewable Energy for Unleashing Sustainable Development*. New York, NY, USA: Springer, 2013.
- [48] V. B. Poala, "Italian involvement in rural electrification: Lessons learnt and way forward," M.S. thesis, Politecnico di Milano, Milano, Italy, 2019.
- [49] E. G. Carayannis and D. F. J. Campbell, "Triple helix, quadruple helix and quintuple helix and how do knowledge, innovation and the environment relate to each other? A proposed framework for a trans-disciplinary analysis of sustainable development and social ecology," *Int. J. Social Ecol. Sustain. Develop.*, vol. 1, no. 1, pp. 41–69, 2010.
- [50] *Leap-Re | Europe-Africa Partnership for Renewable Energy*, LEAP-RE, 2020. [Online]. Available: <http://leap-re.eu/>
- [51] M. Aresti, A. Barclay, P. Cherubini, O. Dubois, F. Eveillé, D. Fioriti, V. Gambino, C. Gertig, A. G. García, A. Micangeli, J. Shehata, S. Stritzke, C. Tacconelli, L. Traini, P. Trotter, S. van den Berg, and T. Volk, "RE-Thinking: Access to energy business models: Ways to walk the water-energy-food Nexus talk in Sub-Saharan Africa," RES4Africa, Rome, Italy, Tech. Rep., 2019.
- [52] G. Falchetta, B. Michoud, M. Hafner, and M. Rother, "Harnessing finance for a new era of decentralised electricity access: A review of private investment patterns and emerging business models," *Energy Res. Social Sci.*, vol. 90, Aug. 2022, Art. no. 102587.
- [53] *Copernicus*, Copernicus, Europe, 2022. [Online]. Available: <https://www.copernicus.eu/en>
- [54] *Open Data*, Our World Data, Oxford, U.K., Mar. 2022.
- [55] X. Li, S. Booth, S. Esterly, I. Baring-Gould, J. Clowes, P. Weston, P. Shukla, J. Thacker, and A. Jacquiau-Chamski, "Performance monitoring of African micro-grids: Good practices and operational data energy 4 impact," Nat. Renew. Energy Lab. (NREL), Golden, CO, USA, Tech. Rep., 2020. [Online]. Available: <https://www.nrel.gov/docs/fy20osti/71767.pdf>
- [56] *Africa's Minigrids*, AMDA, New York, NY, USA, 2020.
- [57] *Powering Productivity. Early Insights Into Mini Grid Operations in Rural Kenya*, Vulcan, Seattle, WA, USA, 2016.
- [58] N. J. Williams, P. Jaramillo, B. Cornell, I. Lyons-Galante, and E. Wynn, "Load characteristics of East African microgrids," in *Proc. IEEE PES PowerAfrica*, Accra, Ghana, Jun. 2017, pp. 236–241. Accessed: Mar. 31, 2022.
- [59] K. Louw, B. Conradie, M. Howells, and M. Dekenah, "Determinants of electricity demand for newly electrified low-income African households," *Energy Policy*, vol. 36, pp. 2814–2820, Aug. 2008.
- [60] A. Azadeh and Z. S. Faiz, "A meta-heuristic framework for forecasting household electricity consumption," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 614–620, Jan. 2011.
- [61] S. Dubey, E. Adovor, D. Rysankova, and B. Koo, "Kenya—Beyond connections: Energy access diagnostic report based on the multi-tier framework," World Bank Group, Washington, DC, USA, Tech. Rep. 157067, Jan. 2020. Accessed: Mar. 15, 2021.
- [62] *Nigeria—Multi-Tier Framework for Measuring Energy Access 2018*, World Bank, Washington, DC, USA. Accessed: Mar. 24, 2022.
- [63] L. Luzi, Y. Lin, B. B. Koo, D. Rysankova, and E. Portale, "Zambia—Beyond connections: Energy access diagnostic report based on the multi-tier framework," World Bank Group, Washington, DC, USA, Tech. Rep., Jun. 2019. Accessed: Mar. 12, 2019, doi: [10.1596/32750](https://doi.org/10.1596/32750).
- [64] *Multi Tier Framework | Tracking Progress Toward Sustainable Energy Goals*, ESMAP, Washington, DC, USA, 2015.
- [65] A. Hasanbasri, T. Kilic, G. Koolwal, and H. Moylan, "LSMS+ program in Sub-Saharan Africa: Findings from individual-level data collection on labor and asset ownership," World Bank Group, Washington, DC, USA, Tech. Rep. 158818, 2021.
- [66] *ISCED 2011 Operational Manual*, OECD, Eurostat, UNESCO, Inst. Statist., Paris, France, 2015.
- [67] *Interactive Webmap for Electrification Planning in Nigeria*, SE4All, Vienna, Austria, 2020.
- [68] F. Riva, "When complexity turns into local prosperity: A system dynamics approach to meeting the challenges of the rural electricity-development Nexus," *Energy Sustain. Develop.*, vol. 59, pp. 226–242, Dec. 2020.
- [69] C. Dominguez, K. Orehoung, and J. Carmeliet, "Understanding the path towards a clean energy transition and post-electrification patterns of rural households," *Energy Sustain. Develop.*, vol. 61, pp. 46–64, Apr. 2021.
- [70] S. Mandelli, M. Merlo, and E. Colombo, "Novel procedure to formulate load profiles for off-grid rural areas," *Energy Sustain. Develop.*, vol. 31, pp. 130–142, Apr. 2016.
- [71] S. Fobi, V. Deshpande, S. Ondiek, V. Modi, and J. Taneja, "A longitudinal study of electricity consumption growth in Kenya," *Energy Policy*, vol. 123, pp. 569–578, Dec. 2018.
- [72] *Scopus Search Engine*, Scopus, Hilversum, The Netherlands. Accessed: 2022. [Online]. Available: <https://www.scopus.com/>
- [73] T. Ferrucci and L. Belloni, "Construction of a harmonized database and training of machine learning models for off-grid demand estimation," M.S. thesis, Politecnico di Milano, Milan, Italy, 2022.
- [74] R. V. Jones, A. Fuertes, and K. J. Lomas, "The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings," *Renew. Sustain. Energy Rev.*, vol. 43, pp. 901–917, Mar. 2015.
- [75] F. Riva and E. Colombo, "System-dynamics modelling of the electricity-development Nexus in rural electrification based on a Tanzanian case study," *Energy Sustain. Develop.*, vol. 56, pp. 128–143, Jun. 2020.
- [76] C. Dominguez, K. Orehoung, and J. Carmeliet, "Estimating hourly lighting load profiles of rural households in East Africa applying a data-driven characterization of occupant behavior and lighting devices ownership," *Develop. Eng.*, vol. 6, Jan. 2021, Art. no. 100073.
- [77] 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.
- [78] A. Allee, N. J. Williams, A. Davis, and P. Jaramillo, "Predicting initial electricity demand in off-grid Tanzanian communities using customer survey data and machine learning models," *Energy Sustain. Develop.*, vol. 62, pp. 56–66, Jun. 2021.
- [79] A. S. Bahaj and P. A. B. James, "Electrical minigrids for development: Lessons from the field," *Proc. IEEE*, vol. 107, no. 9, pp. 1967–1980, Sep. 2019.
- [80] N. D. Rao and K. Ummel, "White goods for white people? Drivers of electric appliance growth in emerging economies," *Energy Res. Social Sci.*, vol. 27, pp. 106–116, May 2017.
- [81] N. J. Williams, P. Jaramillo, K. Campbell, B. Musanga, and I. Lyons-Galante, "Electricity consumption and load profile segmentation analysis for rural micro grid customers in Tanzania," in *Proc. IEEE PES/IAS PowerAfrica*, Cape Town, South Africa, Jun. 2018, pp. 360–365. Accessed: Mar. 31, 2022.
- [82] J. Richmond, S. Agrawal, and J. Urpelainen, "Drivers of household appliance usage: Evidence from rural India," *Energy Sustain. Develop.*, vol. 57, pp. 69–80, Aug. 2020.
- [83] M. Kurata, N. Matsui, Y. Ikemoto, and H. Tsuboi, "Do determinants of adopting solar home systems differ between households and micro-enterprises? Evidence from rural Bangladesh," *Renew. Energy*, vol. 129, pp. 309–316, Dec. 2018.
- [84] D. H. Fabiani, D. P. D. L. Baridó, A. Omu, and J. Taneja, "Mapping induced residential demand for electricity in Kenya," in *Proc. ACM DEV*. New York, NY, USA: Association for Computing Machinery, 2014, pp. 43–52.
- [85] K. Shibano and G. Mogi, "Electricity consumption forecast model using household income: Case study in Tanzania," *Energies*, vol. 13, no. 10, p. 2497, May 2020.
- [86] F. Riva, F. Gardumi, A. Tognollo, and E. Colombo, "Soft-linking energy demand and optimisation models for local long-term electricity planning: An application to rural India," *Energy*, vol. 166, pp. 32–46, Jan. 2019.
- [87] V. Daioglou, B. J. van Ruijven, and D. P. van Vuuren, "Model projections for household energy use in developing countries," *Energy*, vol. 37, no. 1, pp. 601–615, Jan. 2012.
- [88] B. J. van Ruijven, D. P. van Vuuren, B. J. M. de Vries, M. Isaac, J. P. van der Sluijs, P. L. Lucas, and P. Balachandra, "Model projections for household energy use in India," *Energy Policy*, vol. 39, no. 12, pp. 7747–7761, Dec. 2011.
- [89] E. Hartvigsson, J. Ehnberg, E. O. Ahlgren, and S. Molander, "Linking household and productive use of electricity with mini-grid dimensioning and operation," *Energy Sustain. Develop.*, vol. 60, pp. 82–89, Feb. 2021.
- [90] F. Colombelli, "Technical and socio-economic sustainability of rural electrification in Tanzania: The case of Ninga small hydropower plant project," Ph.D. thesis, Politecnico di Milano, Milano, Italy, 2019. Accessed: Mar. 24, 2022.
- [91] *Industrial Contacts of the Authors*.

- [92] M. Yener and A. Theedom, Eds., "Model view controller pattern," in *Professional Java EE Design Patterns*. Hoboken, NJ, USA: Wiley, Jul. 2015, pp. 183–193, doi: [10.1002/9781119209393.ch14](https://doi.org/10.1002/9781119209393.ch14).
- [93] L. Richardson and M. Amundsen, *Praise for RESTful Web APIs*. Sebastopol, CA, USA: O'Reilly, 2013.
- [94] M. Diogo, B. Cabral, and J. Bernardino, "Consistency models of NoSQL databases," *Future Internet*, vol. 11, no. 2, p. 43, 2019.
- [95] S. W. Ambler, *The Elements of UML 2.0 Style*. Cambridge Univ. Press, Dec. 2010, doi: [10.1017/CBO9780511817533](https://doi.org/10.1017/CBO9780511817533).



DAVIDE FIORITI (Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Pisa. In 2018, he was a Visiting Scholar with the Massachusetts Institute of Technology. He is currently an Assistant Professor and teaches energy economics and technologies with the University of Pisa. Since 2021, he has been involved in the PyPSA meets Earth initiative promoting open-source energy planning with Earth scope. He is involved as a Researcher in the EU-financed LEAP-RE project on Long-Term Joint EU-AU Research and Innovation Partnership on Renewable Energy. His research interests include mathematical optimization applied to small- and large-scale energy systems, including microgrids, rural electrification, and least-cost decarbonization strategies applied to country and continental energy systems.



NICOLÓ STEVANATO received the master's degree in energy engineering–power production and the Ph.D. degree in energy and nuclear sciences and technologies from Politecnico di Milano. He is currently a Post-Doctoral Researcher with Politecnico di Milano. During the master's degree, he spent one year with UTFSM, Chile, where he got in contact with the concept of engineering applied to sustainable development, and in particular to tackle the access to electricity issue. He is involved as a Researcher in the LEAP-RE project, the Long-Term Joint EU–AU Research and Innovation Partnership on Renewable Energy financed by the European Commission. His field of research locates in the framework of energy system modeling for policy support and long-term energy planning in developing countries, with a particular focus on the issue of access to electricity in rural areas. In particular, the research investigates the nexus between fundamental human needs and electricity demand, defined as the demand-needs nexus.



PIETRO DUCANGE received the M.Sc. degree in computer engineering and the Ph.D. degree in information engineering from the University of Pisa, Italy, in 2005 and 2009, respectively. He is currently an Associate Professor of information systems and technologies with the University of Pisa. He has been involved in a number of research and development projects in which data mining and computation intelligence algorithms have been successfully employed. He has coauthored over 100 papers in international journals and conference proceedings. His main research interests include explainable artificial intelligence, big data mining, social sensing, and sentiment analysis. He is a member of the editorial board of *Soft Computing*.



FRANCESCO MARCELLONI is currently a Full Professor of data mining and machine learning with the University of Pisa. He has coordinated various research projects funded by both public and private entities. He has also coordinated two Erasmus + KA2 projects. He has co-edited three volumes and four journal special issues. He is the (co)author of a book and more than 240 papers in international journals, books, and conference proceedings. His publications have received more than 7400 citations and he has 43 as H-index. His main research interests include explainable artificial intelligence, federated learning, data mining for big data and streaming data, sentiment analysis and opinion mining, genetic fuzzy systems, and fuzzy clustering algorithms. Recently, he has received the 2021 IEEE Transactions on Fuzzy Systems Outstanding Paper Award and the 2022 *IEEE Computational Intelligence Magazine* Outstanding Paper Award. He serves as an Associate Editor for IEEE Transactions on Fuzzy Systems, *Information Sciences* (Elsevier), and *Soft Computing* (Springer), and is on the editorial board of a number of other international journals.



EMANUELA COLOMBO received the M.Sc. degree in nuclear engineering and the Ph.D. degree in energy from Politecnico di Milano. She has been covering the role of Rector's Delegate to cooperation and development with Politecnico di Milano, since 2005. In 2011, she introduced a new track in the M.Sc. in energy engineering with a focus on energy for development. In 2012, she was named the chairholder of the UNESCO CHAIR on Energy for Sustainable Development assigned to the Department of Energy. She is currently a Full Professor of engineering for cooperation and development and thermoeconomics and energy modeling with the Department of Energy, Politecnico di Milano. She was the scientific coordinator of four European projects and one international tender on green innovation (Egypt), sustainable energy engineering (Kenya, Tanzania, and Ethiopia), water energy and food nexus (Egypt), modern energy services in refugee camps (Lebanon, Somalia, RCA, and Colombia), and capacity building in engineering (Tanzania). She has managed the Africa Innovation Leaders Program funded by the Italian Agency of Cooperation in Kenya, Ethiopia, Niger, Nigeria, Mozambique, and Tunisia. She is also involved in the scientific coordination of Pillar 2 within the LEAP-RE project, the Long-Term Joint EU–AU Research and Innovation Partnership on Renewable Energy financed by the European Commission. She is the author of more than 200 scientific articles, 120 of which are in Scopus with an H-index of 22.



DAVIDE POLI (Member, IEEE) was born in 1972. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Pisa, Italy, in 1997 and 2002, respectively. He is currently a Full Professor of power systems with the University of Pisa, where he teaches energy markets and quality and reliability of power systems. His research activities are mainly related to power system security and smart grids, as well as to production, transmission, and distribution problems in a deregulated context; more recently, he is involved in the optimal sizing and operation of hybrid mini-grids and energy communities.

...