Statistics report

# Energy end-use data collection methodologies and the emerging role of digital technologies

# Abstract

New and digital technologies have been unlocking opportunities to collect, manage and analyse large amounts of data in a relatively cost-effective way. Still, given current challenges, it is prudent that their use for energy statistics is complementary to traditional methods, until issues like data governance, confidentiality or data representativeness are more widely addressed.

This paper aims at exploring the role of new and digital technologies for energy end-use data collection. It reviews applications, strengths, and weaknesses of the major existing technologies, classifying them into three broader categories depending on their purpose: data collection, data management and data analysis.

The analysis is a starting point for energy statisticians and energy efficiency experts across countries in order to guide the design, and/or advise on the implementation of new technologies for data collection based on the case studies reviewed and on the analysis performed.

The research stems from the G20 end-use data and energy efficiency metrics initiative, co-led by the International Energy Agency and the French government through its energy efficiency agency (ADEME), building on established work in developing energy efficiency indicators to monitor energy efficiency progress globally.

# Acknowledgements, contributors and credits

This publication has been prepared by the Energy Data Centre (EDC) of the International Energy Agency (IEA). The research was performed by Ébio Vítor and co-ordinated by Mafalda Silva and Roberta Quadrelli, under the guidance of the Head of Division, Nick Johnstone.

It benefited from important contributions and discussions with other colleagues in the agency including Víctor García Tapia, Jungyu Park, Jeremy Sung, Vida Rosite, Emily McQualter, and Julian Prime.

We also acknowledge input and contribution from Eileen O' Brien (US EIA [Energy Information Administration]), and Ellen Migo (UK BEIS [Department for Business, Energy and Industrial Strategy]).

A few initial phone calls were also very useful during the research stage. We would like to thank specifically Anne Miek Kremer from CBS (Centraal Bureau voor de Statistiek / Statistics Netherlands); and Harald Haas from LiFi Research and Development Centre, University of Edinburgh.

We thank all the participants of the 2019 workshop of the G20 energy end-use data and energy efficiency metrics initiative that took place in Paris (November 2019) for their valuable contributions and stimulating discussions, which also fed this report.

Finally, we would like to acknowledge the French government through its energy efficiency agency (ADEME [Agence de l'environnement et de la maîtrise de l'énergie / Environment and Energy Management Agency]), for leading this initiative together with the IEA and for their voluntary contribution, which made the underlying research for this paper possible.

# Table of contents

## List of boxes

## List of tables

# Introduction

Data are the key to track policies effectiveness and to monitor trends over time, and energy data are no exception. In particular, disaggregated energy demand-side data collection has been a challenge in many countries worldwide, although the role of the demand-side of energy systems, notably of energy efficiency, is widely acknowledged for delivering energy savings and avoiding emissions, and hence contributing to curb climate change.

In order to appropriately track energy efficiency progress, disentangle the different drivers of energy demand (such as activity, structure, and efficiency), and develop appropriate and detailed energy efficiency indicators, it is indispensable to have sub-sectoral or end-use data together with activity data with similar boundaries.

Increasingly, governments and organisations acknowledge the importance of and are committed to developing energy efficiency indicators across sectors, depending on national priorities, and to collecting the relevant data. Traditionally, four main methodologies are widely applied for end-use data collection: administrative sources, surveys, metering and modelling. These are often used on a complementary basis. Each has its own strengths and weaknesses, which are discussed in more detail later in this paper.

In addition to traditional methodologies, new and digital technologies represent an unprecedented opportunity for energy demand-side data collection to fill some of the most challenging data gaps as of now. Overall, new technologies can be categorised into three main types depending on their main purpose: data collection, data management and data analysis. Increasing volumes of data collected in almost real-time, broad connectivity, and advanced data analytics could support end-use data collection and availability, if properly streamlined and structured.

The inherent challenges and difficulties in the use of new technologies have been widely pinpointed in the literature. Huge amounts of data require proper management (including standardising the data collected), ensuring data privacy, raising social acceptance, and allocating proper resources.

This paper is developed under the umbrella of the G20 energy end-use data and energy efficiency metrics initiative, co-led by the International Energy Agency (IEA) and France through the French Environment and Energy Management Agency (ADEME – *Agence de l'environnement et de la maîtrise de l'énergie*). It aims at reviewing traditional end-use data collection methodologies, as well as at exploring and discussing the role of digitalisation/new technologies in energy data collection, while also presenting good practices from different countries in both cases.

The bulk of the material for the section on traditional end-use collection methodologies derives from the IEA's *Energy Efficiency Indicators: Fundamentals on Statistics*, which has been expanded in the IEA Country Practices Database. The digitalisation section results of research work, alongside discussions from the 2019 workshop of the G20 end-use data and efficiency metrics initiative – Uncovering the role of digitalisation for energy efficiency indicators, and a survey conducted by the IEA on the role of digitalisation for end-use data collection.

Our review indicates that digitalisation has strong potential for supporting end-use data collection across sectors, and seems to have been already largely implemented in collecting data particularly for buildings, for example through smart meters. Experts also agree that, at least during a transition stage, new technologies should be a complement to traditional ones, in order to ensure proper quality and statistical representativeness of the data collected.

The paper is structured as follows. The first section focuses on well-established data collection methodologies in line with the IEA efficiency indicators manual. It discusses their respective advantages and drawbacks, and pinpoints examples of good practices based on input provided by countries. Section 2 introduces and describes new/digital technologies from the perspective of their potential use for data collection. It highlights advantages and drawbacks (including barriers for broader implementation) for each technology. The third section discusses and consolidates the review presented above while also showcasing the results from the survey launched by the IEA on the role of digitalisation for end-use data collection. Lastly, the conclusion section summarises the key findings from the paper.

# Methodologies for energy end-use data collection

It is indisputable that energy consumption data is crucial to support governments and stakeholders setting targets, planning and making policy decisions. From a statistical standpoint, robust energy data is important to ensure relevance, accuracy, reliability, timeliness, punctuality, accessibility, clarity, coherence, and comparability across countries (UN, 2019). The inherent costs of data collection increase with higher levels of granularity; still, the more detailed the data available, the more useful it is, giving clearer signals for consumers and policy makers, and helping to determine what should be prioritised.

The way that national statistical systems are organised varies widely across countries. Their effective functioning depends on an effective data governance model; the existence of standardised methodologies and definition of statistical standards; transparent data dissemination; and confidentiality (Masanet et al., 2017).

Some information on the final use of energy is available in the demand-side of national energy balances,[1] typically covering sectoral data: industry and manufacturing; residential; commercial and public services; transport and non-specified energy uses. Still, the development of energy efficiency indicators requires more disaggregated information than the energy balance. These indicators ideally express the relationship between end-use energy consumption and a relevant activity, within each final sector, and can be defined at different levels of detail depending on the data available (for a comprehensive framework on energy efficiency indicators, see IEA, 2014).

An energy end use refers to the final energy required to deliver a given energy service (before accounting for the efficiency of the equipment used) – for example, heating a room, driving a car, or turning on a furnace for steel making. Because energy indicators at end-use (or sub-sectoral) level capture such information, they represent the desired level to track energy efficiency progress, although this level of disaggregation implies higher data collection requirements.

Data collection strategies can be set, given the country-specific situation. As a rule of thumb, it is a good practice to identify existing data, including from across

---

[1] Examples of national energy balances developed by IEA are available at the IEA balances table.

different national institutions, before launching a new data collection programme, in order to avoid redundant expenditure. The **methodologies** to collect end-use energy consumption and activity data across sectors can be grouped into four main categories: **administrative sources, surveying, measuring/metering and modelling** – the latter often being used as a complement of the other three.

The following sections present an overview of each methodology, emphasising their advantages and drawbacks, as well as providing examples of good practices implemented in different countries.

# Administrative sources

Administrative sources can be defined as "the organisational unit responsible for implementing an administrative regulation, for which the corresponding register of units and the transactions are viewed as a source of statistical data" (OECD, 2019). Administrative sources provide official statistics, regardless of the sector.

Within the energy domain, governmental data sources comprise national-, state- and local-level institutions, that usually collect key information either on energy consumption or activity (even if for different purposes), which can be used to develop energy efficiency indicators. For instance, national agencies and statistical offices collect various macroeconomic parameters such as population and value added, which combined with energy consumption data can be used to define the corresponding energy intensities (IEA, 2014). Other examples of some useful administrative sources are those derived from energy utilities, such as billing records and taxes.

The use of administrative sources allows to effectively accessing relevant data collected from official databases, eventually avoiding the duplication of efforts and preventing needless costs. In theory, administrative data could encompass the whole population, thus eliminating the sample error many times observed in survey-related methodologies.

Additionally, there is an advantage that a wide number of records can potentially provide detailed breakdowns. However, it is important to ensure that the individual data collected by statistical agencies are strictly confidential and used exclusively for statistical purposes (UN, 2019). Still, access to such data often requires a certain level of bureaucracy. Even if the administrative data collection system is already in place, the information may not be readily available for statistical purposes, requiring a legal framework for access, either by the public or by other governmental institutions (EUROSTAT, 2013).

**Table 1.1 Summary of advantages and disadvantages of administrative sources**

| Advantages | Disadvantages |
|---|---|
| Avoid duplication by using existing data. | Dependency on third parties |
| Allow more detailed breakdowns. | Potential mismatch between definitions and statistical needs |
| Relative quickly availability. | Time investment in search and matching from existing data sources. |
| No sample error. | May require standardisation of the data collected. |
| Increased synergy between institutions. | Often implies a certain level of bureaucracy. |
| Raise profile and interest of energy efficiency among various services. | May lack transparency and legal framework for sharing data. |

Source: IEA (2014), EUROSTAT (2013).

With regard to access, some reasons why data might not be open include: i) ethical and security concerns, ii) it is time-consuming to provide processing steps to track data, and iii) simple institutional and personal inertia to co-ordinate institutional setups. Consequently, the general lack of access to energy data can lead to irreproducible results in modelling work, along with loss of traceability in decision-making (Hülk et al., 2018).

**Box 1.1 Case study on the use of administrative sources: Client registers from energy companies – Netherlands**

Statistics Netherlands compiles information from a number of sources to derive energy consumption statistics in the services sector: i) the so-called "client files" – registers of the public distribution companies for gas and electricity in the country; ii) the BAG (Basisregistratie Adressen en Gebouwen) – register of all buildings and addresses in the Netherlands; iii) the Dataland[2] – a national register that contains information on building types; iv) the National Business Register, which includes information about all enterprises in the Netherlands; v) Locatus, a national register

---

[2] https://dataland.nl/dataland-draagt-dataverzameling-en-data-ontsluiting-over-aan-het-kadaster/.

of services companies (by service activity); and vi) district heating registers that contain postal codes of district heating usages.

This practice has encountered challenges due to difficult linkages between the various data sources. The client files contain all the connections but do not distinguish, for example, household and business connections. In this case, one useful approach would be linking the client files with the BAG in order to indicate business or domestic use. However, house IDs are often registered in different forms, leading to some inconsistencies and making data linkages challenging. Identifying the building users is also difficult, requiring a match between the National Business Registers and the client files.

Despite the challenges, the client files are a good source for energy statistics, allowing linking data between various sources, and in a spatial way by using Geographical Information Systems (GIS) – refer to the section *"Geographical Information System (GIS)"* for more detail. Client files also allow for plausibility checks, and visual inspection. The outcome of this practice is that more than 98% of natural gas and electricity deliveries were allocated, thus serving as input in the Dutch Energy Balance.

# Surveys

Surveys are a cost-effective methodology to gather information on energy use patterns, as well as factors affecting energy demand. The data acquired from surveys may not necessarily provide breakdowns on energy end use, but are useful to perform well-informed estimations, or to complement other methods like modelling (using additional parameters), then allowing for further levels of disaggregation.

Production and consumption surveys are frequently used methods for energy data collection. While the former generally focus on suppliers and measure the quantities of fuels produced for/supplied to the market, the latter collect information on end-use consumption and the consumer characteristics associated with energy use (Yu et al., 2014).

For example, residential surveys are based on questionnaires collecting specific information on household occupancy patterns and activity within the building (sometimes can inquire occupants' perception of comfort). In terms of activity data, surveys can compile dwellings' characteristics – such as floor areas, type and age of dwelling, appliance diffusion within households, and household occupancy characteristics (IEA, 2014).

The main steps of designing a survey include: deciding what data are required and to what frequency; who is best placed to respond to the survey in order to avoid duplicate data and bias; costs of running and responding to surveys; statutory requirement or voluntary participation. In order to ensure representativeness of data collection, requirements such as sampling method, data validation, and data dissemination should be accounted for.

The sampling strategy (including sample size and stratification) is a key aspect of the survey, largely depending on the target population, and level of confidence pursued. UN (2008) gives some practical guidelines on sampling and more broadly on survey design. Constraints such as data protection and restricted access must also be considered in order to comply with local legislation related to the protection of collected information (EUROSTAT, 2013).

As an alternative to traditional paper-based surveys, surveys can also be conducted online, or through smart phone applications, often implying lower costs. Still, face-to-face or telephone consumer surveys are common approaches for statistical purposes. Given the extensive amount of information that can be collected, surveys are relatively cost-effective and often provide acceptable statistical significance (Table 1.2). Another important step of the surveying process is data validation, in order to provide comprehensive, accurate and good quality data.

**Table 1.2 Summary of advantages and disadvantages of surveys**

| Advantages | Disadvantages |
|---|---|
| Representativeness/statistical significance. | Time-consuming and resource intensive |
| Comprehensive and good quality data when combined with a validation process. | High respondent burden. |
| Can be used directly and as inputs for model calculations. | Potentially high absolute cost. |
| Ad hoc design of items collected based on purpose. | Requirement of staff training. |
| Relatively cost-effective, given extensive information collected. | Risk of incomplete and low-response rates, biases, sampling errors. |
|  | Need for further estimation work. |

Source: IEA (2014), EUROSTAT (2013).

On the downside, surveys can be resource-intensive and requiring qualified staff. There is a risk of incomplete responses and biases – as surveys can often be personally sensitive. The subjective nature of some assessments however implies that a degree of variability is inherent between surveyors in some of their judgements (EUROSTAT, 2013). Additionally, data gathered from surveys typically must be further treated for estimation purposes, which may be time-consuming.

### Box 1.2 Case study on the use of surveys: Energy consumption of the tertiary, agriculture and construction sectors – Germany (2011-13)

The Fraunhofer Institute for Systems and Innovation (ISI) in co-operation with other German institutions has carried out surveys in the services sector between 2011 and 2013 with the aim of improving the coverage of this sector under the official energy statistics, by collecting physical buildings characteristics, energy consumption by equipment, energy expenditure, and hence tracking energy consumption over time.

This survey was designed using a stratified sampling approach of around 2100 workplaces, and the target respondents were most knowledgeable people regarding facilities/enterprises energy consumption. The main focus was on sub-sectors, such as agriculture, construction, airports, as well as certain number of companies within the tertiary sector.

Data collected from this survey included: main building function, building floor area, building age, number of occupants, energy bills, number of lights, energy used, among others. Concerning end uses, it encompassed space cooling/heating, water heating, office equipment, lighting, auxiliary motors, process heat/cold.

The main challenge reported for this practice consisted in inconsistent responses. Besides, this survey unraveled key best practices for enhancing data collection, such as the development of a building typology based on the survey results, and a personal check and energy audit by qualified engineers in some of the assigned workplaces.

### Box 1.3 Case study on the use of surveys: Commercial Buildings Energy Consumption Survey (CBECS) – United States

The Commercial Buildings Energy Consumption Survey (CBECS) is a sample survey of national reach in the United States, which collects information on the stock of U.S. commercial buildings, including their energy-related building characteristics and energy usage data, such as consumption and expenditures. Commercial buildings are those which at least half of floor space is used for non-residential, industrial or agricultural purposes. Following this criteria, CBECS might encompass building types such as schools, hospitals, stores, restaurants, offices, among others (EIA, 2019).

The first CBECS was conducted in 1979 and, since then, has been occurring every four year; the eleventh and last survey was fielded during 2019. The main purposes of this survey are to: collect physical building characteristics, track diffusion of equipment, track energy consumption over time, understand building occupant characteristics, and collect energy expenditure. The sample design consists in an area-probability sample survey.

Elements collected include: main building function, building floor area, building age, number of occupants, energy consumption, diffusion of office equipment, and diffusion of lighting by type. The methodologies implemented to collect data comprise paper form sent by mail, computer-assisted personal interview and telephone interview.

Within the buildings sector, the end uses covered are: space cooling, space heating, water heating, office equipment, lighting and cooking. In addition, the main challenges reported from this practice refer to inconsistent responses, response qualities as well as quality of the interviewing staff.

### Box 1.4 Case study on the use of surveys: Monthly report of the current survey of energy consumption for the industry sector (chemicals, pulp and paper, iron and steel, aluminium, others) – Japan

In 2016, The Ministry of Economy, Trade and Industry (METI), has launched the survey of energy consumption in the Japanese industrial sector. The main goals of this survey are to understand the current consumption of petroleum and other types of energy by industry, as well as to obtain a groundwork for policy-making concerning petroleum consumption in the industry (this share has been progressively decreasing in time).

The sample design is based on a purposive sampling[3] in a paper version (sent by mail), comprising approximately 1600 establishments. The nine industrial categories that the survey comprises are: Pulp and paper; chemical fibres;

---

[3] Purposive sampling is a sampling design that is not intended to offer a representative sample but rather to hone in on particular phenomena and/or processes.

petroleum products; glass products; chemical industry; ceramics; clay and stone products; iron and steel; non-ferrous metals; and machinery.

There are three main elements collected from this survey:

- Fuels —receipt, generation, recovery and production, consumption, shipment, and month-end inventory.
- Electricity —electricity purchased, private power generation, consumption, and electricity sold.
- Steam —receipt, steam generation, consumption, and shipment.

This practice occurs on a monthly basis and it is mandatory – except from smaller establishments in some industries.

# Measuring and metering

Measurement or metering is not widely applied as a data collection method yet, due to relatively high costs (IEA, 2017). For this reason, samples are usually small, and thus they may lack national representativeness. Such high costs are mostly due to the costs of equipment and their maintenance, and of personnel. With the fast-paced improvements in technologies, costs are expected to gradually decrease, offering more opportunities for setting up data collection programmes via metering.

Although measuring methods may not provide large-scale samples, the data collected can serve as a complement to the other data collection approaches, especially surveys. This procedure improves also links between energy use patterns and behavioural aspects. In addition, the detailed information gathered from metering campaigns (sometimes on daily and hourly basis) present higher quality and accuracy if compared to other collection methodologies. Therefore, measurements are a key instrument to improve the existing knowledge about energy use, for example in the household sector (EUROSTAT, 2013), and in the industry sector (e.g. energy audits).

In the buildings sector, direct metering can provide useful end-use data especially for electric appliances. Particularly for decision-making, it can support demand-side management strategies and allow for smarter planning of power systems capacity (for instance, through load shifting programmes); while for households, it provides a better understanding of how energy has been consumed, which is of utter importance to raise awareness of the potential savings – either by behavioural change or technological upgrades (IEA, 2017).

Whereas metering is most commonly applied to electricity (or other energy products) consumption, it can be also used for different purposes. For instance, measuring the temperature in homes can provide great acumen of heating needs, which could be used for modelling energy demand and monitoring buildings performance (EUROSTAT, 2013).

According to Parker et al. (2015), a metering plan should address data collection, data loss (due to interruption, corruption and interception), project tracking, and data security. With regard to frequency, suitability may depend on what is measured. In some cases, a daily or hourly data collection would be convenient in order to perform processing and benchmarking routines, with data quality checks easily implemented for this time span. On the other hand, higher-resolution data would need aggregation into larger time spans, in order to provide representative statistics.

**Table 1.3 Summary of advantages and disadvantages of measuring and metering**

| Advantages | Disadvantages |
|---|---|
| Detailed information on daily and hourly consumption at end-use or equipment level. | High cost of both technical assistance and proper equipment. |
| Information on habits of the user. | Small sample of population and time (lack of representativeness). |
| High quality and accuracy of data collected. | Difficulties in finding volunteers willing to participate in the monitoring programmes. |
| Complementary to surveys and/or modelling. | Possible malfunctioning of equipment. |
| Can provide feedback to manufacturers about the real operation of equipment. | Privacy concerns regarding consumers' data collected. |

Source: IEA (2014), EUROSTAT (2013).

### Box 1.5 Case study on the use of measuring: Steel Industry Measures from the Iron and Steel Federation – Japan (2010)

A measuring initiative entitled "Nippon Keidanren Voluntary Action Programme" was carried out by the Sumitomo Metal Industries Ltd., covering iron and steel industries, in order to: i) conserve energy with more efficient steel production processes; ii) contribute to energy conservation outside the steel industry; and iii) develop new technologies to reduce $CO_2$ emissions.

This programme aims at: measuring energy use at various stages of production within a facility; measuring non-use energy used of fuels (e.g. feedstocks) within a facility; optimising fuel utilisation within a facility production process; and calculating energy efficiency of an industrial process.

The sample included all steel producers who agreed to submit data, and the measurements were taken by energy auditors. Elements collected in this practice comprise: energy use at various stage of production, heat generation/losses, combustion process efficiency, production volumes by type of product, non-energy use of feedstocks and efficiency rating of processes within production flow.

A key point of this initiative is that measurements were undertaken using the latest technology available. Besides, is highly recommended that all measurements should be made using the same methodology in order to keep data consistency.

### Box 1.6 Case study on the use of measuring: Understanding appliance utilisation patterns/residential consumption patterns (Analysis of Monitoring Campaign in Europe) – various countries (2006-08)

The Intelligent Energy – Europe (IEE), under the framework of the European Commission, launched the REMODECE initiative aiming at understanding residential energy consumption in Europe for the different types of appliances and identifying demand trends.

A simple random sampling was designed through various sources, such as lists from energy suppliers, lists of addresses/telephone numbers, and volunteers. The end uses covered from this practice comprise: space heating/cooling, water heating, refrigerators, freezers, dishwaters, washing machines, clothes dryers, televisions, computers, etc.

Among the main challenges encountered, it is worth highlighting: equipment set-up difficulties, quality control, malfunctioning equipment, insufficient number of devices, legal constraints in some countries, and monitoring longer than foreseen.

As a key recommendation, difficulties in installing on-site metering devices should be considered. Also, regular checks and calibration should be performed to ensure data collection reliability and avoid repetition of measurements. A tool was developed to analyse the metered data – appliance data were combined to survey data on a number of appliances to generate a typical load-demand curve on an hourly basis.

### Box 1.7 Case study on the use of measuring: Canadian Vehicle Use Study (CVUS) – Understanding fuel consumption trends/Reduce consumption and expenses on fuel – Canada (2014)

Transport Canada launched between 2009 and 2014 the Canadian Vehicle Use Study (CVUS), which is divided in two components: light vehicles, and medium and heavy vehicles. For light vehicles it aimed at measuring vehicle-km, passenger-km, fuel consumption, speed, fuel consumption ratio, etc. In the case of medium and heavy component the project looked at collecting cargo-tonnage, tonne-km, vehicle configuration, type of cargo, etc.

The study, conducted each quarter, combined trip data collected automatically from the electronic device without the driver intervention, and a small sequence of questions for better characterisation of the trip (driver-ID, the number of passenger by age group, the purpose of the trip and if fuel was bought).

CVUS was the first study collecting vehicle activity directly from the vehicle using electronic collection methods only. It used an electronic data logger to reduce reporting burden; to increase accuracy and the amount of data collected; and to allow speeding up the data processing.

The sample design (for distribution of the devices) resorted to vehicles registers (from national census and plate registrations) and followed a stratified sampling. The elements collected allow characterising: trip patterns, fuel consumption over time, carrier utilisation pattern and its impact on fuel economy, as well as impact of fuel switching on vehicle fuel economy.

The main challenges of this data collection initiative consisted in malfunctioning equipment, lack of understanding and proper training as well as communication difficulties with respondents. Costs may also be an important challenge in this kind of studies.

# Modelling

Given the complexity of the demand-side of energy systems, accurate energy consumption data for final consumption sectors are often difficult to obtain. Modelling stands as one important data source, providing estimates based on a range of input data and assumptions. The set of output stemming from models can yield stand-alone and readily available estimates or, as mentioned above, be used to complement/refine data collected through other methods, such as surveys and administrative sources. Modelling is not a data collection methodology in itself, but it is widely used across countries as a way to provide the best available picture of the energy use across different final consumption sectors. For this reason, it is included as the last method of this section.

In a broad sense, a model is a tool for analysing or investigating some aspects of the real world. It is usually a quantitative method, system or approach which applies statistical, economic, financial, or mathematical theories, techniques and assumptions to process input data into estimates. There are typically three parts to a model: inputs, in the form of data and assumptions; a processing component, often through calculations; and outputs (Department for Transport UK, 2014).

The main steps of modelling include: establishing a comprehensive framework, setting the assumptions required, inputting data, running the model, validating the outputs and assessing the results. The quality of the generated outputs is critically impacted by the quality of input data and the validity of the assumptions (IEA, 2014).

Historically, the two main approaches used to model end-use energy and monitor energy efficiency are the so-called “top-down” and “bottom-up”. Such nomenclature is a reference to the position in which data inputs are hierarchically placed compared to the whole system. Grubb et al. (1993) stated that the top-down approach is associated with an economic paradigm, while the bottom-up approach is associated with an engineering paradigm – therefore it is many times referred as to the engineering approach.

Top-down models typically estimate energy demand through aggregate economic indexes – such as GDP and price elasticities. Top-down models can not explicitly represent the technologies available, thus usually underestimate the potential for efficiency improvements. In contrast, bottom-up models represent technologies using disaggregated data but often disregard market thresholds, potentially overestimating efficiency improvements (Van Beeck, 2000).

With more detailed data, bottom-up models have a strong ability to model technological options, but at the same time they require a greater amount of input data, which makes them often significantly more complex (Swan and Ugursal, 2009).

The strength of top-down models lies in the fact that only aggregate data are needed, which is more available and simple to obtain.

Overall, the application of modelling approaches has key advantages, such as resource savings, reduction of respondents' burden, quick results, and a complementary usage to improve survey results (Table 1.4). Inherent risks include, for example, the use of default values from another region/country due to unavailability of local data (EUROSTAT, 2013). Attention must be drawn to the validity of modelling assumptions, to the quality of input data, and respective impact on the model outputs.

**Table 1.4 Summary of advantages and disadvantages of modelling**

| Advantages | Disadvantages |
|---|---|
| Allows quantification of variables that cannot be measured. | Estimated data with variable accuracy |
| Cost-effective. | Relies on availability and quality of input data (no stand-alone methodology). |
| Designed based on purpose. | Depends on assumptions made. |
| Quick results. | Transparency and traceability may be an issue. |
| Can be used to improve/complement other data collection methods. | |
| Allows validation of bottom-up estimates against national energy statistics. | |

Source: IEA (2014), EUROSTAT (2013).

Transparency is also invaluable for improving the modelling quality. Firstly, it encourages those producing publicly available content to undertake checks that are more rigorous. Then, it opens up approaches and results to a wide range of experts that would be able to raise insightful debate around it.

### Box 1.8 Italy: A case study on the use of modelling – Estimate energy consumption by end use/Estimate residential energy load profile/Complement data collected through surveying

A modelling practice in the residential sector was carried out by the Institute of Studies for the Integration of Systems in order to: estimate diffusion of residential appliances, estimate energy consumption by end use, estimate residential energy load profile, and complement data collected through another survey.

The model consists in a statistical and partly engineering bottom-up tool, and the assumptions and results obtained were compared and calibrated against data from Eurostat and ODYSSEE at the reference year.

Key model inputs include: space/water heating systems, lighting, household occupancy, energy prices, macroeconomic data, and technology life cycle. Model's outputs, on the other hand, consist in energy consumption from space/water heating, appliances, lighting, and total residential sector.

Main challenges were reported, such as quality control issues in the model and input data, lack of input data, choice of assumptions, and lack of good model documentation.

### Box 1.9 Case study on the use of modelling: The Cambridge Housing Model (United Kingdom)

The Cambridge Housing Model (CHM) is a domestic energy model for Great Britain and the United Kingdom. The model was used to generate estimates of energy use for the former Department of Energy and Climate Change (DECC). The primary source of input data for the CHM was the English Housing Survey (EHS) of 2011, which provided data on around 16 000 representative English dwellings (cases). The model couples EHS data to a SAP-based energy calculator, in order to estimate energy use and CO2 emissions for all homes, broken down by final use (GOV.UK, 2015).

The CHM was developed under non-domestic energy model, which allowed assessing the effect of upgrading buildings on energy use and carbon dioxide. This special attention to non-buildings is due to the fact that this type of buildings have long been neglected in energy modelling, partly because their diversity – which makes them hard to model – and partly because there is much less reliable data available compared to domestic buildings (Cambridge Energy, 2019).

# New technologies and their potential role for data collection

New technologies, in particular digital ones (such as sensors, connected devices, network equipment), are progressively emerging as potential tools for data collection, management and analysis due to their fast-paced deployment, including applications in the energy sector. In the light of digital advancements, increasing amounts of data are being generated – requiring innovative ways to treat larger data volumes but also offering unprecedented opportunities to fill current data gaps. Technologies such as smart devices, cloud computing and artificial intelligence, to name a few – have been enabling to collect and connect important information from energy systems.

Digitalisation stands for the expanding of ICT (information and communications technologies) applications across economies, including on energy systems. It can be described as the increasing interaction between the digital and physical worlds (IEA, 2017). Its ongoing and continuous development can be seen as a confluence of three major factors:

- increasing volume of data acquisition due to a decline in the costs of sensors and storage capabilities
- rapid growth in advanced analytics and disruptive computing
- greater connectivity and data transmission enabled by lower costs.

Particularly considering energy demand-side, digitalisation provides an unprecedented potential to track energy efficiency progress by collecting, managing and analysing large volumes of data, ultimately with a significant potential to inform policy making. It also has the potential to enhance the quality, timeliness and availability of energy data, which may have a significant role in the support and development of official energy statistics, particularly as end-use data collection has been historically very challenging due to its scattered nature.

There are still several challenges associated with the take-up of digital technologies more broadly. One key challenge is the urge to properly structure and deal with the large amounts of data generated. Besides, social acceptance (closely related to data privacy and security), high deployment costs as well as lack of structured regulatory frameworks may be important barriers for a broader adoption of new technologies.

For instance, in recent years, the transport sector is becoming increasingly smarter and connected, mostly for improved safety and efficiency standards. Vehicles are now able to generate geospatial data that, if properly collected, can provide huge and accurate amounts of information; hence, enabling a better analysis of travel patterns and energy efficiency of vehicles. This can overcome a hurdle for collecting accurate and detailed disaggregate travel data, which refers to the limitation of traditional methods used to follow the travellers' choices both in space and time (Allström et al., 2017).

In buildings, both residential and commercial, digitalisation developments mainly focus on heating and cooling end uses, supported by smart thermostats and sensors (IEA, 2017). These, along with smart meters and smart lighting (refer to the respective sections under *New technologies and their potential role for data collection* for more detail), have been enabling data collection in innovative ways and with high levels of granularity. Digitalisation in buildings is often applied through the so-called smart "building energy management systems (hereafter BEMS)."[4]

Also, industries have historically been taking advantage of digital technologies for the optimisation of their processes to improve safety and increase production through automation (IEA, 2017), which can also enable a data collection improvement.

Undoubtedly, ongoing digital transformation and the increasing flow of big data pave the way for a myriad of opportunities across sectors and end uses. Nevertheless, this digital progress should go hand-in-hand with data security, with the new types of system vulnerabilities to be carefully considered (UNIFE, 2019). Government, agencies, private sector, among other agents that collect very granular or individual data must comply with regulations and laws on data privacy, also with the aim of increasing public acceptance. Leakage risks and security threats need to be very well understood and cautiously managed (Singh et al., 2014).

In order to better analyse and understand the role of new/ digital technologies, these have organised into three main categories, according to their main applications: data collection, data management and data analysis. Although the boundaries between these three categories may not be always clear, as one technology may be used for more than one purpose, the sections below try to best describe them. They examine key factors and technologies opening new

---

[4] From the energy efficiency perspective, BEMS aim at enhancing energy use; for instance, this can be accomplished by improving energy services, raising awareness of costumers about their consumption patterns, reducing peak loads, and monitoring energy performance of buildings.

possibilities for each of these applications, with a focus on energy end-use level and potential for energy efficiency indicators work.

Note that, even though this section focuses on data collection from new technologies, there may be some overlap with the methodologies described in the previous section, especially administrative sources and measuring/metering.

# Data collection

Data collection refers to the process of acquiring and gathering a number of data sets, either by implementing a new system or by updating an existing one. With new technologies/digitalisation, ever-growing volumes of data have been generated from various sources. This section presents examples of new technologies used for data collection, with a focus on energy end uses.

## Global Positioning System (GPS)

Positioning technologies, such as GPS, allow capturing raw location data from a device. Positioning data has a major relevance for the transport sector, on one hand because GPS is widely used for transport navigation/ mobility support, and on the other hand because they can help track origin and destinations of specific trips, allowing to capture information such as distances travelled. This information is not only relevant as activity data for efficiency indicators (e.g. vehicle-kilometres[5] - *vkm*), but also as input data for modelling purposes that can help estimate energy use for travel. GPS is often embedded either in smartphones – a widespread gadget in the daily life of most people – and in vehicles. This turns GPS into an unprecedented opportunity for large-scale data collection, as long as privacy is guaranteed and data can be anonymised.

GPS data has also been used along with sensor data for traffic management. For example, IBM worked with the City of Dublin to identify and solve traffic congestion in Dublin's public transport network. By combining GPS information with data collected from the city's network of sensors, traffic is monitored and managed in real time. Dublin's Roads and Traffic Department collects and uses data streaming in from a variety of sources – bus timetables, inductive-loop traffic detectors, closed-circuit cameras, and GPS of buses – this is depicted into a digital map of the city with real-time positioning of buses (IBM, 2013).

Additionally, Rio was the first city in the world to collect real-time data from drivers using the Waze navigation app, and from users of the public-transportation app

---

[5] vkm = number of vehicles x average distance per vehicle (km).

Moovit. Both companies compile these data into a network of sensors that municipalities can access for predicting traffic issues, risks, and sending alerts too (Forbes, 2014). In order to access Moovit's data, municipalities download a web interface that gives them an overall view of where Moovit users are going. In exchange, the city's transport department feeds Moovit's database with real-time GPS information for buses and trains. Therefore, such type of data can be complementarily used with administrative sources and serve as input for modelling purposes.

Waze is able to capture driving patterns through phones' GPS, which makes data extremely accurate. This potentially enables the collection of relevant activity data such as vkm (vehicle-kilometre). In terms of data security, apps are sharing data anonymously, through alias, which enhance data privacy (Forbes, 2014).

### Box 2.1 Case study on the use of GPS data to manage traffic flow – Philippines

In the Philippines, the ride-share platform Grab launched the so-called OpenTraffic initiative, in partnership with the World Bank and the Department of Transportation and Communications, aiming at providing real-time data to manage traffic flows, through low-cost tools to evaluate and make sense of the data collected.

From Grab's raw data, the initiative collects the following information for each GPS point: Vehicle ID, date and time, latitude and longitude. These data are generated by more than 64 000 vehicles and updated about every 6 seconds. The project aims at developing and testing a methodology for estimating the cost of congestion – e.g. fuel consumption, GHG emissions, and economic costs (The World Bank, 2015). An important challenge is how to transform the data collected into meaningful travel information.

### Box 2.2 Mobile application for collecting data on daily trips – Slovenia

In Slovenia, a mobile application (app) was developed as a pilot within a Eurostat project "Passenger mobility statistics and road traffic statistics" with the main goal of collecting data about single-user trips. Data were collected anonymously between October and December of 2019 from 153 users, totalling 3 938 trips and 26 606 km travelled. The preliminary results provide a daily average of trips per person, travel distance per person, and travel time per person. Also, the days were categorised in: working days, non-working days and all days.

Several advantages of this pilot were reported: i) data capture, by providing accurate spatial and time data, real-time continuous measurement, and a faster data collection process; ii) improved data quality, with less measurement errors; iii) reduction of burden on respondents; and iv) cost-effectiveness, less expensive data collection on a large number of units.

Challenges included a possible sampling bias due to the choice of the respondents, and how to encourage people to participate. The installation of the app needed the authorisation of users, which also raised some questions about privacy and security and the necessity to inform about general rules and instructions. There were also reported issues related to measurement and technical capabilities of the phone: missing data, accuracy of the GPS, systematic bias and consumption of the mobile phone battery depending on accuracy.

In any case, given infrastructure limitations, there will always be situations where GPS will fail to collect data due to difficult satellite communication. Hence, this technology can be complemented by paper-based, mobile-phone based and web-based surveys (Bachu et al., 2010; Ohmori et al., 2005; Bourbonnais and Morency, 2013). In addition, it is important to bear in mind that even though positioning systems can record accurate time and geographical position, the users still need to be willing to provide and verify entities and attributes (Allström et al., 2017).

## Smart devices

Smart appliances or devices are increasingly a feasible option for households to enhance comfort, security and energy management. The Internet of Things (IoT) has led to an increasing amount of smart devices available in the market.

Those devices hold the potential of providing consumers with real-time energy consumption data and energy cost information, as well as helping utilities to understand and anticipate customer behaviour. According to Lobaccaro et al. (2016), systems and new feedback initiatives that enable visibility of energy resources to households must require a combination of well-designed programmes that successfully inform, engage and motivate users; in summary, data collection, data processing, data representation, and control/interaction capabilities.

Besides, data can be displayed to the user both through direct and indirect feedbacks. While the former represents the data collected typically in real-time, the latter are derived from post-processing tasks, for instance energy audits and enhanced billing (Lobaccaro et al., 2016).

Examples of smart devices for collecting data include smart meters, wireless sensors, smart thermostats, and smart lighting and plugs. The subsections below describe recent developments in end-use data collection (mainly in the residential sector) obtained through smart devices, also enabled by their decreasing deployment costs.

## Smart meters

Since the 2000s, many countries are increasingly adopting smart meters. These devices record and transmit information of electricity and gas consumption. Further, it is possible to store these data on a server and calculate consumption fees and statistics.

Smart meters can be useful to track the energy demand at a high level of granularity, e.g. at end-use level. Still, further disaggregation of smart meter data is needed given that smart meter readings do not provide energy use from different equipment. For example, additional approaches include: i) further estimation (e.g. from modelling), and ii) coupling it with dedicated sensors (such as plug sensors, heat flow meters, among others) that give individual end-use demand (Firth and Palmer, 2013).

Smart meters represent a real opportunity for data collection, although their implementation needs to be properly backed by a good regulatory and institutional framework to minimise or eliminate some possible risks, such as data privacy and security issues – as data are transmitted through a network, and may contain sensitive information about the user. Smart meters generate a vast amount of data, creating high traffic between devices and servers and consequently requiring extensive memory for data storage. Asghar et al. (2017) summarises issues and solutions related to data transmission, storage and processing, in association with data privacy issues in smart meters.

### Box 2.3 Smart Metering Programme – United Kingdom

The Smart Metering Implementation Programme was initiated by the UK government in 2011 and is led by the Department for Business, Energy and Industrial Strategy (BEIS), regulated by the Office of Gas and Electricity Markets (Ofgem), and delivered by energy suppliers across Great Britain. The Data and Communications Company (DCC) was granted a license to establish and manage the communications network for smart meters, as well as the data collected. In Autumn 2019, BEIS consulted on a policy framework for smart metering from January 2021 (consultation webpage available), which would apply when the current duty on energy suppliers ends

(BEIS, 2019). BEIS is carefully considering the range of responses and evidence that has been submitted and will publish the government response in due course.

This initiative has been allowing British households to have near real-time information on their energy consumption through the use of an in-home display unit (IHD)[6] supplied with the smart meter. On the other side, energy suppliers have access to accurate billing data, while network operators use consumption data to manage reinforcement activity. This initiative has been increasingly improving the volume and granularity of energy consumption data: smart meters are capable of storing at least thirteen months of half-hourly gas and electricity consumption (BEIS, 2018).

The government initially defined a standard called SMETS1[7] to ensure a minimum common functionality across smart meters from different suppliers. To enable SMETS1 meters to retain their smart functionalities in case of a switch to different energy suppliers, DCC are currently enrolling them into the national smart metering communications infrastructure, which enables full interoperability among all energy suppliers, network operators, authorised service users, and devices. The transition to installing second generation smart meters (SMETS2), which are connected to the DCC network from the point of installation, is nearly complete.

The government established the smart metering Data Access and Privacy Framework to safeguard consumers' privacy, whilst enabling proportionate access to energy consumption data (BEIS, 2018). The central principle of the Framework is that consumers have control over who can access their energy consumption data, how often and for what purposes, except where this is required for regulated purposes, such as billing. Smart Metering has been developed in close co-ordination with experts in industry and government including Government Communications Headquarters's (GCHQ) National Cyber Security Centre to ensure that robust security controls are in place with respect to data and privacy (National Cyber Security Centre, 2016).

Aiming at strengthening national statistics, some countries have been developing initiatives to couple new technologies with existing data collection programmes, with varying scopes, representativeness, and development level across countries.

---

[6] IHD allows consumers to see which appliances or end use activity is using energy and how much it is costing, including information about the amount of energy previously used in the past day, week, month and year (Department of Energy and Climate Change, 2015).

[7] SMETS stands for Smart Meter Equipment Technical Specifications.

For instance, the United States, through the Energy Information Administration (EIA) has launched a digital measurement research program in order to identify suitable technologies to collect end-use data and enhance the consolidated RECS initiative (see Box 1.13).

**Box 2.4 U.S. EIA digital measurement research program**

The Energy Information Administration (EIA) from the United States has started a residential sub metering program with the aim of testing new technologies for collection end-use data, hence identifying a technology that is sufficiently suitable. This would also be used to confirm coverage of end uses, accuracy of modelled end uses as well as to collect sufficient data to provide insight and feedback to the RECS (Residential Energy Consumption Survey) modelling process.

The proposed analysis plan from EIA consists in comparing usage patterns from sub meters results against self-reported respondent data, comparing reported behavioural characteristics against self-report respondent data, and – at the household level –, comparing actual end-use consumption from sub metered data to the 2015 RECS results.

The project – whose contract was awarded by the Pacific Northwest National Laboratory (PNNL) – started with an evaluation of disaggregation to improve RECS, followed by Phases I and II. The need of disaggregating energy consumption suggested different sources of inputs, such as CT and EMF sensors installed at the breaker panel, utility meter adapter/recorder, utility meter wireless sensor and smart meters.

The Phase I consisted mostly in the staff recruitment, homeowner agreements and subcontractor selection. Three load disaggregation products were analysed, and one demonstrated strong potential to support EIA, and therefore warranted further investigation. The recommendations from Phase I based on the pilot experience were moved into the Phase II, which included the option of exploring Advanced Metering Infrastructure (AMI[8]) for disaggregation. The second pilot was implemented under the scope of the End Use Load Research (EULR) project and installed a selected technology in nine homes. Data sharing is based on the already existing agreement with the Department of Energy (DoE).

---

[8] AMI electricity meters measure and record usage data at an hourly minimum and provide usage data to consumers and energy companies at least once daily. In 2018, the U.S. had about 87 million AMI meters, of which about 88% were residential installations (U.S. EIA, 2019).

As a conclusion, homeowner acceptance, communications reliability, cost and accuracy varies between load disaggregation products. In addition, over the short term, sub meter data collection might be potentially included on a subsample of RECS 2020. Over the long term, AMI data disaggregation Is said to may offer a scalable and cost-effective solution for understanding end-use consumption.

Another possible drawback related to smart metering devices is the inevitable data loss at some point in the reporting system. According to Park et al. (2015), typical root causes for data loss include:

- Data transfer interruption, due to loss of power to buildings/communication nodes, loss of phone connection, or radio frequency interference.
- Data corruption, translated by no data, partial data, or suspicious data, due to errors and/or communications interference.
- Data interception that, even improbable in the case of building energy-use, encourages data transfers to be encrypted to avoid security issues.

## Wireless sensor networks (WSNs)

WSNs have been advancing fast due to widening technology improvements in the fields of micro-electro-mechanical systems (MEMS) and wireless communications. These sensor nodes are capable to sense, measure and gather information – such as temperature, humidity and light intensity – from the environment, and wirelessly transmit the data to the targeted user (Yick et al., 2008).

WSNs can have diverse applications including within industry and business, military, agriculture, intelligent buildings, smart grids, and for predictive maintenance of various equipment (Ali et al., 2017). Chen et al. (2011) installed WSNs in four convenience stores in Taipei to study the reduction of air conditioning energy use. It recorded, variables such as indoor temperature, relative humidity, air velocity and illumination values.

WSNs are economically feasible to deploy and the control system can be managed centrally. However, WSNs may also have a limited memory, little computational ability, limited/ not rechargeable battery. In order to support their wider use, it is also important to ensure network security, setting a regulatory framework by statutory governing bodies, and raising public awareness (Ali et al., 2017), for example on the opportunity for innovative applications of WSNs, like reducing energy consumption (Elshrkawey et al., 2018).

## Smart thermostats

A smart thermostat is a programmable device through which a user can adjust heating or air-conditioning loads according to a given set point and/or schedule (U.S. Department of Energy, 2019). Although based on the same operational system of a conventional thermostat, a smart thermostat has a series of sensors and functionalities that allow making decisions itself in order to increase indoor comfort, and eventually saving costs for heating/cooling (Bustamante et al., 2017). In other words, smart thermostats can automatically learn from using additional data such as presence, occupancy and weather forecasts to improve energy demand management (IEA, 2017). All these data processed by the smart thermostat could be useful both for tracking/estimating thermal energy uses (e.g. space heating and cooling), but also to develop corresponding indicators in relation, for example, to occupancy information.

### Box 2.5 The Nest Learning Thermostat

The Nest Learning Thermostat, developed since 2011, is a type of thermostat that learns users' cooling and usage behaviour to help manage scheduling and energy consumption (Hernandez et al., 2014). The data it collects include: setup information provided by the user (such as location), environmental data from the Nest's sensors, heating and cooling uses, and other technical information (e.g. sensor status, Wi-Fi connectivity, and battery charge level) (Nest, 2019).

Nest is designed to control an air conditioning/heating unit based on heuristics and learning ability. A coupled Wi-Fi module connects it to the user's network and interfaces with the cloud, allowing for remote control (Hernandez et al., 2014). A study from Nest reported that energy savings ranged from 10-12% for heating and 15% for cooling (Nest, 2015).

## Smart lighting

Smart lamps are lamps embedding technology that allow for automatic control (e.g. wireless) by devices such as smartphones or remote controls. Specific functionalities of smart lamps include lighting sensor, dimming possibility, demand response readiness, presence detection, lighting scheduling, remote control, interaction with smart home hubs, among others (Serrenho and Bertoldi, 2019).

There are four different types of smart lamps: i) domestic – wireless colour tuning, dimming, integrated audio speaker, etc. ii) data delivery – connectivity enabling lamps (for example, these can track movements around a room with activation of location-specific services depending on occupancy), iii) professional features and iv) economising lamps – including sensors and controls in order to optimise use and save energy (IEA, 2016). For end-use data collection, data delivery and economising categories are the most relevant, since the former can track activity while attempting to provide improved data security, and the latter allows for energy savings and, consequently, boost energy efficiency.

In buildings, smart lighting can also help monitoring energy consumption for other end uses, such as heating and cooling. This is because it can be equipped with sensors and microprocessors, enabling to collect activity data related to occupancy and weather conditions.

Smart lighting communication uses various wireless protocols, such as user interface, user link, gateway and lamp link (IEA, 2016). Wi-Fi and Bluetooth are the most used communication technologies, due to their high rate of data transmission and easy accessibility. Recent LED technologies can be directly integrated into the building's Ethernet, turning lighting into an information tool, of which an interest example is the Light Fidelity (Li-Fi).

### Box 2.6 Light Fidelity: A promising technology for energy end-use data?

Light Fidelity (Li-Fi) is a wireless communication technology that uses infrared and visible light for high speed data communication (Haas, 2018). The technology provides: an increased data rate per square meter, very high speed, high degree of granularity, limitation of interference compared to Wi-Fi, and more security than Wi-Fi –allowing for more reliability and network capacity.

In the industrial sector, Li-Fi has been applied to operational dockyard and to the car industry by providing datasets to calibration and maintenance tasks. In the transport sector, applications include: car-to-car communications, connection of cars to infrastructures, capturing of data traffic patterns, as well as identifying traffic accidents. On the other hand, in buildings, Li-Fi can be deployed to track user occupancy, indoor movements and connection of appliances (whether certain devices are used or not), for example.

Still, from a physical standpoint, light waves can be blocked and fail to penetrate thick barriers as other waves do (Wi-Fi). Also, the user becomes dependent of lighting sources for Internet access. Due to the many different wireless protocols, the users may still have to rely on different apps to control activities; it may cause a lack of interoperability, which restricts the user to a single manufacturer in order to ensure compatibility between equipment.

## Smart plugs

Smart plugs are devices with characteristics between an ordinary plug and an appliance, with the function of turning non-smart appliances into smart ones due to their incorporated intelligent features (Serrenho and Bertoldi, 2019). Such devices can be used for efficient energy management and load monitoring in the home environment, for instance, as a viable solution to establish compatibility between smart meters and Phasor Measurement Units (PMUs[9]) (Suryadevara and Biswal, 2019).

According to Lee and Yang (2017), all the studies regarding smart plugs are based on their two major components: hardware and a management platform. In terms of data collection, consumption data and voltage are collected via hardware and sent to the management platform through communication modules, such as Wi-Fi, Bluetooth and Zigbee,[10] for later analysis.

# Web crawling

Web crawling techniques have been quickly developing in recent years. Through effective algorithms, it is possible to scrape large quantities of data from publicly available data online sources (e.g. administrative sources) in real time. In the context of energy efficiency, realms such as energy labelling and eco-design have benefited with the possibility of tracking products available in the market and comparing those (Bennich et al., 2017). Typical information such as model name, energy use, energy performance/energy label, functionalities, and purchasing price can be gathered and used for policy evaluation as a way to assess how the market is positioned regarding cost- and energy-efficient options.

### Box 2.7 The “NordCrawl” project

In this project, created under the Nordic Council of Ministers, market surveillance authorities (MSAs) from Nordic countries are developing a web-based application for MSA to monitor their national markets. MSAs have been co-operating to share the burden of data gathering, which can lower inherent costs, but increase administrative

---

[9] PMUs are high-speed sensors that measure voltage and current synchro phasors of the power system with the accuracy in the order of one microsecond (Gabbar, 2017).

[10] Zigbee is used in smart lighting, appliances and other products that involve wireless communication. The ZigBee protocol is maintained and developed by the ZigBee Alliance, which is an open, non-profit association of approximately 450 members. The characteristics for Zigbee are: mesh network protocol, low data rate, short distance operating range, and low power consumption (IEA, 2016).

constraints on the other hand. For this reason, the effectiveness of using web-scraped data for market analysis is currently being investigated.

The "NordCrawl" process of web scraping comprises the following steps: i) download the webpage; ii) extract raw data; iii) process and clean the raw data from redundant information; iv) process and organise data in a database. After all, the data can be accessed from the application, which is run by the MSAs and used for monitoring the market.

Since the project was launched in 2015, the web crawler has been scraping data from a variety of products within the Nordic market, e.g. vacuum cleaners, washing machines, combined fridge/freezer, and lighting. From the data collected – product type, model, number, technical specifications and energy performance – the following insights can be grasped: a general snapshot of the market, and its development over time, compliance with eco-design/energy labelling requirements, price development, among others.

The initial project has been continued as "NordCrawl2". Under this scope, methods to estimate market share were described, tested and evaluated. The results suggest that this method can be helpful to find out the market share of non-compliant products, which is a good input for policy-making (Mogensen et al., 2019).

The implementation of web crawling techniques in some pilots has already highlighted many advantages that this technology may offer for data collection purposes, including: reduced costs (replace purchased data), higher coverage and better representativeness of the market, data collection in real time, and improved sampling strategies. On the other hand, legal concerns have been regularly raised, such as legal possibilities and obstacles for data collection and storage from e-commerce websites; data ownership; public access to official records (Bennich et al., 2017).

# Data management

In this paper, data management refers to the methods/ tools used for processing and structuring the data (for example deleting null entries, adding/removing fields in databases), as well as data storing and retrieval, without including an in-depth analysis. It can also refer to a specific platform or software that enables these processes to take place.

Defining good quality data is a complex and broad task, regardless the sector. According to EUROSTAT (2003), data quality assurance follows six criteria:

i) accuracy, how well the data depicts the reality; ii) coherence, whether the different flows from the data collected are consistent among themselves; iii) timeliness, if the reference period for the data is the latest possible and not outdated; iv) relevance, how do data fit in the purpose; v) accessibility and clarity, how easily users can access data and understand it; and vi) comparability, the extent to which data is comparable (definitions), for example across countries.

Countries usually define their own frameworks for quality assurance and validation of the data collected, which may vary significantly.

## Database management systems

Organisations are facing new challenges in recording, updating and tracking their data on an efficient and regular basis. After collection, a proper data management becomes imperative before proceeding to a subsequent analysis. New technologies may also have a role to play in this context.

Database management systems are nowadays crucial to qualitatively store and treat data. For instance, linking end-use data with its metadata, describing the definitions of the data series, as well as the relationships and links within the database.

A useful database management system allows not only for entering new data, but also for updating existing data and correcting irregularities within a given dataset. An efficient database must be able to bring stored data into a format that would allow for further display and interpretation. These management systems can embed plausibility and quality checks to support the improvement of data quality.

### Box 2.8 SIMATIC: Energy data management in industry

An online platform solution designed by Siemens (SIMATIC) has been applied to manage energy data in different industrial sectors. An extension of this tool, the SIMATIC B.Data, offers some options for detecting and correcting irregularities within the datasets collected, such as plausibility checks and filling gaps in a series of measurement values.

A cellulose manufacturer is using this tool to deal with high volumes of data coming from different energy flows. For instance, data were available in different formats, macros were complex and difficult to understand, some data were available on multi places and historical growing system was requiring high maintenance efforts.

SIMATIC B.Data was used to acquire energy flows without gaps and track all measuring points in real-time for monthly energy accounting (Siemens, 2012).

Among others, the platform helped the manufacturer carrying out data validation checks; and ultimately reducing complexity, and increasing data quality and transparency. Besides, it enabled to provide detailed presentation of energy consumption and costs, as well as to better process understanding in terms of end-use consumption.

As an example of the importance of database management, it has been raised the need for a central database of Energy Performance Certificates (EPCs), in order to obtain representative data at national level that is good quality, comprehensive and comparable (see box below).

EPCs have been widely implemented across countries, constituting an invaluable source of data on buildings' energy performance. This makes available a large amount of data on the theoretical energy performance of buildings for key end uses (e.g. heating and cooling) – it cannot be considered proxy to energy consumption data, as the real consumption depends on additional factors such as occupants behaviour. However, it can be a useful input to derive actual consumption.

### Box 2.9 Energy Performance Certificates in Denmark: A role model for an EU central database?

The Danish authorities have been using EPCs and other public information for energy management in the buildings sector. Denmark was among the first European countries to create a central EPCs register, with around 600 000 building certificates, and covering almost 40% of the total national building stock.

Currently, European EPCs have scattered data due to the fact that regional databases present different designs, formats and access levels, with many EPC registers without public or with restricted access. In that light, an eventual design of a central European database relies on the improvement of national EPC registers, as well as the improvement of data availability, timeliness and spatial resolution.

The Danish experience could eventually support the design of a central European EPC database by enhancing the data from the Building Stock Observatory (BSO). This could use open standards and emerging technologies that provide direct data exchange, such as Application Programming Interfaces (APIs).

Databases also face some performance constraints. With increased data volumes, the speed of data processing becomes an important factor. Linked to performance, data also needs to be processed in a way that ensures that important information is not lost. Thus, ideally databases would bring together a good processing power and transactional guarantees.

## Cloud computing

Increasing volumes of data requires cost-effective storage solutions. Cloud technologies provide companies and institutions with access to large and real-time data, as well as the ability to store multiple data types for different needs and from various sources.

Cloud computing is based on the following service models (Rittinghouse and Ransome, 2017).

- Software-as-a-Service (SaaS): applications that are publicly available over the Internet.
- Platform-as-a-Service (PaaS): delivers hardware and software tools and gives customers the possibility to create their own tools.
- Infrastructure-as-a Service (IaaS): offers hardware, software, servers, IT infrastructure components over the Internet.
- Data-as-a-Service (DaaS): where data is available to customers.
- Communication-as-a-Service"allows for messaging tools.
- Monitoring-as-a-Service (MaaS): used to provide security services.

In the residential sector, technological advancements such as enhanced internet connectivity and cloud computing have significantly improved the concept of BEMS. The so called "smart grid environment" comprises multiple devices on a common platform, for instance smart meters, sensors, and communication network devices. Consequently, there is a need for proper protocol models to provide support to those devices altogether (Hashmi et al., 2011).

For example, in the case of traditional building management systems, data is extracted into a virtual cloud environment, in order to be stored and managed. With growing deployment of smart meters, power utilities face the problem of processing and storing the incoming data. Lohrmann and Kao (2011) proposed the use of Infrastructure-as-a-Service (IaaS) clouds and frameworks in order to meet the requirements to process smart meter data streams from the utility-side perspective.

In the transport sector, cloud-computing technologies have been used to track travel history from GPS data. Google's application – Google Cloud Platform (GCP) –

provides a platform to deal with connected vehicle data such as vehicle location (GPS co-ordinates), drivetrain metrics, vehicle environment status, and custom sensors. A connected vehicle combined with cloud data storage and analytics can be useful to enable predictive maintenance and freight tracking, for instance (Google Cloud, 2019).

Authentication and specific authorisation methods may help delivering better data/user privacy and security (Dakkak et al., 2018). Cloud-based services have also been allowing dealing with issues related to storing massive data, such as data losses, threat detection and cyber-physical attacks. Identity encryption, signature requirements and proxy re-encryption have also been implemented to overcome some security bottlenecks (Baek et al., 2015).

# Data analysis

Subsequent to data collection and management, the process of analysis includes checking and examining the data in order to describe facts and trends, identify patterns and understand drivers. This may also include data quality checks, statistical analysis, modelling, and interpretation of results.

For example, smart devices like smart meters often needs to couple data collected with some additional estimation, typically based on analysis with varying degrees of sophistication. Firth and Palmer (2013) suggest as an alternative approach to estimate the individual end-use energy demand using smart meter readings coupled with additional measurements from extra sensors. For instance, readings from a gas smart meter combined with measurements of indoor temperature can improve the estimative of space heating demand.

Often different types of scattered data coming from new technologies can complement each other in order to derive more robust estimations (both on end-use and activity data), that could potentially be used to feed official energy statistics.

Data analysis can be very diverse, as categories here included have very different scopes and applications. Some of the most relevant for the analysis of energy demand-side, and in particular end-use data are presented below.

## Geographical Information System (GIS)

GIS applications use different types of geo-referenced objects, associated with a geometric entity suitable for representation. These tools became a suitable instrument for managing and analysing large datasets. GIS allow for the acquisition, archiving, analysis and visualisation of various levels of geo-referenced/ spatial data (Ascione et al., 2013). Given the spatial distribution of energy users, like buildings and industries, GIS stands out as an effective approach to capture, process, store, and visualise energy-related data (given availability) as well as to track use patterns.

**Box 2.10 Case study on building and dwelling register for the production of geostatistical data – Switzerland**

A collaborative approach between the Swiss Federal Statistical Office and the Federal Office of Topography has been carried out with the aim of creating a single harmonised and open residential dataset for all partners at national level – The register of building and dwelling (RBD). At the end of 2018, the housing stock in Switzerland comprised 1.7 million residential buildings.

Data on buildings with residential use and dwellings comprise (Federal Statistical Office, 2020):

- For buildings: geo-coordinates, building category, period of construction, type of heating, energy source for heating and hot water, number of floors, number of dwellings, number of inhabitants.
- For dwellings: number of rooms, floor space, and occupancy status.

However, developing this dataset requires a common definition and ontology about buildings and addresses, as well as data sharing agreements and processes across different national institutions. This has been done through a collaborative process for acquisition and update of the data and unified processes for data exchange (including technical specifications).

GIS has become a common and effective tool for urban applications, especially in the fields of energy management and efficiency, as it combines different layers of information in a spatially explicit way. Nowacka and Remondino (2018) created a GIS streetlight platform – a comprehensive tool for efficient streetlight management and design. It includes the following data: i) geospatial information of urban areas; ii) streetlight inventory data, such as localisation of each luminaire, type of luminaire, lighting type, pole and mounting criteria; iii) consumption data per each luminaire or lamp power; and iv) photometric data for spill light and light pollution.

In addition to its ability to integrate software, hardware and data in order to capture, manage, analyse and display diverse information, GIS comprises other advantages, such as improving data consistency, systematic global coverage and regular repeat cycles. Nonetheless, although there are open source software, GIS can be often considered an expensive option; and it requires a degree of expertise from the staff using the tool.

## Online platforms/dashboards

Online platforms may serve different purposes. For example, online visualisation tools and dashboards are becoming increasingly popular as an effective way of data communication and disseminations. They may often have embedded functionalities that allow for an automatic processing/analysis of data available in order to produce more insightful graphs and other formats.

E.ON UK – the third-largest utilities' company of the United Kingdom – has built an online tool called "The Saving Energy Toolkit" to provide personalised advice and products to help customers track their energy consumption, compared to their average peer households. The application provides charts that detail energy use patterns on a monthly basis and how energy consumption is split between heating, lighting, appliances, and other uses (Bigliani, 2015).

**Box 2.11 eMARC: Monitoring and Analysis of Residential Electricity Consumption in India**

eMARC is a new initiative launched by the Indian group Prayas (Energy Group) with the aim of monitoring and analysing electricity consumption across a sample of homes in India. GPRS[11] enabled monitors are installed on the main line of residences and selected appliances in order to track electricity consumption at every minute (Prayas, 2018). The main motivation of this initiative is to overcome accuracy-related issues from consumption data collected through surveys.

The eMARC initiative has started in the city of Pune and it is expected to extend to other urban, semi-urban and rural areas in India. Two monitors are typically installed at each residence – one in the main line and the other connected to one specific appliance, such as refrigerator or air-conditioner. Both monitors measure electricity data, such as consumption (kWh), active power (kW), voltage (V), and power factor (pf) per minute, and send this information to a server via GPRS enabled monitor.

This scattered information is later aggregated and shown in an online dashboard depicting different charts showing monthly electricity consumption of all households and individual appliances. Charts can be sorted by: appliance type, number of people in a residence, and number of rooms. The dashboard also provides daily household load curves at different times throughout the day, with 15-minute block averages.

---

[11] General Packet Radio Service.

eMARC stores personal data separately from the household features and energy use data for increased security, and does not share individual data unless required for legal purposes.

**Box 2.12 The Helsinki Energy Atlas**

As part of the mySMARTLife project, an open energy atlas of Helsinki was created through the compilation of energy-related and building stock data (around 77 000 buildings). The information within the atlas has been structured in three major categories (Nowacka and Remondino, 2018):

- Basic building information: use, height, number of floors, building materials, year of construction, etc.
- Building energy and repair information: heating mode, source of energy, airflow rate, energy certificates, energy efficiency index, etc.
- Building consumption: consumption for district heating, electricity and water.

Likewise, The Energy Atlas of Berlin was developed for energy planning addressing decision makers at different levels. The toolset integrates high resolution data on energy resources and infrastructure at a range of scales in the city. It facilitates the integration and analysis of relevant energy information, such as sources, production and consumption. Key indicators for energy analysis include, for instance, the estimation of heating energy consumption at a building and neighbourhood scale (Krüger and Kolbe, 2012).

## Data mining

Data mining finds statistically reliable, previously unknown and actionable insights from data (Elkan, 2001). This has been rapidly developing over the last years and progressively applied in official statistics (Hassani et al., 2014). The application of data mining to official data can be defined as “retrieving data from different surveys or administrative sources and properly interpreting them as measures of observed phenomena” (D'Angiolini, 2002).

The application of data mining techniques in the energy context has been increasingly important to cope with large amounts of data generated from a myriad of sources. According to Ghodsi (2014), data mining has the potential to explore data to unravel unknown trends, patterns and relationships. Kim et al. (2011) noted that rising volumes of data generated in various energy-related sectors limit the applicability of traditional data analysis methods.

**Box 2.13 Wi-Fi pilot system in the London underground**

Transport for London (TfL), the local body responsible for most components of the Great London Area's public transportation network, operates the largest smart ticketing scheme in England through Oyster card, which records the location, date and time of when cards are used (POST, 2014). Such information might possibly enable the assessment of passengers' activity data, but it does not give a complete picture of travelling patterns across Londoner Tube stations. To fill this gap, TfL implemented a pilot in 2016 using a Wi-Fi system in order to have a greater understanding of how people travel beyond the gateline, by anonymously collecting passengers' data through the MAC addresses of their smartphones.

This system would allow to complement traditional surveys – that can be costly and time-consuming – to gather detailed customer information. Concerning data transparency and privacy, a number of controls were taken to make sure the pilot fully complied with the Data Protection Act 1998, by making customers sure about the purposes behind such data collection, followed through an extensive Data Protection Impact Assessment (TfL, 2017).

Singh and Yassine (2018) proposed a data-mining model (Bayesian network) to analyse behavioural energy consumption patterns in order to uncover appliance-to-appliance and appliance-to-time energy use patterns derived from energy time series of a household. This also allowed to forecasting multiple appliance usages on the short- and long-term. The results could be directly used to develop energy efficiency indicators for appliances.

Two main data mining categories can be considered: statistical models and artificial intelligence. The section below explores the latter in more detail.

## Artificial intelligence

Artificial intelligence (AI) relates to machines' ability to learn, identify and generalise patterns in large datasets, which opens a broad scope of analytical possibilities. According to the OECD (2016), AI is the ability of machines and systems to acquire and apply knowledge and to behave intelligently.

Predicting energy use and modelling the behaviour of an energy system can be very challenging; in buildings, for instance, systems are dependent on external factors, such as weather, building materials and occupants' activities, with complex multivariate inter-relationships between those factors, that are not easily captured by traditional models (Kalogirou, 2006; Zhao and Magoules, 2012).

Artificial intelligence methods, such as artificial neural networks (ANN), support vector machines and fuzzy logic have been broadly used to analyse energy consumption. Such methods are typically faster and more accurate compared to traditional methods – engineering and statistical (Khosravani et al., 2016).

Aydinalp et al. (2004) have used an ANN to model residential end-use consumption at both national and regional levels, concluding that ANNs are able to accurately model consumption for appliances, lighting, and space cooling in the residential sector. Later work used two ANN models to estimate consumption on space and domestic hot water heating in the Canadian residential sector (Aydinalp et al., 2004).

There are several applications of ANN for predicting energy use of non-residential buildings, including at end-use level (e.g. electricity for cooling, heating, lighting, and total building), by using as inputs data on weather and building design (e.g. Wong et al., 2010). Other studies, like the one from Bagnasco et al. (2015) use short-term electrical consumption data and outdoor ambient temperature to forecast through a multi-layer ANN the day-ahead load in a hospital facility. Results shown that model can also be integrated into a Building Management System (BMS), as well as be applied to forecast other types of loads, such as domestic and industrial.

For the industrial sector, Azadeh et al. (2008) successfully used an ANN for predicting annual electricity consumption for large-consumption industry sub-sectors, such as chemicals, basic metals and non-metal minerals, using historical data between 1979 and 2003.

# Implications of new technologies for national energy statistics

From the review above, it is evident that digital technologies are increasingly embedded into all sectors of the society and people's lives. The astonishing rate at which new technologies emerge and others become obsolete creates a level of uncertainty regarding those that may or may not be applied in a systematic way and endure in the years to come.

The technologies included in this paper have been around for some time, and given their increasing widespread applications are not expected to become obsolete (at least in their essence) in a mid- to long-term future.

The way that new/ digital technologies can be applied in national statistics, though, is even more difficult to foresee. As national statistics strive to be comprehensive, robust, representative and sound, the current scattered application of new technologies needs to be streamlined and integrated into the existing data collection methods.

This paper does not attempt to provide a definite answer to this complex question, but it aims at providing some guidance on what is currently being done in different corners of the world, and sharing considerations gathered from our national counterparts. The main lessons learned are as follows:

- New and digital technologies are here to stay and it is better to progressively start incorporating them into the national data collection systems than ignoring and missing the opportunity.
- As of now, and likely still in the future, they should applied in combination with traditional data collection methodologies.
- Data quantity or volume (from big/smart data) is not the only relevant aspect. The term *thick data* has also been raised, as data should also provide in-depth insights rather than a large amount of disconnected information.
- Data privacy is one of the concerns on digitalisation, which has more broadly addressed across countries. With an appropriate underlying regulatory framework, and required level of aggregation for anonymising the data, it seems that many countries have been able to overcome this challenge.

- New/digital technologies seem to have been applied widely in buildings (residential and commercial). If a starter country wants to invest on new technologies (e.g. smart meters, smart devices, GIS, etc.) for data collection leveraging on existing experiences, this would probably be the sector to start with.

The following section will present the results from the survey launched by the IEA, to help better grasp how countries have been applying new technologies.

# The IEA survey on digitalisation

Under the framework of the "G20 energy end-use data and energy efficiency metrics" (EEUDEEM) initiative, the IEA launched a survey on digitalisation across G20 countries and beyond. The survey explored the potential role of new technologies/digitalisation to collect end-use and efficiency data at national level. The results informed the third IEA EEUDEEM workshop (held in November 2019), they provided insights for this paper.

This survey had responses from 31 respondents from 21 countries. Its main findings are as follows:

- Three main data applications of emerging technologies were considered: data collection, data management and data analysis. Eighty-three percent of the respondents confirmed that new/digital technologies have been applied for data collection in their countries, while data applications for data management and analysis have been acknowledged by 70% and 37% of respondents, respectively.
- For data collection purposes, recurrent data collection has been applied in residential, services and industry sectors according to around 50% of the respondents. One-time data collection has been carried out in all sectors according to around 27% of the respondents, while pilot/testing has been carried out mostly in residential and services (33%).
- In terms of technologies, smart meters have been the most popular, widely applied across sectors (i.e. around 70% in residential, 60% in industry and 50% in services). Naturally, GPS have been dominantly applied in the transport sector (mentioned by 68% of the respondents), while cloud computing has been mostly used in services (45%). Overall, for end-use data collection, smart meters have been pointed by far as the most used technology (75% of respondents), followed by GPS (39%) and wireless sensors networks (35%).
- In addition, at a more disaggregated level, respondents have identified a larger potential for digitalisation to be more practically implemented for end-use data collection in heating/cooling (52%), appliances (42%), and manufacturing (29%).

- Lastly, the major constraints pinpointed in using digitalisation for data collection, management and analysis include security/data protection and the high associated costs/lack of financial incentives. These major bottlenecks have been identified by 61% of the respondents.

In summary, new/digital technologies have been applied with different purposes (collection, management and analysis), and varying scopes – even though the challenges seem to be common across countries. This highlights the importance of initiatives like the G20 EEUDEEM initiative, aiming at fostering sharing of good practices and international collaboration around this topic.

The last workshop was rich in interesting examples and initiatives from countries aiming at strengthening national statistics by introducing new technologies with existing data collection programmes. For example, this is the case of the US EIA digital measurement research program aiming at identifying suitable technologies to collect end-use data and enhance the consolidated RECS initiative (Box 2.4).

On the other hand, as more opportunities for data collection have been unlocked, the private sector has also been venturing in business models that offer platforms to collect, manage, analyse and display energy consumption data from final consumers. For example, the N'Gage platform, created by the French company Energisme, can be used by either businesses or individuals to collect energy data, detect anomalies, optimise facilities and supply contracts, identify savings potential and the impact of energy efficiency actions. The platform facilitates the acquisition of large and heterogeneous data, while providing the user with relevant information (like energy use per square meter) and comparing consumption between different buildings in real-time.

Similarly, Deepki has created solutions to automate data collection by aggregating data in a common format data such as energy bills, equipment lists, and site descriptions. The so-called "Deepki Collect" platform is claimed to extract the data from the documents and check the completeness and consistency of the dataset.

In order to better systematise the review above, Table 3.1 identifies the potential opportunities and barriers of each of the technologies examined throughout the paper. This aims at helping and informing stakeholders' choices for the different methods.

### Table 3.1 Key opportunities and barriers across technologies for leveraging the use of digitalisation into official energy statistics

| Technology | Potential opportunities | Potential barriers |
|---|---|---|
| **GPS** | Because GPS is widely present in smartphone devices, it unlocks unprecedented opportunities to collect data in the transport sector. Capturing positioning data allows for more accuracy over the data collected as well as more coverage, thus improving representativeness. Also, as GPS collects data at real-time, timeliness of the data collected is considerably satisfactory. | GPS might fail in the absence of necessary infrastructure, for instance, Internet access. In addition, as positioning systems record the exact position, users need to be willing to provide and secure data accuracy, tapping into societal acceptance of the technology. The huge amount of data that can be collected by GPS applications might make it difficult to select useful data and make sense of it. |
| **Smart meters** | The potential to collect consumption data at a high very level of granularity makes smart meters an invaluable option to improve accuracy of consumption data, and so increase coverage as more smart meters are installed in buildings. | Smart meters rely on a structured regulatory and institutional framework in order to tackle major concerns regarding data security and privacy. For this reason, societal acceptance may be a hurdle for the smart meters implementation, once consumers might not feel comfortable about sharing their consumption data with energy suppliers and authorities. |
| **Other smart devices** | Other technologies such as WSNs, smart lamps, smart thermostats and smart plugs present the opportunity to improve accuracy of the data collected as well as provide information at real time, hence enhancing timeliness. | Those technologies alone are not capable of collecting data and need to be coupled with smart meters or/and sensors. Besides that, their use are still not very well widespread, thus more societal awareness is necessary. Additional barriers may refer to little storage memory and limited computational ability. |
| **Web crawling** | The capabilities of this technology have the potential to scrape large quantities of sorted information at real-time, hence increasing relevance, coverage (representativeness), sampling strategies and timeliness. | On the downside, legal concerns have been regularly raised, as it still lacks a regulatory framework for web crawling. This includes legal hurdles for data collection and storage from e-commerce websites, doubts about data ownership and bottlenecks for accessing official records. |
| **Database management systems** | Databases often have the ability to manage the data collect with the aims of improving data quality, by detecting errors and standardising data in the same format, thus enhancing clarity and coherence as well as allowing for better comparability. | With increasing data volume, the speed of data processing might be negatively affected. In addition, it is imperative to ensure that data will not be lost, so making it necessary to ensure transactional guarantees. |

| Technology | Potential opportunities | Potential barriers |
|---|---|---|
| **Cloud computing** | Cloud computing provides access to large and real-time data. The ability to store multiples data types from various sources in the cloud improves data coverage and accessibility. The implementation of authentication and authorisation protocols provides an opportunity for ensuring data security and privacy. | Increasing volumes of data requires cost-effective cloud-based services, so prohibitive costs and limited storage capacity may consist in important bottlenecks. Security concerns are also critical because – apart from storing data – the solution must be able to prevent data losses and detect threats/ cyber-physical attacks. |
| **GIS** | GIS has the ability to enhance data consistency, systematic global coverage (due to geospatial information) and timeliness (regular cycles). Such technology presents invaluable opportunities for urban applications. | The large amount of data that can be collected might make it difficult to define which information is accordingly useful. Large amount data also implies in storage solutions. The technology complexity may require a high degree of expertise in order to manipulate the tool. |
| **Artificial Intelligence** | Artificial intelligence opens up wide opportunities in the field of analytics. Predicting energy use and performing modelling tasks considering external factors may increase the relevance of analyses, accuracy of models, data consistency and comparability. | Lack/ need to adapt professional skills; quality of output data derived from AI algorithms. |

In the future one possible way of addressing some of the barriers for the adoption of digital technologies from governments could be to leverage on private initiatives and to develop partnerships that can help sharing risks, and investment costs, but that can be valuable while providing users with advanced data/ information services.

# Conclusions

This paper, drafted under the G20 end-use data and efficiency metrics initiative, aimed at compiling and sharing good practices across countries on end-use data collection for the development of energy efficiency indicators across sectors. It has a special focus on digital technologies, exploring how these have been implemented so far in G20 countries and beyond, as well as trying to understand how they can potentially be used in a systematic way to help national administrations to improve energy demand-side statistics.

Traditionally, one can consider four main methods for end-use data collection: administrative sources, surveys, measuring and modelling. Each method has its own advantages and downsides, and can complement one another. For example, while many countries collect data through surveys or use administrative sources, these methods are often used together with modelling in order either to interpolate between survey years, or to derive energy end-use information from the non-energy data collected from surveys. Data accuracy and representativeness may also depend on resources invested (for example sample size).

New and digital technologies have been unlocking possibilities to collect, manage and analyse large amounts of data in a cost-effective way, compared with traditional methods. There are still a number of challenges to deploy digitalisation for end-use data collection (e.g. data security, standardisation, representativeness, upfront costs, lack of appropriate regulatory frameworks and societal acceptance). Still, as raised during the workshop, we may miss an exciting decade ahead if we let the bottlenecks stop us from taking advantage of new technologies.

However, even though new technologies are increasingly present in the society across all energy consumption sectors and may have an important role in helping to fill some of the most challenging data gaps, at least for the moment, their use for energy statistics needs to be complemented by traditional methods. Country experts supported that the best approach is to continue using a combination of both in order to collect reliable and timely data. However, how to best achieve this is not yet clear. In addition, many experts state that supporting data gathering can be often cost-prohibitive for governments, thus being necessary private sector engagement.

Although there is no single recipe to deal with existing challenges, learning from successful implementation cases from across the globe is expected to help informing and fostering the adoption of new technologies to improve energy statistics.

The applications of new technologies can fit into three main types depending on their purposes: data collection, data management and data analysis. The results from the survey conducted by the IEA show that most countries are currently applying new technologies mainly for data collection and management, while data analysis seems to have fewer applications.

From the collection perspective, smart devices and GPS, for instance, are playing a central role in improving data collection across end-use sectors. After proper collection, technologies such as cloud computing and artificial intelligence have been used to tackle the challenge of translating data into meaningful information, by providing cost-effective ways for data management and analysis, respectively. Survey results also show that, for instance, GPS has been largely applied in the transport sector, whereas smart meters are used mostly in residential and services.

Both from the survey and workshop discussions, it is evident that digital technologies have been implemented for data collection especially in buildings, for example, through the deployment of smart meters. Nonetheless, digitalisation may also have a huge potential in supporting data collection in the transport sector (especially road transport), which has been among the most challenging to collect data at a detailed level, due to its scattered nature. Positioning technologies have been playing a key role in tracking mobility data.

In any case, opportunities provided by digitalisation to improve official energy statistics can only be seized with an adequate regulatory framework in place, which establishes not only responsibilities and duties of different parties, but also necessary requirements like data access, privacy, and transparency (IEA, 2017).

Finally, it is also acknowledged that countries are at different levels when it comes to the deployment of emerging technologies for energy data collection. Therefore, there is a continued need for sharing practices among countries, and the IEA is keen to help facilitating the discussion as well as leveraging on countries' efforts for a broadly benefiting of innovative ways for collecting, managing and analysing data, and possibly extending the current Country Practices Database to include a section on digital technologies.

# References

Ali, A., Y. Ming, S. Chakraborty and S. Iram (2017), A comprehensive survey on real-time applications of WSN. *Future internet*, *9*(4), p. 77.

Allström, A., I. Kristoffersson and Y. Susilo (2017), Smartphone based travel diary collection: Experiences from a field trial in Stockholm. *Transportation research procedia*, *26*, p. 32-38.

Ascione, F. et al. (2013), Analysis and diagnosis of the energy performance of buildings and districts: Methodology, validation and development of Urban Energy Maps. *Cities*, *35*, p. 270-283.

Asghar, M. R. et al. (2017), Smart meter data privacy: A survey. *IEEE Communications Surveys & Tutorials*, *19*(4), pp. 2820-2835.

Aydinalp, M., V.I. Ugursal and A.S. Fung (2004), Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Applied Energy*, *79*(2), pp. 159-178.

Bachu, P. K., T. Dudala and S.M. Kothuri (2001), Prompted recall in global positioning system survey: Proof-of-concept study. *Transportation Research Record*, 1768(1), pp. 106-113.

Baek, J. et al. (2014), A secure cloud computing based framework for big data information management of smart grid. *IEEE transactions on cloud computing*, *3*(2), pp. 233-244.

BEIS (Department for Business, Energy and Industrial Strategy) (2019), Closed consultation: Smart meter policy framework post 2020: https://www.gov.uk/government/consultations/smart-meter-policy-framework-post-2020. Accessed on 31 March 2020.

BEIS (2018), Smart Metering Implementation Programme. Review of the Data Access and Privacy Framework: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/48804/2549-smart-meter-rollout-domestic-ia-180811.pdf. Accessed on 17 Sep. 2019.

Bennich, P. et al. (2017), Using webcrawler techniques for improved market surveillance—new possibilities for compliance and energy policy. *ECEEE Summer Study*.

Bigliani, R. (2015), E. ON UK Rebuilds Trust with Customer Engagement and Digital Transformation, *IDC Energy Insights*.

Bourbonnais, P.L. and C. Morency (2013), Web-based travel surveys, in Zmud, J., Lee-Gosselin, M., Munizaga, M., Carrasco, J.-A. (eds.), Transport survey methods: Best practice for decision making, p. 207–223. Emerald, Bingley, UK.

Bustamante, S. et al. (2017), Smart thermostats: An experimental facility to test their capabilities and savings potential. *Sustainability*, *9*(8), p. 1462.

Cambridge Energy (2019), Cambridge Housing Model (BEIS). https://cambridgeenergy.org.uk/project/cambridge-housing-model-decc/. Accessed on: 25 November 2019.

Chen, C. S. and D.S. Lee (2011), Energy saving effects of wireless sensor networks: A case study of convenience stores in Taiwan. *Sensors*, *11*(2), pp. 2013-2034.

Dakkak, O. et al. (2018), From grids to clouds: Recap on challenges and solutions. In *AIP Conference Proceedings* (Vol. 2016, No. 1, p. 020040). AIP Publishing.

D'Angiolini, G. (2002), Developing a Metadata Infrastructure for Official Data: The ISTAT Experience. In *ECML/PKDD Workshop on Mining Official Data*. Helsinki.

Department of Energy and Climate Change (2015), Smart Metering Early Learning Project: Domestic Energy Consumption Analysis. Report and Technical Annex. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/407542/2_ELP_Domestic_Energy_Consumption_Analysis_Report.pdf. Accessed on 17 Sep. 2019.

EIA (Energy Information Administration) (2019a), About the Commercial Buildings Energy Consumption Survey, https://www.eia.gov/consumption/commercial/about.php. Accessed on: 15 Nov. 2019.

EIA (2019b), How many smart meters are installed in the United States, and who has them? https://www.eia.gov/tools/faqs/faq.php?id=108&t=3. Accessed on: 07 Jan. 2020.

Elkan, C. (2001), Magical thinking in data mining: lessons from CoIL challenge 2000. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 426-431). ACM.

Elshrkawey, M., S.M. Elsherif and M.E. Wahed (2018), An enhancement approach for reducing the energy consumption in wireless sensor networks. *Journal of King Saud University-Computer and Information Sciences*, *30*(2), pp. 259-267.

EUROSTAT (2013), Manual for statistics on energy consumption in households. Publications Office of the European Union, Luxembourg.

EUROSTAT (2003), Methodological Documents – Definition of Quality in Statistics. Working Group “Assessment of quality in statistics”.:https://ec.europa.eu/eurostat/documents/64157/4373735/02-ESS-quality-definition.pdf. Accessed on: 15 Jan. 2020

Federal Statistical Office (2019), Construction and housing, https://www.bfs.admin.ch/bfs/en/home/statistics/construction-housing.html. Accessed on: 15 Jan. 2020.

Firth, S. and J. Palmer (2013), The Potential for Smart Meters in a National Household Energy Survey. *Cambridge Architectural Research Limited & Loughborough University*.

Forbes (2014), Why Google’s Waze Is Trading User Data With Local Governments.: https://www.forbes.com/sites/parmyolson/2014/07/07/why-google-waze-helps-local-governments-track-its-users/#2db6d2e639ba. Accessed on: 25 Jul. 2019.

Gabbar, H. A. (2017**)**, Smart energy grid infrastructures and interconnected micro energy grids. *In Smart Energy Grid Engineering* (pp. 23-45). Academic Press.

Ghodsi, M. (2014), A brief review of recent data mining applications in the energy industry. *International Journal of Energy and Statistics*, *2*(01), pp. 49-57.

GOV.UK (2015), Cambridge Housing Model and user guide. https://www.gov.uk/government/publications/cambridge-housing-model-and-user-guide. Accessed on: 25 Nov. 2019.

Grubb et. al. (1993), The Cost of Limiting Fossil-Fuel CO2 Emissions: A Survey and Analysis. in: Robert H. Socolow et. al. (eds.) *Annual Review of Energy and the Environment*. vol. 18, 1993. Annual Reviews, California.

Haas, H. (2018), LiFi is a paradigm-shifting 5G technology. *Reviews in Physics*, *3*, pp. 26-31.

Hashmi, M. H. S. M. K., S. Hänninen, K. Mäki (2011), Survey of smart grid concepts, architectures, and technological demonstrations worldwide. In *2011 IEEE PES conference on innovative smart grid technologies latin america (ISGT LA)* (pp. 1-7). IEEE.

Hassani, H., G. Saporta and E. S. Silva (2014), Data Mining and Official Statistics: The Past, the Present and the Future. Big Data, 2(1), pp. 1–10.

Hernandez, G. et al. (2014), Smart nest thermostat: A smart spy in your home. *Black Hat USA*, pp. 1-8.

Hülk, L. et al. (2018), Transparency, reproducibility, and quality of energy system analyses–A process to improve scientific work. *Energy strategy reviews*, *22*, pp. 264-269.

IBM (2013), Big Data Helps City of Dublin Improve its Public Bus Transportation Network and Reduce Congestion. https://www-03.ibm.com/press/uk/en/pressrelease/41097.wss. Accessed on: 23 Jul. 2019.

IEA (International Energy Agency) (2017), Digitalization & Energy.

IEA (2016), Solid State Lighting Annex: Task 7: Smart Lighting – New Features Impacting Energy Consumption**.**

IEA (2014), Energy Efficiency Indicators: Fundamentals on Statistics.

Kalogirou, S. A. (2006), Artificial neural networks in energy applications in buildings. *International Journal of Low-Carbon Technologies*, 1(3), p. 201-216.

Khosravani, H. et al. (2016), A comparison of energy consumption prediction models based on neural networks of a bioclimatic building. *Energies*, *9*(1), p. 57.

Kim, H., A. Stumpf and W. Kim (2011), Analysis of an energy efficient building design through data mining approach. *Automation in construction*, *20*(1), pp. 37-43.

Krüger, A. and T.H. Kolbe (2012), Building analysis for urban energy planning using key indicators on virtual 3D city models – The energy atlas of Berlin. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *39*(B2), pp. 145-150.

Lobaccaro, G., S. Carlucci and E. Löfström (2016), A review of systems and technologies for smart homes and smart grids. *Energies*, *9*(5), p. 348.

Lohrmann, B. and O. Kao (2011), Processing smart meter data streams in the cloud. In *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies* (pp. 1-8). IEEE.

Masanet, E. et al. (2017), Leveraging smart system technologies in national energy data systems: Challenges and opportunities.

National Cyber Security Centre (2016), The smart security behind the GB Smart Metering System.https://www.ncsc.gov.uk/information/the-smart-security-behind-the-gb-smart-metering-system. Accessed on: 17 Sep. 2019.

Nest (2019), Privacy Statement for Nest Products and Services.https://nest.com/legal/privacy-statement-for-nest-products-and-services/. Accessed on: 27 Aug. 2019.

Nest (2015), Energy Savings from the Next Learning Thermostat: Energy Bill Analysis Results.

Mogensen, K., T. Fjordbak and L. Blomqvist (2019), Nordcrawl2. Nordic Council of Ministers.

Nowacka, A. and F. Remondino (2018), Geospatial Data for Energy Efficiency and Low Carbon Cities: Overview, Experiences and New Perspectives. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

OECD (Organisation for Economic Co-operation and Development) (2019), Glossary of Statistical Terms.https://stats.oecd.org/glossary/. Accessed on: 15 Aug. 2019.

OECD (2016), OECD Science, Technology and Innovation Outlook 2016.

Ohmori, N., M. Nakazato and N. Harata (2005), GPS mobile phone-based activity diary survey. *In Proceedings of the Eastern Asia Society for Transportation Studies* (Vol. 5, pp. 1104-1115).

Parker, S. A. et al. (2015), Metering Best Practices. A Guide to Achieving Utility Resource Efficiency. *Release 3.0* (No. PNNL-23892). Pacific Northwest National Lab.(PNNL), Richland, WA (United States).

Parliamentary Office of Science and Technology (POST) (2014), Big and open data in transport. *Parliamentary Office of Science & Technology, POSTnotes Post-PN-472*.

Prayas (Energy Group), 2018. Introducing eMARC: Remote monitoring of electricity consumption patterns in Indian Homes.http://emarc.watchyourpower.org/blog/introducing-emarc-remote-monitoring-of-electricity-consumption-patterns-in-indian-homes/. Accessed on: 26 Nov. 2019.

Serrenho, T. and P. Bertoldi. (2019), Smart home and appliances: State of the art. Energy, Communications, Protocols, Standards. Europen Commission. JRC Technical Reports.

Siemens (2012), Utilize savings potential with intelligent strategies. https://c4b.gss.siemens.com/resources/images/articles/e20001-a1100-p200-x-7600.pdf. Accessed on: 24 Oct. 2019.

Singh, V., I. Srivastava and V. Johri (2014), Big data and the opportunities and challenges for government agencies. *International Journal of Computer Science and Information Technologies*, *5*(4), pp. 5821-5824.

Singh, S. and A. Yassine (2018), Big data mining of energy time series for behavioral analytics and energy consumption forecasting. *Energies*, *11*(2), 452.

Swan, L. G. and V. I. Ugursal (2009), Modeling of end use energy consumption in the residential sector: A review of modeling techniques. *Renewable and sustainable energy reviews*, 13(8), pp. 1819-1835.

Transport for London (TfL) (2017), Review of the TfL WiFi pilot. http://content.tfl.gov.uk/review-tfl-wifi-pilot.pdf. Accessed on 18 Aug. 2019.

UN (United Nations) (2019), United Nations National Quality Assurance Frameworks Manual for Official Statistics. Department of Economic and Social Affairs. Studies in Methods, Series M No.100. New York.

UN (2008), Designing household survey samples: practical guidelines (Vol. 98). United Nations Publications.

UNIFE (Union des Industries Ferroviaires Européennes / European Rail Industry) (2019), Digital Trends in the Rail Sector. http://unife.org/component/attachments/attachments.html?id=1011. Accessed on: 19 Aug. 2019.

U.S. Department of Energy (2019). Thermostats. https://www.energy.gov/energysaver/thermostats. Accessed on 27 Aug. 2019.

Wong, S. L., K. K. Wan and T.N. Lam (2010), Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, *87*(2), pp. 551-557.

Van Beeck, N. (2000), Classification of energy models. Tilburg University, Faculty of Economics and Business Administration.

Yick, J., B. Mukherjee and D. Ghosal (2008), Wireless sensor network survey. *Computer networks*, 52(12), pp. 2292-2330.

Yu, S., V. Roshchanka, M. Evans and B. Liu (2014), International Best Practices on Energy Data Management. Insights for an Indian Roadmap. U.S. Department of Energy.

The World Bank (2015), Combining Taxi GPS Data and Open-Source Software for Evidence-Based Traffic Management and Planning. http://pubdocs.worldbank.org/en/513661445369530688/Open-Traffic-Completion-Report-1.pdf. Accessed on 19 Aug. 2019.

Zhao, H.-X. and F. Magoules (2012), A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16, pp. 3586–3592.

# INTERNATIONAL ENERGY AGENCY

The IEA examines the full spectrum of energy issues including oil, gas and coal supply and demand, renewable energy technologies, electricity markets, energy efficiency, access to energy, demand side management and much more. Through its work, the IEA advocates policies that will enhance the reliability, affordability and sustainability of energy in its 30 member countries, 8 association countries and beyond.

**IEA member countries:**

Australia
Austria
Belgium
Canada
Czech Republic
Denmark
Estonia
Finland
France
Germany
Greece
Hungary
Ireland
Italy
Japan
Korea
Luxembourg
Mexico
Netherlands
New Zealand
Norway
Poland
Portugal
Slovak Republic
Spain
Sweden
Switzerland
Turkey
United Kingdom
United States

The European Commission also participates in the work of the IEA

**IEA association countries:**

Brazil
China
India
Indonesia
Morocco
Singapore
South Africa
Thailand

International Energy Agency
Website: www.iea.org

IEA Publications
International Energy Agency
Website: www.iea.org
Contact information: www.iea.org/about/contact

Typeset in France by IEA - October 2020
Cover design: IEA

Statistics report