# Data gathering and architecture aspects of a major EU wide energy efficiency project for SMEs

Neil Brown<sup>(1)</sup>, Paul Fleming<sup>(1)</sup>, Nicoletta Favaretto<sup>(2)</sup> and Niall Sandford<sup>(1)</sup>

(1) Institute of Energy and Sustainable Development, De Montfort University, Leicester, UK. (2) Eurochambres, Brussels, Belgium

## Abstract

"Support and Training for an Excellent Energy Efficiency Performance" is a 3-year European project helping over 600 European cross-sector small and medium sized enterprises (SMEs) to reduce their energy use and become more energy-efficient. Companies participating in STEEEP benefit from tailored training and guidance on effective energy management tools and best practices provided by an established network of energy advisors from Chambers of Commerce and Industry (CCIs) in 10 different countries. SMEs in many EU countries employ over 90% of the workforce, so improving the energy efficiency of EU SMEs is therefore compelling, with clear advantages for the European economy. Energy efficiency in SMEs previously received less attention than in larger companies, the public sector and dwellings. Previously, policymakers had little energy (and related) data for SMEs, making prioritising ways to support energy conservation difficult. In addition staffing resources and knowledge levels within SMEs frequently determine the level of commitment to energy efficiency and implementing EU energy and climate policy, with a dedicated or even part time energy manager for many SMEs a rarity. The STEEEP project aims to help this by introducing training to SMEs via CCIs, and monitoring savings and providing feedback to SMEs. Crucial to this is the benchmarking of energy use: Basic data about the SMEs, the SME's energy consumption, and information about the SME policies and procedures relating to energy were gathered form each of the over 600 participants. Managing these data is a considerable task, notably in several languages, using combinations of numeric, free text and other data, gathered through questionnaires. It is not merely fiscal metering data, and supporting information that are gathered, we ask for from occupancy, to building types, and to complete the energy management matrix. We describe how this is done; the data processing . survey design, initial data gathering, benchmarking, and database architecture. Energy use is gathered as the project progresses, with interventions and changes captured. This paper describes the methods used and presents lessons learnt. This include the process of collecting, storing and analyzing the data from over 600 SMEs in 10 different countries. It identifies how barriers were overcome and how information from the data collection is being used by Chambers of Commerce and Industry to help reduce energy use of SMEs

## Introduction

Within the context of EU Energy and climate policies, the case for energy efficiency is compelling. The STEEEP (Support and Training for an Excellent Energy Efficiency Performance) project began in March 2014, involving EUROCHAMBRES, 35 Chambers of Commerce and Industry (CCIs) from 10 different European countries and the Institute of Energy and Sustainable Development, De Montfort University (DMU) [1]. These would provide 630 cross-sector SMEs with tailored training and guidance on effective energy management practices and tools, taking into account specific regional needs, enabling SMEs to approach energy management in the way that traditionally larger organisations have been able to [2].

A key part of the project is to be able to measure progress of SMEs, of different sizes, in different sectors in different countries. Our approach to this is to produce a series of benchmarks from the data that we collect.

Crucial to the process of producing a European benchmark for SMEs, is to receive a representative sample of each with respect to industrial classification. While utility datasets are large, information for energy efficiency improvements is rarely gathered, data being restricted in the main to fiscal billing. Published data for benchmarking for nondomestic buildings has tended in the past in most countries to concentrate on the public sector, so buildings such as council offices, government buildings, schools, are usually well represented [3]. However, the manufacturing sector has been frequently under represented with often insufficient samples in national nondomestic stock databases to produce sensible figures for benchmarking [4]. This a major contribution of the STEEEP project. The number

of SMEs involved means that this may be regarded as significant - from this it should be possible to pursue various levels of disaggregation with industrial classification codes to compare companies against their peers, as long as we have sufficiently representative sample numbers.

The approach described here in terms of data gathering is aimed at partner countries, this is due to the project structure since CCIs in partner countries receive project data, and while data will then be fed back to partners the same way, the benchmarking process itself may disregard countries and concentrate on industrial classifications. This builds enough sample numbers within the dataset for each classification type to perform useful benchmark calculations. The NACE descriptors [5] for industrial classification represent an extensible system of taxonomy for business types of SMEs. The codes can be used to describe the general area of activity, for example manufacturing, or finance, and subsequent extensions to the code may be used to describe company activities in more detail. Therefore, should a low sample number exist for a particular kind of disaggregation of company activity, we can achieve a broadly sensible result for benchmarking, albeit with slightly reduced precision, by moving up to the next level of aggregation for benchmarking purposes. For example. from mining organisation we may move from the mining of a specific mineral as described in the classifications, to, for example, general opencast mining. A reduced version of the codes was applied to improve the functionality of the long guestionnaire, to reduce the load on translation for each country's version, and to remove unnecessary activity types which would not apply to the companies recruited for the project.

#### **Qualitative Benchmarking**

The Energy Management Matrix (EMM) came into existence in the mid-1990s as a means of assessing the state of organisation of businesses with regard to certain key areas for responding to energy issues [6]. A major advantage of EMMs, is that they can quickly identify areas of excellence, and likewise, areas that needs improvement. By continually reassessing and using an EMM, performance can be tracked over time. EMMs can also be used to benchmark a business's organisation against that of other businesses. Despite the functionality of this tool, there is precious little literature relating specifically to EMMs, evenoutside the context of the EU. There are some case studies that highlight the importance of

EMMs, Zastava, a Serbian car manufacturer was able to achieve a remarkable 25% reduction of total energy use in the factory. The EMM was credited as a very effective way of quickly identifying areas in most need of attention [7]. A study conducted on the work of Envirowise notes the utility of EMMs. They noted that when EMMs were incorporated into a survey of UK businesses, 30% of companies had a strong energy management policy, yet only 10% were taking effective action [8] [9]. A sample EMM is shown below in figure 8.

LEVEL	POLICY AND SYSTEMS	ORGANIZATION	MOTIVATION	INFORMATION SYSTEMS	TRAINING AND AWARENESS	INVESTMENT
4 Multi- national	Formal energy / environmental policy and management system, action plan and regular review with commitment of senior management of senior management of senior esuprate enategy.	Energy / environmental management fully integrated into management structure Clear delegation of responsibility for energy use.	Formal and informal channels of communication regularly extincted by energy /cr/ricontrental manager and staff a all levels	Comprehensive system sets targets, monitors materials and energy consumption and westes and emissions, identifies faults, quantifies costs and savings and provides budget tracking	Marketing the value of material and energy efficiency and the performance of energy / environmental management both within the organization and outside it.	Positive discrimination in farout, of energy / environmental saving schemes with detailed investment appraisal of all new build and plant improvement opportunities
3	Formal energy / environmental policy, but no formal management system, and with no active commitment from top management	Energy / environmental manager accountable to energy committee, chaired by a member of the management board	Energy / environmental committee used as main channel together with direct contact with major users	Monitoring and targeting reports on individual premises assed on sub- metring / neglering, but savings not reported effectively to users	Programme of staff training awareness and regular publicity campaigns	Same pay back criteria as for all other investments. Cursory aperaisal of new build and plant improvement apportunities.
2 Family owned	Unadopted / informal energy / environmental policy set by energy / environmental manager or senior departmental manager	Energy / environmental manager in post, reporting to ad-hoc committee but line management and authority are unclear	Contact with major users through at hoc committee claired by senior departmental manager	Monitoring and targeting reports based on supply meter /measurement data and invoices. Env. / energy staff have ad-hoc involvement in budget setting	Some ad hoc staff awareness and training	Investment using short term pay eack criteria mostly
1	An unwritten set of guidelines	Energy / environmental management the part-time responsibility of someone with only limited influence or authority	Informal contacts between engineer and a few users	Cost reporting based on invoice data. Engineer compiles reports for internal use within technical department	Informal contacts used to promote energy efficiency and resource conservation	Only low cost measures taken
0	No explicit policy	No energy / environmental manager or any formal delegation of responsibility for env / energy use.	No contact with users	No information system. No accounting for materials and energy consumption and waste	No awareness raising of energy efficiency and resource conservation	No investment in increasing environmental performance / energy efficiency in premises

Figure 1 - Energy Management matrix [8]

#### **Data Gathering**

The project consortium or Eurochambres, the CCIs and DMU agreed on a common methodology to collect relevant data from SMEs, using two questionnaires - data were collected via the submission of an initial online questionnaire, with updates on energy use and any changes made, via a shorter online form, every two months.

## (Initial) Long Questionnaire

Energy benchmarking methodologies and data are relatively scarce for industry, and even more so for SMEs, with some areas of industry represented little in non-domestic energy datasets [3]. The first in a set of two questionnaires focuses on the nature of the SME and its energy and building use. Specifically, it asks for information on the participating company (e.g. sector of activity, level of activity), descriptions and readings of its energy meters, information on past energy use, basic energy use and relevant production data, or any other quantifiable data that energy use may be measured against [10]. To make accurate energy saving recommendations, it is necessary to understand existing demands on energy, notably if they vary from one year to the next [11], so questions are included on local climate, building occupancy and building use, indeed, all data needed to produce an energy benchmark from a potentially disparate set of data [4]. Technical building services, such as ventilation and compressed air, can use more energy as a constant load than production in some cases, so the questionnaire asks about these, also the presence of any heavy machinery or plant [12], and building characteristics [13]. The second part of the questionnaire gathered qualitative data on the organisational culture of supported SMEs with a view to reducing energy use. Mere technical interventions only take us part of the way towards serious energy use reduction, and taking human factors into account allows us to give better advice on energy efficiency, including maintaining it . Questions relating to the Energy Management include presence and type of energy policy. organisation, the level of staff training that may or may not have been given, types of performance measurement, level of communication and type of investment.

Further questions would analyse motivations, attitudes and perceptions of control of participating SMEs. Questions will show if SMEs are motivated to change their energy use, have a positive attitude towards saving energy, and their level of knowledge on energy efficiency. Comparing attitudinal behaviour with actual energy data will support the related evaluation and recommendations for effective energy strategies.

The questionnaires provide 490 columns of data, some which may be empty, since it allows for up to 10 streams (from different meters) of energy (building) data, the same for energy (production) data, and multiple buildings, with subsets of data for each, such as fuel type - in general, most SMEs tend to have a maximum of two energy meters (gas and electricity), with many having electricity only. The main question groupings are summarised below in table 1.

Question Group	Question types
Operational	Company name, contact information, location (address)
Benchmark	Building details, square meterage, company type, location (lat, long)
Building Physics	Building construction, height, plant, treatments (e.g. air conditioningg)
Energy (Building)	Main meters, meter reference numbers, meter types, previous data,
	fuel type, connected buildings or areas
Energy (Production)	Main meters, meter reference numbers, meter types, previous data,
	fuel type, connected buildings or areas
Qualitative	Energy Management Matrix and supporting data
Supporting data	Multi use buildings, seasonal occupancy or use

#### Table 1: Long Questionnaire (up to 490 data entries in total)

#### Short questionnaire

This short questionnaire was designed to be much quicker to fill in, asking seven questions on energy use, production or similar performance indicators and weather (Table 2). Also, SMEs report relevant information on any energy conservation intervention or investment, and any events which may have affected the pattern of energy use. Operational details, such as change of the main contact on site, will also be gathered. The form is integrated on the STEEEP website. All companies have a personal account and password to access the short questionnaire. After the first submission, businesses may see the history of their entries and, consequently keep track of their own energy use. Both questionnaires have been translated by partners in 10 different languages to facilitate SMEs to report the requested data. Apart from English, translations were made into Croatian, Dutch, Estonian, French, German, Hungarian, Italian, Latvian, Romanian and Spanish.

Question group	Question types				
Operational	Company name, company ID number				
Energy (Building)	Main meters, meter reference numbers, meter data				
Energy (Production)	Main meters, meter reference numbers, meter data				
Supporting data	Any changes to be made to Long Questionnaire data, details of any energy savings				

## Table 2: Short Questionnaire

## Data handling

A spreadsheet was used as the main 'terminal' to the software, such that hand-cleaning of all data is straightforward. However, it must be remembered that no calculations are to be carried out within spreadsheets, which are then useful to the flat file database. The spreadsheet therefore is simply used as a visualisation and editing too, with GNU Octave used for the detailed analysisThe following physical data flow diagram (figure 2) shows how software fits together for processing of long questionnaire data.



Figure 2 - Benchmarking dataflows

Floor areas are extracted from the data using a custom algorithm (area\_clean\_f in figure 2). These data need cleaning since many errors arise, for example sometimes text is mixed in with data, and a reference to be appendix will show parts of software which are used to remove frequently occurring words in all languages. Sometimes an SME may have indicated, for example if a floor area is approximate, e.g. "3453 SQ.M APPROXIMATIV", although the position of numbers within text descriptors is not a given. However, while such details are preserved in a copy of the full\_data array (because it is good practice to preserve raw data), text cannot be processed by machine of course to input to the equation which calculates benchmarks. So floor areas are checked by hand just in case any text data is given which could hamper data processing.

The ISO format for presentation of numbers is not adhered to routinely across Europe, What is certain is that the variation in presentation of numbers between countries can be significant, for example a "." or ",", may be used as the decimal separator in some countries, and a corresponding "," or "." may be used as 1000 separator in these same countries. Conversely, most software for processing numbers available from the UK and the USA uses a "." as a decimal separator, and sometimes a "," as 1000 separator by default. Conversion between these systems is easy to manage. However, to compound this, sometimes a space is used as 1000 separator, anathema to processing of text to numbers, and in some participating countries, symbols such as "" may be used with 1000 separator too. To this end, floor areas, energy usage and other key variables are machine cleaned, but subsequently checked by hand as part of the project's quality control strategy.

Cleaning of data for meter reading dates is rather more complicated. Date data have been provided whereby it is usable, but sometimes a string of numbers or text were entered by some SMEs - a text

to date function (nice\_datenum\_f in figure 2) runs in an error trap wrapper, which has a fairly high processing overhead and runs slowly, but only needs to run once, the whole dataset processed in around 10 minutes - the function exports dates from and to the meter reading purposes, so that we can count the number of days between meter readings and use this to calculate benchmarks or so tables are produced to let us know whether or not the entered dates are usable. Variables are then placed in the Octave workspace, whereby they can be viewed, and saved/backed up. The full data set exported to HTML compilation includes the full data itself, cleaned data tables, and all arrays which indicate whether or not the data from a company is usable. This is so that benchmarks are not generated from missing data, with simply an error message sent back to the SME.

The main data types which are captured are for electricity, gas, oil, liquefied petroleum gas, biofuels/ biomass, none, or to query the energy type. We noticed that some SMEs are still using coal so this will need updating. A mixture of upper and lowercase characters appear in the survey data here. which cannot be processed normally, so all data cells must be extracted and made into lowercase before processing. The software then translates fuel types from every language in the project (energy types f shown in figure 2), which is relatively straightforward since many similar letters appear in each country's word for electricity or gas, so we can run searches on abbreviated versions. An output log file is also generated which is available to read for anybody who may find it useful showing a detailed description of all meters and meter data types for every company. The function energy data f (figure 2) is used to check whether meter readings can be used for benchmarking, and to check where the units are in kilowatts, cubic metres, and so forth. When an empty cell is encountered, should further metering data be added in the short questionnaire, a message is generated that this needs to be done. Where text appears in the multiplier type, then this is logged. So, as the software is running, it reports on finding empty cells, whether a cell contains text, if there is a validated meter reading for benchmarking, or if something else appears in the meter reading cell. This function also generates a log file which is available for SMEs, or CCIs. Company names cannot be used as the primary key to search for their details, since companies may type names in differently between the long and short questionnaires. Assigning a number, while errors may still take place, reduces ambiguity [Some Aspects of a Framework for Energy Data., Brown et al., 2012].

New companies are best grouped by country, and physically grouped in data tables with their counterparts. Sequential numbering would mean that companies could only be added at the end of the main data block, making any visualisation or direct printing difficult. The solution to this therefore is to leave a gap between ID numbers, such as was done in the early days of programming where line numbers would increase by multiples of 10, so that spare lines could be inserted. Should any particular line need to be ignored this is not deleted from the main data table, or rather is represented by a signal in a separate array which lets the software know whether it should process or ignore the entry for a particular company. This also is used to remove duplicate or test entries, but at the time of final processing only. It follows that grouping by country may be allowed for each identifier as an extra data quality check. The quickest way to do this is simply to add the international dialling code for each company.

Thus the identification structure is: Incremental ID, country dialling code.

### Output

Benchmarking software produces summary files in HTML (Figures 3 and 4) - this enables connection between any further glue code, databases, and word processors with formatted printed output for SMEs - also it means that automated import of graphics such as plots of energy use are possible. Once all data were cleaned, it was possible to load these processed results into a PostGreSQL database, with a skeleton copy of the Octave processing code used to continue to produce HTML containing long questionnaire data (and processed data / information) for SMEs, including energy use, energy benchmarks, where available, and this is also where any energy-saving recommendations may appear. To extend this process to handle short questionnaire data is a relatively simple process requiring an extra import, and an extra processing module. This will then be used to produce a series of benchmarks for each company, whereby it will be possible to track any increase or decrease in normalised energy use.



Figure 3 (left) Sample Page 1 of HTML, Figure 4 (right) Sample Page 2 of HTML

# Data snapshots

This is a method paper but nevertheless it is still interesting to present results at this stage, notwithstanding of course final results from interventions remain to be seen. Figure 5 shows the distribution cohort wide of SME types by NACE descriptors, and shows a healthy sample size for manufacturing, which has been previously under represented in many datasets.



Figure 5 – Cohort by NACE grouping

## **Quantitative Benchmarking**

Sample benchmarking results are shown below in figures 6 and 7 for gas and electricity use across the whole cohort by count. These data refer to 2014 and before.





## Qualitative benchmarking

Figure 8 shows the results of an EMM survey of the whole cohort, on first viewing, confidence in investment appears to stand out as higher than other factors, but the overall picture is not especially clear - what becomes fascinating is when grouping all data by country and averaged EMM, as shown in figure 9, which shows clear needs for training and communication. This suggests strongly that these data validate the ethos of increasing training in energy efficiency in SMEs.







Figure 9 - EMM values averaged by country

Analysis by country and by NACE sector has shown that in general, EMM data are too noisy to draw firm conclusions geographically or by sector, which makes the mean values particularly interesting. Future work will involve looking for correlations between qualitative and quantitative benchmarking, as well as benchmarking on production.

# Conclusions

We have presented an effective methodology for initial data gathering and energy benchmarking for over 600 EU wide SMEs. Crucial to any project of this scale is to budget effectively for data preparation and cleaning, notably when data are collected in many languages. A hybrid approach to this may automate much data preparation using scripting languages (such as Octave) before data are loaded into a relational database. The usefulness of a joint approach combining quantitative and qualitative benchmarking is to be explored as benchmarking progresses.

The effectiveness of gathering data by questionnaire for the project has been proven, and engagement from SMEs is very encouraging, with participants continuing to regularly submit data. Enthusiasm for energy efficiency amongst SMEs strongly suggests that this area of energy efficiency, arguably overlooked in some cases in the past, is an area where resources for saving energy whilst maximising productivity may be well spent. As the effects of training and interventions become known, the next step for data analysis is a longditudinal study to analyse energy savings.

# Acknowledgements

The work described in this paper was undertaken as part of the project STEEEP : Support and Training for an Excellent Energy Efficiency Performance, project leaders are EUROCHAMBRES, Avenue des Arts, 19 A/D, B1000 Brussels, Belgium, Duration: 01/03/2014 to 28/02/2017, EU funding Contract number: IEE/13/844/SI2.675838. The project website is <u>www.steeep.eu</u> The authors would like to thank all project partners, CCIs and SMES for their cooperation.



Co-funded by the Intelligent Energy Europe Programme of the European Union

The sole responsibility for the content of this paper lies with the authors. It does not necessarily reflect the opinion of the European Union. Neither the EASME nor the European Commission are responsible for any use that may be made of the information contained therein.

#### References

- STEEEP Support and Training for an Excellent Energy Efficiency Performance www.steeep.eu EUROCHAMBRES, Avenue des Arts, 19 A/D, B1000 Brussels, Belgium, Duration: 01/03/2014 to 28/02/2017 Contract number: IEE/13/844/SI2.675838
- [2] Webber, P., Chauhan, D., Morris, N., Fleming, P., Stuart, G. and Ferreira, V. (2011) Intelligent metering for energy savings in Small and Medium Sized Enterprises : experiences at the local level in 5 European countries. Proc ECEEE Summer Study 2011
- [3] Steadman P, Bruhns H R, Rickaby P A, 2000, An introduction to the national Non-Domestic Building Stock database. Environment and Planning B: Planning and Design 27(1) 3 10
- [4] Brown, N Greenough, R., Vikhorev K. Precursors to using Energy Data as a Manufacturing Process Variable. IEEE-DEST-2012 Conference on Digital Ecosystems, Campione d'Italia
- [5] NACE Rev. 2 Statistical classification of economic activities in the European Community. Luxembourg: Office for Official Publications of the European Communities, 2008 ISBN 978-92-79-04741-1
- [6] S. Fawkes, Outsourcing Energy Management: Saving Energy and Carbon Through Partnering, 1 ed., Aldershot: Gower Publishing Limited, 2007.
- [7] D. Gordic, M. Babic, N. Jovicic, V. Sustersic, D. Koncalovic and D. Jelic, "Development of energy management system – Case study of Serbian car manufacturer," Energy Conversion and Management, vol. 51, no. 12, pp. 2783-2790, 2010.
- [8] M. Gibson, "The work of Envirowise in driving forward UK industrial waste reduction," Resources, Conservation and Recycling, vol. 32, no. 3-4, pp. 191-202, 200135
- [9] Chemical Industries Association, "Energy Efficiency Best Practice Programme Energy Management Matrix," Chemical Industries Association, Harwell, 1998.
- [10] Brown, N., Greenough, R., Vikhorev K. Some Aspects of a Framework for Energy Data. Improving Energy Efficiency in Commercial Buildings, Frankfurt, 2012
- [11] Brown, N., Wright, A.J, Shukla, A., Stuart, G., Longitudinal analysis of energy metering data from non-domestic buildings, Building Research & Information, Volume 38, Issue 1, 2010
- [12] CIBSE CIBSE TM46 Energy Benchmark. The Chartered Institution of Building Services Engineers 222 Balham High Road, London SW12 9BS, 2008.
- [13] Software Engineering for Building Energy Efficiency N Brown, N., Snape, R., Khattak, S., Greenough, R. Improving Energy Efficiency in Commercial Buildings, Frankfurt, 2014