Some Aspects of a Framework for Energy Data

Neil Brown, Rick Greenough, Konstantin Vikhorev

Institute of Energy and Sustainable Development, DeMontfort University, Leicester, UK.

Abstract

Little doubt exists that technologies for precisely, and automatically measuring energy use are timely. Pressure to reduce greenhouse gas emissions, the potential for a generation shortfall, and rising energy prices, modern building complexities, mean that estimating building energy use patterns, profiling of building energy use, and energy failure mode identification can help to maintain energy efficiency. However, further uses exist beyond operational management; These include urgently required meta-analysis of building stock by sector, up to national stock levels, to inform policymakers, since a dearth of national stock data exists at government level in many countries. Many existing systems use radio telemetry, often producing unclean data. Analysis of energy data for large datasets becomes expensive due to incompatible formats, hampering use of old data on new systems, and data from different systems should we acquire new data, or as we acquire physical buildings. We argue that a standard should include basic specifications for fundamentals such as date formats, but a secondary scalable layer will allow future-proofing of datasets for longitudinal study, and open the door to advanced analysis techniques such as complex event processing. Disaggregation to plant level, as well as building related activities, such as manufacturing activity, also becomes possible with a scalable data structure. This paper, proposes a framework for an energy data standard from a data analysis perspective built around four areas: Temporal, accuracy & precision, operational and energy documentation.

1. Introduction

In the UK, 44% of total national energy use is consumed as part of activities within buildings as reported by UK Government [1]. Advanced metering gathers building energy data remotely without requiring site visits (typically half-hourly) [2], and is increasing in use due to market pressures [3]. It is mainly currently used for billing purposes, although offers very useful datasets to study energy use [4-5]. It offers massively streamlined and enhanced opportunities for identifying energy savings [6]: It has been found that building control faults can be identified [5, 7] causing building services to run unnecessarily (e.g. over-cooling, unoccupied heating), in some cases up to 30% of an estate showing unoccupied (night or weekend) heating [4-5]. Electrical baseloads can be identified, and reduced by up to 20% [8]. Meta analyses are also possible to gauge the energy efficiency of entire commercial sectors [9]. Analysis of these data offers a potentially massive benefit to building users, designers, utilities and policymakers in understanding building energy wastage [5-6, 10-13]. It becomes clear that the uses of high frequency energy data are numerous, yet a common problem has been one of compatibility between datasets, data quality, and completeness as identified in VDI 4602 [14].

Energy management standards are becoming well documented, showing procedures and performance metrics [14-15], and energy efficiency prognosis [16]. Data interchange standards ensure reliability of telemetry systems in diverse applications, However, converting between formats is time consuming, and there is a major reliability issue, due to inconsistencies in documentation, and data quality [5]: When analyzing meta datasets, inconsistencies become clear in even methods of notation for such fields as building addresses [4], or metering point numbers, such that results (and their input to government policies) may be skewed by e.g. data duplication and double counting. Also, critical data may frequently be ignored, because inadequate documentation formatting means that a database server is not capable of processing it. Data quality issues arise since radio telemetry systems are not 100% reliable, and data may contain gaps, or artifacts from transmission inconsistencies [4]. To make matters worse, it's fairly standard practice to interpolate between missing data points to continue with analysis, but no agreed method exists for indicating that modified data is used in profile analysis. As new techniques emerge for energy data analysis [11], a consistent data format and inbuilt functionality for checking data quality lends itself to automated processing [6, 11-12]. Compatibility, over time becomes an issue such as when a supplier of advanced metering hardware and software changes their own data structure. A basic framework for an energy data interchange standard, albeit at a higher level of abstraction, but broadly functionally similar to a standard such as BS EN 61850-7-420:2009 [17], is needed to ensure sideways, and forwards and backwards compatibility.

2. Existing Data Standards

Clearly the issues surrounding an energy data standard at first glance appear incredibly complex, but they fall into four main areas: Temporal (e.g. time and date issues, handling time driven data errors), Accuracy and Precision (including sensing devices and energy data quality), Data documentation (e.g. multipliers for kWh, building zones), and operational documentation (e.g. known down-time, data transmission issues). A literature search has revealed existing approaches which address at least in part some or most of these areas:

The International Standard for spectral data exchange [18] gives a good example of a data structure . specifically aimed at complex data transmission. For example, timestamp formats are specified as DD-MMM-YYYY, HH:MM (exactly the kind of detail, not currently made standard in many energy datasets). We must remember that timestamps are likely to require resolution down to seconds, (e.g. advanced building controls, PV monitoring, energy data rates used for manufacturing), and we must ensure, or at least be sensitive to compatibility with week numbers. The International standard 8601 [19] allows us a portable method of representing dates and times including these, as well as time intervals. (It could be argued that when dealing with, for example half hourly data, the sensible approach taken in manufacturing of representation by week numbers circumvents the perennial problem of processing date of the calendar months. Many of which have different lengths). Crucial at this phase is to specify the accuracy and precision of timestamps, and latency as in EN 61970-407:2007 on time series data access [20]. Following on from timestamps, there may be periods of non-transmission and we need a standard method of reporting this. This would require similar error codes to those described in BS EN 60255-24:2001 [21].

Low-level data transmission as described effectively in BS 15231:2006 on data communication in building automation [22], shows how reporting of transducer types (such as fiscal meters or current transformers), and equipment status is carried out using appropriate keywords. The standard contains appropriate codes for phenomena such as power, power factor, current and voltage, and fiscal meters may be treated as a credible reference since they offer stability, precision and accuracy suitable for billing [23] indeed, it is possible to specify measuring procedure down to appliance level as in ISO 12174:2003 [24].

It becomes clear that an extensible data structure offers clear advantages, notably for disaggregation down to plant level. The draft BS EN 62714-1 Engineering data exchange format for use in industrial automation systems engineering [25] offers a way into appropriate descriptors, albeit to describe hierarchies of manufacturing elements as part of a production line. Just as the Celenec report CLC/TR 50403 [26] represents supply-side disaggregation, it seems clear that an extensible way of representing hierarchies is ideal for representing disaggregated energy data. An extensible data structure with event reporting capability is described in three standards [20, 27-28], which while aimed at supply-side electricity distribution, comes very close to an appropriate system for energy event reporting, hierarchy descriptors within buildings, etc. Taxonomies also exist for room and building use types, enabling profile comparison, or to examine more precisely, energy use by sector, and an appropriate link should be documented. It may be also necessary to represent building physics, and appropriate date for performance evaluation (such as degree day calculation, climatic data [29]). Clearly a useful set of standards already exists (often with some overlap), but this is currently in a fragmented state when looking at data analysis for energy and buildings. A new standard structure can build on this useful work.

3. Energy Data Entities

A possible cause of slow development of a data standard is that solutions have largely been developed in-house by software companies, with little perceived need for portability. Also the link has not been made, or disparities identified between the [level of] data complexity required by an energy manager and software provider, meaning underlying data structure is almost certainly not seen as important when purchasing energy data software. Clearly a dataset should address data portability between core energy data tables. It would arguably be a missed opportunity, to ignore routine peripheral data, such as building characteristics, plant, machinery, building use, and occupancy.

These wouldn't necessarily need to be described in as much detail, but a standard could and should make recommendations for inclusion of the types of documentation needed for analysis.

Figure 1 shows the entity relationships for a typical energy dataset. On first glance, these data may be complex to envisage, although when logically presented, we see that the data structure falls into four sections: temporal data, accuracy and precision of data, supporting documentation, and energy data documentation.

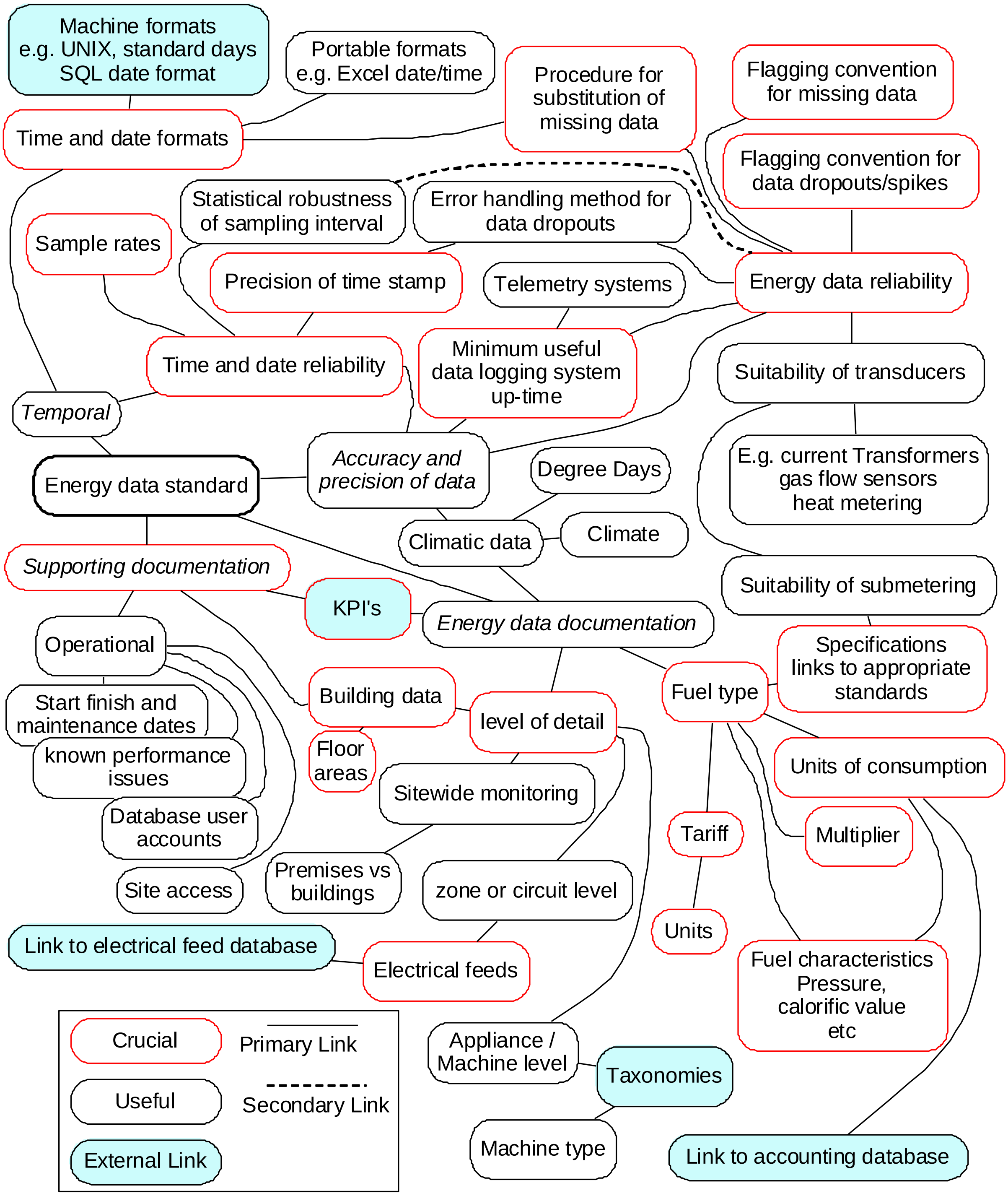

Figure 1. Entity relationship diagram for main components of an energy dataset

4. The four key aspects of a standard dataset in detail

The entities from Figure 1 may be distilled thus:

1. Temporal. Date formats, time stamping and sample rates, are important parts of energy data, which are often error prone, so some data may not be analysed. Often it is usual to substitute data from similar periods, or interpolation is used with some loss of detail [6]. A minimum level of acceptable data quality could be stipulated for time stamping in both reliability and precision, but also procedures must be put into place for handling of data errors, and logging such events. Basic telemetry systems reliability should ideally be stipulated, whereby should data dropouts cross a threshold, cleaning is abandoned in favor of corrective action on hardware or software.

2. Accuracy and Precision of Data. One shortcoming in conventional energy datasets, is that accuracy and precision of time stamping and energy data is rarely quantified. One upshot of this is that half hourly data may not be recorded on the half-hour, hampering analysis. The basic nature of (low power) electronics in conventional energy data loggers has been in the past such that errors can creep in to timestamps.

Transducer accuracy and precision is also important. For example, current transformers (CTs) for electricity monitoring, are unreliable for certain low currents, depending on CT peak load (in other words, a 400A CT may be unreliable at less than 10 amps). Clearly transducer specifications should be recorded, or at least useful operating range.

Climatic data is widely used calculation of predicted loads. The location of outside air temperature (and other weather) probes should be documented, also degree day algorithms.

In radio-based [including low-power radio and 3G based telemetry] is that where data transmission dropouts have occurred, local metering will continue to log integrated consumption until the channel for data transmission reopens. This produces in data, apparent zeros for energy consumption, followed often by a very large accumulated spike when transmission resumes. This hampers analysis considerably, but could be circumvented with a flagging convention, to allow skipping over faulty data, or to trigger data cleaning.

3. Operational Documentation – is required such as naming conventions from buildings, energy feed descriptors, and database specific operational data, as well as performance issues such as downtime for radio telemetry apparatus. Documentation should also be stored in one place, such as user accounts, permissions and privileges for database access, details of access to metered buildings or sites and scheduled maintenance of data gathering hardware and software.

4. Energy Data Documentation - Energy data documentation should be stored within the dataset, notably energy types (fuel types). (e.g. gas caloric values), and conversion factors to kWh. Any data structure must be open to the inclusion of tariff information, which also raises the possibility of data suitable for variable rate tariffs and load leveling.

Suggested data types for inclusion within an energy data structure

Sample datatypes are described below and expanded upon in Table 1 which have been used very effectively in the analysis of half-hourly data by the authors [4]. Since this paper is mainly from an energy data processing and software engineering perspective, these are offered as a point for debate, rather than the last word on a standard datasets contents. However, it is hoped that this gives a general indication of data types required for cross compatibility.

FT – Free Text. The free text data type, to describe for example, the types of telemetry systems used, or operational data such as database usernames. A very comprehensive standard could even dictate database table names to be used, and a naming convention for fields.

FN – Fixed number. A specification. e.g. 'there should be no more than 3 missing data points in 168 hours data'. A fixed number, would be a specification - another example being: time stamping should be within X percent of an agreed value for accuracy and precision.

UV – User Value. User determined value (e.g. sample rate), which should be documented in the dataset. Another example would be calorific value of gas [30].

MP – Method or Procedure. (e.g. for flagging missing data, a recommendation is made that in a separate field, codes are used to distinguish between failure modes, and suggested codes are given), as well as certain practices for handling energy data handling. A perennial issue is data dropouts (caused by a gap in radio transmission), and a method or procedure would be stipulated for handling these. This may also for example suggest a method for interpolation or substitution. Flags should be set to indicate estimated data.

FM – Format. (e.g. for a date format, which standard format(s) to present e.g. mm/dd/yyyy hh:mm:ss, pivot year etc.) A data standard would be well within its rights to suggest formats for certain kinds of data, one perennial issue is the formatting of building information, such as date formats and site data. It can be argued that considerable time and money is spent unnecessarily in writing code merely to convert between formats for energy data, and date format processing forms a major part of this.

OD – Other Documentation, (e.g. scanned floor plans, wiring diagrams, factory layout, office electrical feeds.) Such documentation for data analysis should be included, or at least stored accessibly. Thiese would include for example, building floor plans or machine layout diagrams.

LK – Reference or link to other standards or documentation. Finally, references or links to other standards would be included in any comprehensive energy dataset, such as for example, a taxonomy of machine classes or types for analysis of energy consumption at individual machine level.

The following table is not a data dictionary, or an exhaustive list of data types which would appear in a standardised energy *database,* since data design is only represented at sub-context level. What it does represent is a workable grouping of core energy data types for a basic useful dataset. Links to other standards or references are crucial where cross-comparison of plant or machinery performance is required, (e.g. Key Performance Indicators) for comparisons within sectors usually considered beyond the scope of building energy analysis (such as efficiency of server farms, technical building services for factories, and ultimately appliance/subcircuit/machine use).

Area	Primary Subgroups	Secondary Subgroups	Subject for Standard and data type	Relates to
1.Temporal	1.1 Time and Date Reliability 1.2 Time and Date Format	----------------- -----------------	1.1.0.1 Sample rate UV 1.1.0.2 Precision of Timestamp FN 1.1.0.3 Substitution of missing time data MP 1.2.0.4 Machine formats FM 1.2.0.5 Portable formats FM 1.1.0.4 Statistical Robustness of Sampling Interval FN	2.1.1.3 Substitution of missing energy data
2. Accuracy and Precision of Data	2.1 Minimum useful system up time	----------------- 2.1.1 Energy Data Reliability	2.1.0.1 Telemetry Systems MTBF, design life. FT, FN 2.1.1.1 Flagging Convention for Data Spikes MP, FM 2.1.1.2 Flagging Convention for Missing Data MP, FM 2.1.1.3 Substitution of missing energy data MP 2.1.1.4 Suitability of Transducers MP,FX 2.1.1.5 Suitability of Sub metering MP, FX	1.1.0.3 Substitution of missing time data
3. Operational Documentation		-----------------	3.0.0.1 Storage of start, finish, maintenance dates. MP, FM 3.0.0.2 Known performance issues. FT 3.0.0.3 database user accounts. FT 3.0.0.4 site access. FT	
4. Energy Data Documentation	4.1 Level of Detail	4.1.0 building data 4.1.1 Site wide	4.1.0.1 Building physics, UV, LK 4.1.0.2 floor areas, UV 4.1.1.1 premises vs. buildings, FT, OD	

	 4.2 Fuel Type 4.3 KPI's 4.4 , climatic data	monitoring 4.1.2 Zone or circuit level 4.1.3. Appliance or machine level ----------------- ----------------- 4.4.1 degree days 4.4.2 other climatic data	4.1.2.1 electrical feed database , OD 4.1.3.1 taxonomies , FT 4.2.0.1 fuel characteristics, FM, MP, FT, LK 4.2.0.2 Tariff and tariff Units 4.2.0.3 Multiplier, UV 4.3.0.1 Key performance indicators, LK 4.4.1.1 Degree day data, FM 4.4.1.2 degree day data standards, LK 4.4.1.3 degree day calculation methods, LK 4.4.1.4 outside air temperatures, FM 4.4.2.1 links to other data and standards for e.g. wind speed and direction, precipitation, humidity, etc. LK	

Table 1. Energy datatypes

Discussion

A basic data structure is proposed which offers the software engineer, the energy manager, and ultimately a standards body a 'way in' to constructing compatible energy datasets. Clearly an extensible structure is required to add extra data where appropriate. The next step must be cooperation between all interested parties to achieve consensus on energy data for portability, functionality, and quality. While this paper looks at energy data from an operational and software engineering perspective, other analysts in the field may notice gaps or improvements in functionality:

These may include modelers of non-domestic stock energy, who may use energy data for compiling statistics on energy use by sector, whereby it is emergent that bottom-up modeling provides a solution to analysis of disparate stock, not least because of small sample numbers when grouping building use by type, where disaggregation by zone use within buildings enables more effective cross-sector comparison. An example is the analysis of UK manufacturing data which are based on small sample numbers when looking at vertical sectors [9].

Many analysis techniques commonplace in manufacturing are finding application in energy analysis [6], and additional data may be required to describe, for example, production schedules, and machine types. As exergy analysis becomes more popular too, more data will be required to describe fuel types, and material throughput, from office supplies to manufactured items, all of which affect energy use in commercial buildings, not least by effects on technical building services.

As analysis techniques mature, we must be mindful that techniques such as AI, spectral analysis and complex event processing will place more exacting requirements on data quality and documentation. Finally and topically, manufacturers who increasingly see energy as a manufacturing process variable, will be looking to improve compatibility between manufacturing data systems and energy data systems.

Conclusions

The idea that four main areas should be of importance when designing energy datasets has been stated, these being temporal, accuracy and precision of data, operational and energy supporting documentation. We have described some of the intricacies of a core dataset for building energy analysis, and some aspects of data which must be documented for enhanced functionality, not least to describe in more detail energy use patterns caused by activities within buildings. A data design model has been presented as an example for a portable energy dataset.

The benefits for carbon reduction from the currently fairly basic analysis of advanced meter data are considerable. As analysis becomes more precise and datasets expand, analysis will become more problematic, yet still offer considerable insights into energy saving opportunities. Issues of compatibility, reliability and accuracy within datasets if addressed would mean that potentially useful data need not be abandoned. As advanced meter data becomes more widely available, it seems clear that cross compatibility between datasets will be highly beneficial to analysts, energy managers, software vendors, and ultimately utilities and policymakers.

References

1. DEFRA, *e-Digest Statistics about: Climate Change*. 2006.
2. Lunzer, H., *Intelligent metering.* 2007 5th Ieee International Conference on Industrial Informatics, Vols 1-3, 2007: p. 1215-1219
3. Stephenson, P. and M. Paun, *Consumer advantages from half-hourly metering and load profiles in the UK competitive electricity market.* Drpt2000: International Conference on Electric Utility Deregulation and Restructuring and Power Technologies, Proceedings, 2000: p. 35-40
4. Brown, N., Wright, A.J, *DELIVERABLE 15. Report on Half-Hourly Metering Profiles* 2008, Institutute of Energy and Sustainable Development, DeMontfort University, Leicester, LE1 9BH, UK.
5. Brown, N., et al., *Longitudinal analysis of energy metering data from non-domestic buildings.* Building Research & Information, 2010. **38**(1): p. 80 - 91.
6. Brown, N., Wright, A.J, *New Approaches to the Gathering, Analysis and Interpretation of Half Hourly Energy Metering Data from Buildings.* International Journal of Low Carbon Technologies, 2007. **2**(1).
7. Younger, W., *Using interval meter data for improved facility management.* Energy Engineering: Journal of the Association of Energy Engineering, 2007. **104**(5): p. 21-33.
8. Brown, N., et al., *Novel Instrumentation for Monitoring After-Hours Electricity Consumption of Electrical Equipment.* Energy and Buildings, 2011.
9. Liddiard, R., *Characterising Space Use and Electricity Consumption in Non-domestic Buildings*, in *Institute of Energy and Sustainable Development*. 2012, DeMontfort University: Leicester, UK. p. 218.
10. Stuart, G., *Monitoring Energy Performance in Local Authority Buildings*, in *IESD*. 2011, DeMontfort University: Leicester.
11. Ferrera, V., *The analysis of primary metered half-hourly electricity and gas consumption in municipal buildings* in *IESD*. 2009, DeMontfort University, Leicester. Instituto Superior Técnico, Lisbon.
12. Stuart, G., et al., *Rapid analysis of time series data to identify changes in electricity consumption patterns in UK secondary schools.* Building and Environment, 2007. **42**(4): p. 1568.
13. Wright, A.J., Brown, N., *The Analysis and interpretation of half hourly utility data in UK buildings.*, in *Fourth International Conference Improving Energy Efficiency in Commercial Buildings*. 2008: Frankfurt.
14. Anon, *VDI 4602 Energy Management Standard: Energy Management. Terms, definitions*. 2007.
15. Mse, A., *A Management System for Energy*. 2008.
16. Chu Pulido, H.A. and I. Sanchez Ramos, *Present conditions and perspectives of the standardization of the energy efficiency in Mexico; Actualidad y perspectivas del programa de normalizacion de la eficiencia energetica en Mexico*. 1999.
17. ECS, B., *BS EN 61850-7-420:2009 Communication networks and systems for power utility automation. Basic communication structure. Distributed energy resources logical nodes.* 2009.
18. ISO, *ISO 22029: Standard file format for spectral data exchange* 2003.
19. ISO, *ISO8601: Data elements and interchange formats Information interchange Representation of dates and times*. 2003.
20. BSI, *BS EN 61970-407:2007 Energy management system application program interface (EMS-API) Part 407: Time Series Data Access* 2007.
21. BSI, *BS EN 60255-24:2001 IEC 60255-24:2001 Electrical relays Part 24: Common format for transient data exchange (COMTRADE) for power systems*. 2001.
22. BSI, *DD CEN/TS 15231:2006 DRAFT FOR DEVELOPMENT Open data communication in building automation, controls and building management Mapping between Lonworks and BACnet*. 2006.
23. BSI, *PD CEN/TR 16061:2010 Standards Application - Gas meters - Smart Gas Meters*. 2010.
24. ISO, B.E., *BS ISO 12174:2003 Space data and information transfer systems. Telecommand. Architectural specification for the data management service.* 2003.

25. BSI, *Draft BS EN 62714-1 Engineering data exchange format for use in industrial automation systems engineering (AutomationML) - Part 1: Architecture and General Requirements*. 2012.
26. CENELEC, *TECHNICAL REPORT CLC/TR 50403 Standardization and the liberalization of the energy market* 2002.
27. ECS, B., *BS EN 15232:2007 Energy performance of buildings. Impact of building automation, controls and building management.* 2007.
28. BSI, *BS EN 61970-405:2007 Energy management system application program interface (EMS-API) - Part 405: Generic eventing and Subscription (GES)*. 2007.
29. BSI, *BS EN ISO 15927-4:2005 Hygrothermal performance of buildings Calculation and presentation of climatic data Part 4: Hourly data for assessing the annual energy use for heating and cooling*. 2005.
30. ECS, B., *BS ISO 15112:2007 Natural gas. Energy determination.* 2007.