Scottish Public Sector Energy Benchmarking Tool (SPSEBT)

Technical Summary

Prepared by: Dr Jon Stinson, BRS

Contents

List of figures

List of tables

Declaration

This is a technical report for Zero Waste Scotland and the Scottish Energy Officers Network (SEON) and is the Technical Summary associated with a series of other technical papers and handover documentation; this declaration applies to all documents submitted in fulfilment of the Scottish Public Sector Energy Benchmarking Tool project deliverables. Building Research Solutions Ltd (BRS) have taken all reasonable steps to evaluate the details enclosed within, and to ensure that the content and information contained in this document is correct in all material respects. However, the Recipient is recommended to interpret the content with caution and ensure that geographic and context-specific elements are carefully considered when applying the findings. While every effort is made to ensure the accuracy of all that is contained within, BRS cannot accept liability for loss or damage however caused arising from the use of the information supplied. BRS do not provide Expert Witness services for use in any kind of legal proceedings.

Nothing presented within this document is intended to be or should be interpreted as an endorsement or recommendation of any supplier, service, product or method. BRS are independent of any and all materials, technologies or methods pertaining to use in the construction or energy performance of buildings. BRS are neither partially or wholly owned by any organisation which sells or deals with the manufacture and/or installation of any building material and/or technology. All findings contained within this technical paper are specific to the locations stated, i.e. Scotland, and are not necessarily applicable to other unspecified locations.

BRS took every reasonably practicable step to ensure the quality and accuracy of the inputs and results reported in this technical paper and the associated handover documentation. However, given the inconsistency in data availability from the surveys, volume and delivery dates of AMR data, and the novel method employed for the AMR Profilers, results should be interpreted with an understanding of the limitations behind the underlying data. These limitations have been reported throughout the handover documentation that accompanies the Scottish Public Sector Energy Benchmarking Tool.

Executive Summary

Background

In 2018, SEON, which comprises Local Authority representatives along with representation from other parts of the public sector, undertook an energy benchmarking exercise, also referred to as the Public Sector Benchmarking Tool (PSBT) Phase 1. The analysis covered 27 Local Authority property portfolios and included over 9,000 individual sites and produced a set of energy benchmarks for 24 categories of buildings. The Scottish Public Sector Energy Benchmarking Tool (SPSEBT 2022) builds on this previous work undertaken by SEON by establishing a comprehensive set of Scotland-specific energy benchmarks for all building types within the wider public sector. This project also explored the potential for creating high resolution (15- and 30-minute) consumption profiles for individual building activity types (BATs) for electricity, gas and water.

The SPSEBT 2022 supersedes the PSBT Phase 1 by:

- Including a wider range of public sector organisations, beyond local authorities
- Expanding and redefining Primary Building Categories, into 63 Building Activity Types. These BATs harmonise more closely with the National Calculation Methods (NCM) Simplified Building Energy Model (SBEM) building databases, Energy Performance Certificates building types, and other Government tools, e.g. the Energy Heat Map, EPC non-domestic register, etc.
- Includes benchmark values for operational carbon dioxide emissions
- Ensuring data anonymity for all buildings and organisations used to create the benchmarks, presented in an accessible, secure, online browser-based platform
- Harmonises the benchmarks for heat consumption to include electricity
- Expanding filters to include retrofit, HVAC types, BEMS, and heat distribution types
- Increasing usability and user experience through an optimised set of data visualisation methods
- Providing an AMR profile for utility data relating to electricity (non-heat), gas, and water consumption.

The provision of these benchmarks supports public sector building managers to improve energy efficiency and produces a standardised dataset for all Scottish Public Sector buildings that will be of ongoing benefit in responding to current and future net-zero carbon focused initiatives.

Scope of work

The initial stage began with a stakeholder engagement and recruitment event in Summer 2021 to collect the necessary breadth and depth of data, which was a collaborative exercise. From this event, 150 public sector organisations received the first official email survey to build both the benchmarks and AMR profiles.

All Scottish Public Sector organisations were contacted in April 2021 with an overview of the planned initiative and an invitation to attend a stakeholder engagement and recruitment event in May 2021 to:

- inform participants of the goals and aspirations of the project
- define and agree the necessary breadth and depth of data required
- highlight actions required of participants and the appointed consultant.

At the project initiation stage, 150 organisations had confirmed participation from across the Scottish Public Sector property portfolio.

Significant support was provided by SEON, Zero Waste Scotland, Sustainable Scotland Network (SSN), NHS Health Facilities Scotland, Environmental Association of Universities and Colleges, and Community Leisure UK for access to specific sectoral groups.

A spreadsheet-based questionnaire, relating to building information and utility consumption, was issued to all participants, with returns consolidated, verified, validated, and conditioned. A comprehensive list of benchmarks was subsequently developed and refined and, to aid accessibility and understanding, an associated benchmarking dashboard was created.

Benchmarking Dashboard

The benchmark dashboard is a free to use, browser-based platform, which defines the electricity consumption, heat energy consumption, water consumption and associated carbon dioxide emissions for 63 building activity types (BATs). The values for each BAT were statistically extracted from data from 6331 Scottish Public Sector buildings.

The definitive benchmark table is presented in Appendix 1 of this document (which can also be downloaded from the SPSEBT's export feature). This table provides statistically accurate values for each BAT.

The SPSEBT dashboard provides filtering options to refine the benchmark values, allowing the user to 'build their own benchmark' so the value can be calibrated, as far as reasonably practicable, to provide a more accurate comparison to building specific factors, e.g. age, heating systems, applied energy efficiency works, etc.

Appendix 2 details a full list of data fields contained within the SPSEBT, defined by application category and further by whether the data was source from participants or calculated/derived.

AMR Profiler

The AMR profiler is a free to use spreadsheet tool, allowing users to insert a building's half-hourly or 15-minute AMR data. The AMR Profiler allows for a comparison of a building's weekly electricity, gas, and water consumption, to a statistically generated profile, thereby allowing identification of consumption patterns which are outwith normal or typical expectations for a building of that type. The AMR Profiler provides insights with two data visualisation systems:

1. Average week per month overlay chart
2. Annual half-hourly or 15-minute colour coded heatmap.

The AMR Profiler is currently only available for 2 building types: primary schools and open plan offices.

Lessons learnt

Below is a list of the main lessons learnt from both the methods employed to collect, condition, process, and analyse the data during this project and also from the participants who provided the raw data:

1. **Standardisation of the data collection** – The data capture questionnaire was optimised to collect the minimal required information to produce robust benchmarks. However, requests for certain information presented an obstacle to many organisations, i.e., utility data held separately to building characteristic information (e.g., building age, floor area, HVAC systems, retrofit level, etc).

2. **AMR data capture** – The procedures to handle, condition, process, and analyse AMR data proved to be resource heavy. It was evident that the utility providers required additional support during the AMR data collection phase. For future iterations, either the utility provider should be stakeholders within the project, whereby the quality and delivery of the AMR data is either guaranteed by a certain date or processed and conditioned by their in-house staff. Or a data sharing agreement is setup, between the research team and the utility providers in advance of any data analysis phase, this agreement could take the form of automated data sharing on a weekly or monthly basis.

Three items are paramount for future iterations of the SPSEBT AMR Profiler:

    a. Utility consumption data requirements should be defined at an early stage to ensure correlation between all data provision sources, e.g. participants and utility providers.

    b. AMR data formats should be defined and approved by all relevant parties at an early stage to avoid issues with automated integration of large data sets,

    c. Requests for AMR held by the organisation should be made to the organisation during the next LoA survey. This would serve to streamline the AMR collection process, reduce dependency on utility companies, and potentially increase the number of AMR datasets, to include organisations who are either not on the same framework as the utility provider, and/or who do not have HH or 15-miinute AMR accounts for some or all buildings.

3. **Harmonising public sector reporting** – For many participating organisations, and those who declared an inability to participate, overlapping reporting commitments was the primary reason provided for delays, partial returns, and drop-outs. It was observed that the Public Bodies Climate Change Duties Annual Reporting Requirements deadline fell over the same period that the SPSEBT 2022 data collection surveys were issued. Future public sector benchmarking data capture activities would benefit from either:

    a. Aligning with SSN data capture calendar, where a unified questionnaire survey or new reporting platform could be optimised to reduce pressure on organisations. or

    b. Selection of an optimal period to avoid, as far reasonably practicable, times of multiple reporting

Next steps – Phase 3

To provide Scotland's Public Sector organisations, stakeholders, and policy makers with the most robust decision-making tools available, the SPSEBT online dashboard and AMR Profiler should and must evolve. To enable this evolutionary process, some key aspects for consideration are noted below:

1. **Wider coverage of buildings** – Participation of organisations in underrepresented sectors and building activity types. Scotland has approximately 20,000 public sector buildings. Data captured in Phase 2 covered 7,243 sites (36% of the total public sector building stock); of these, 6,331 were suitable for statistical analysis (32% of the total public sector building stock).

2. **Data Reporting Integration** - Further investigation into how the Benchmarking Tool can be used to improve annual mandatory climate change reporting (coordinated on behalf of Scottish Government by SSN) and how mandatory reporting could be used to drive on-going data gathering to aid the creation of future benchmarks while avoiding duplication of effort across the public sector.

3. **A SPSEBT central portal** – A centrally managed and secure database of information held on currently captured buildings, with remote, organisation level specific login features, i.e. a database portal. Thereby streamlining the data capture process to remove the need for the, somewhat limited, Excel questionnaire surveys, and eliminating the transfer of potentially sensitive data via email. This portal could be designed to hold building information for buildings already collected, whereby future information requests would be a matter of an organisation checking and updating their information rather than recreating it.

4. **'Design year' AMR Profiler** – the SPSEBT AMR Profiler is built using AMR data from 2019. Future iterations of this tool should seek to secure multiple years of AMR data from the participating utility companies. In a similar fashion to dynamic simulation modelling (DSM) for energy demand calculations of new build and retrofit projects, the AMR Profiler would benefit from being built from a 'design year'. Whilst not intended to be used for forecasting or predictive projections, a design year profile would increase the confidence of the bandwidth and reduce its dependency on local factors that have affected 1 particular month in one particular year.

5. **AMR Profiler add-in** – the next evolutionary stage for the SPSEBT AMR profiler is to potentially move calculation processes from a method that requires the user to paste the AMR data into a specific file, to allowing the user to do a profile comparison within their own Excel sheet, i.e. not having to paste data from one sheet into another.

1 SPSEBT

The Scottish Public Sector Energy Benchmarking Tool (SPSEBT) comes in two parts:

1. The browser-based, build your own benchmark dashboard
2. An AMR (automatic meter reading) Profiler.

Part 1 of the SPSEBT is a free to use, browser-based platform, which defines the electricity consumption, heat energy consumption, water consumption and associated carbon dioxide emissions for 63 building activity types (BATs). The values for each BAT were statistically extracted from data from 6331 Scottish public sector buildings. The definitive benchmark table is presented in Appendix 1 of this document, which can also be downloaded as an .xlsx file from the SPSEBT's export feature. This table provides the statistically accurate values for each BAT. The SPSEBT provides filtering options to refine the benchmark values, allowing the user to 'build their own benchmark' so the value is calibrated, as far as reasonably practicable, to provide a number (or numbers) that more accurately represent a given real-world building. The values in the Appendix 1 table should be used when referring to an unfiltered benchmark value for any of the 63 BATs, the values on the SPSEBT should be used after any of the available filtering options are applied to refine the benchmark for any of the 63 BATs. Appendix 2 presents the list of fields used to create the SPSEBT. This shows the variables that were collected directly from the participating organisation, and those that were derived to power the SPSEBT.

Part 2 of the SPSEBT is a free to use spreadsheet tool, allowing users to insert a building's half-hourly or 15-minute AMR data. The AMR Profiler allows for a comparison of a building's weekly electricity (non-heat), gas, and water consumption, to a statistically generated profile, thereby allowing the user to identify times when the building's consumption is 'out of profile'. The AMR Profiler provides insights with two data visualisation systems:

3. Average week per month overlay chart
4. Half-hourly or 15-minute colour coded heatmap.

The AMR Profiler is currently available for 2 BATs: primary schools and open plan offices. At this stage, SPSEBT Part 2 is usable within the Microsoft Excel ecosystem.

This executive summary provides a concise overview of the SPSEBT and the project activities that led to the delivery of the Tool. Here, we introduce both parts of the Tool and describe important elements of the methods applied to deliver them. A library of documentation is available which describes, in greater detail, each of the topics presented in this executive summary. These are available for each of the sub-sections in Section 2, upon request from Zero Waste Scotland.

1.1 Purpose of the tool

Energy benchmarks provide stakeholders with actionable, comparative values that act as an industry calibrated gauge, which help to understand a building's performance on a national scale. The SPSEBT 2022 provides access to a robust and wide set of building related benchmarks, designed to support the energy, sustainability, building management, and construction sectors. SPSEBT provides reliable, locally sourced, and refinable benchmarks to support of wide variety of local and national agendas relating to energy efficiency, decentralised heat networks, retrofit rollout, and decarbonisation of energy infrastructures.

The SPSEBT 2022 replaces any previously applicable benchmark values for energy, water and carbon related to buildings owned and operated by public sector organisations in Scotland, i.e. PSBT Phase 1 delivered by UCL.

1.2 How to use the tool

Presented on a browser accessible webpage, Part 1 of the SPSEBT uses a data driven, graphical user interface to access the benchmarking data. The information is presented as:

- A map of Scotland
- Dot charts, showing 3 values per building activity type
- Pie charts and column charts, showing a breakdown of various energy related variables
- Tables.

The SPSEBT was designed and refined to be accessible and intuitive to suit a wide range of users, and their requirements and expectations. Therefore, the information is presented over 2 levels (shown in Figure 1):

- Level 1 – Main level, shows benchmarks for 10 building activity types that group the 63 BATs seen on Level 2, these 10 are referred to as Parent BATs. This main level is designed for an at-a-glance overview of the whole benchmark dataset, aimed to provide higher level information for those seeking a national level view.

- Level 2 – Detailed level, shows benchmarks for 63 BATs, referred to as Sub-BATs. This level is designed for users who want a comparative benchmark, which includes certain specific characteristics. Benchmark values for Sub-BATs can be filtered, using the Tools' top filtering ribbon, to be presented per Parent BAT, which greatly improves the readability of the values. The SPSEBT filters provide users with greater insights, allowing them to 'build their own benchmark', which will more accurately provide a comparative benchmark for the building or buildings that they wish to gauge. When filters are applied, SPSEBT provides statistically robust and authoritative benchmarks for any available combination of building characteristics. The definitive 63 Sub-BAT benchmarks, where no filters are applied, should be taken from Appendix 1 of this document, or from the SPSEBT export feature.
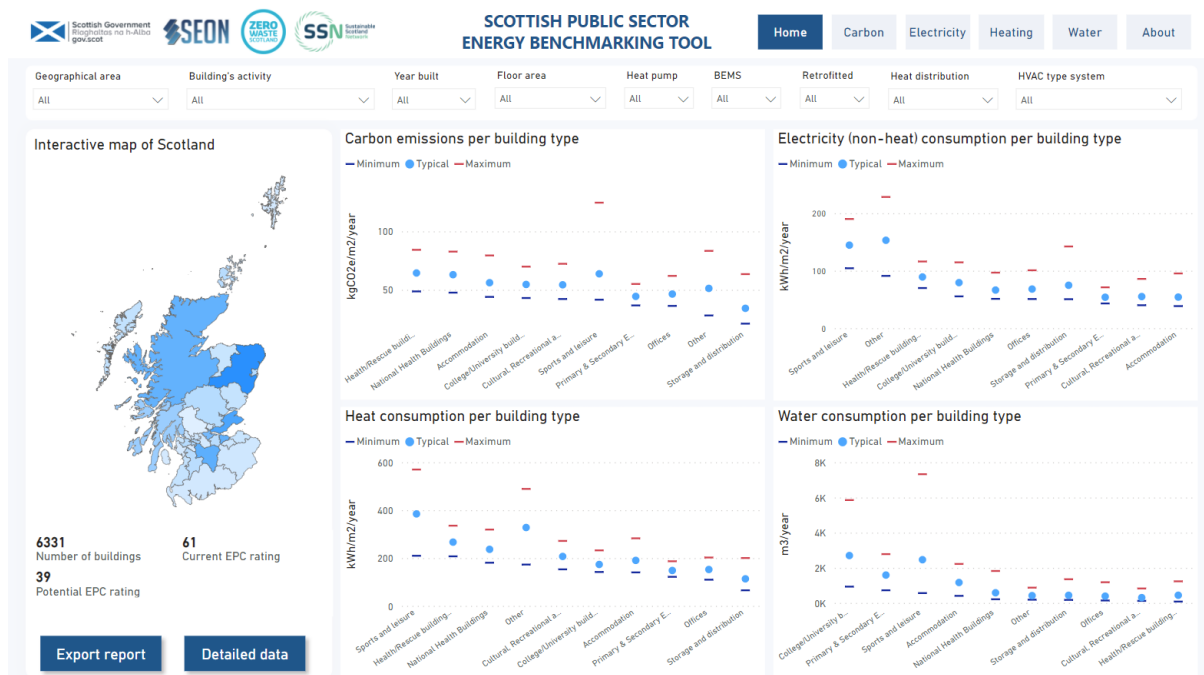
Figure 1: Screenshots of the SPSEBT home page for Level 1 (top) and Level 2 (bottom)

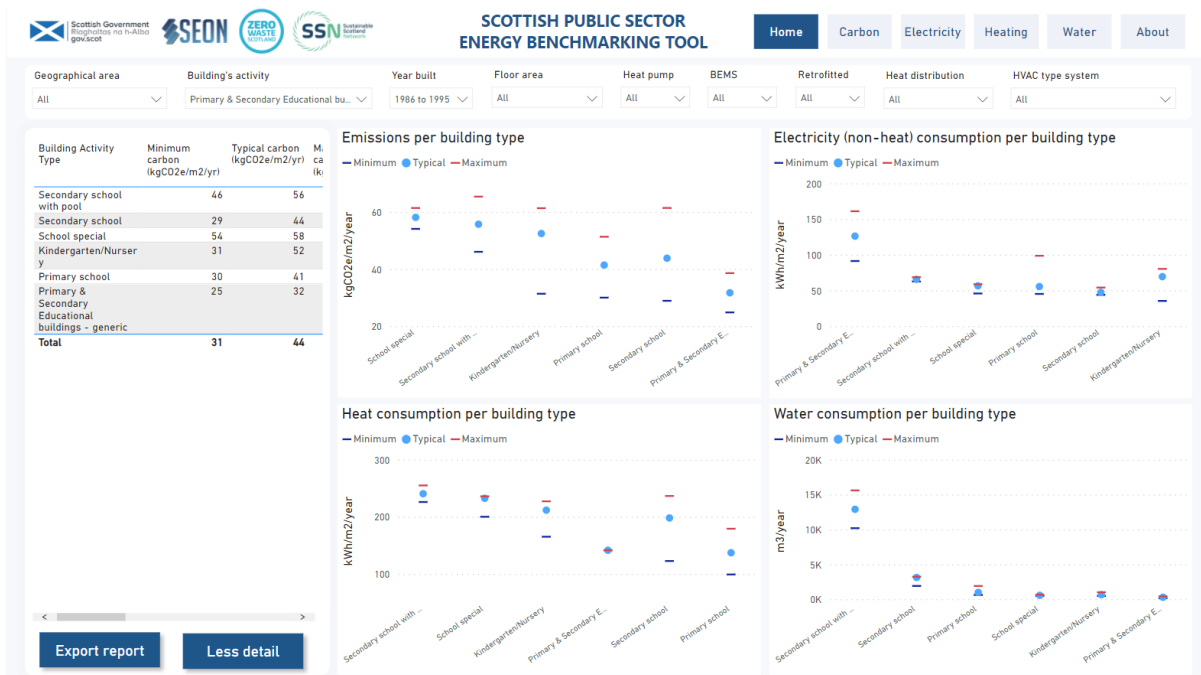Each level provides access to benchmarks for electricity (non-heat) consumption, heating consumption, water consumption, and the associated carbon dioxide emissions. All values contained within are specific to operational energy and carbon, they are not related to embodied values.

All charts, tables, and maps within the SPSEBT are interactive, meaning that the values in the infographics will respond to the filters selected from the Tool's top filter ribbon. Derived from 6331 buildings, collected from 78 Scottish public sector organisations, representing 7 main sector groups, the filters provide the user with the ability to create a benchmark based on any combination of these characteristics:

1. Parent Building Activity Type
2. Geographical area
3. Age band of construction
4. Floor area band
5. With or without a heat pump
6. With or without a building energy management system (BEMS)
7. With or without thermal fabric retrofit measures
8. Various heating, ventilation, and air conditioning types
9. Various heat distribution systems.

The benchmark information presented in Levels 1 and 2 provide reliable, statistically robust, and authoritative values. With the variety of variables used to create the SPSEBT (see Appendix 2), Level 2 is best used when the filters (above) are applied. To ensure that a statistically representative range is available for all 63 Sub-BATs, the benchmark values available in Appendix 1 of this document, and from the SPSEBT export feature, should be taken as the definitive benchmark values for any Sub-BAT where a filter is not applied.

SPSEBT Part 2 provides results by comparing a user's AMR data to that of a statistically generated profile. These profiles are generated and visualised providing two layers of insight. The first layer presents the results in a heatmap format, where each value for each time stamp is coloured to show the building's consumption level in relation to the profile. Colours are used to visualise times and days when the building's consumption is 'out of profile'. The second layer presents results over 12 charts; each chart plots the daily consumption, averaged for a week per month. These charts visualise the

absolute profile line and a profile bandwidth; with the building's AMR data overlain, these charts provide greater insights to when the building's profile is within or outwith the statistical profile. Figure 2 shows an example of the About page for the AMR Profiler, Figure 3 and Figure 4 provide an example of the heatmap and chart, respectively.

This **Gas AMR profiler** was created as part of the Scottish Public Sector Energy Benchmarking Tool (SPSEBT) project delivered by Building Research Solutions Ltd for Zero Waste Scotland, the Scottish Energy Officers Network (SEON), and Sustainable Scotland Network, and funded by the Scottish Government. The aim of this AMR profiler is to allows users to compare their AMR data with that of the public sector benchmarks, for specific building activity types. This profile is for "Primary Schools" which was done as a proof of concept, along with "Open Plan Offices", and as such it has limited functionality (see notes below). Greater functionality can potentially be added in Phase 3 of the SPSEBT project.

This spreadsheet contains 3 tabs:
- **About** - which explains the existence of the AMR profiler and how to use it/caveats (described below).
- **User Input** - allows users to enter their specific building AMR data for their reporting year. This tab also provides a heat map view of the users AMR data, compared against the public sector benchmarks, for 2019.
- **Chart** - presents a visual comparison of the users AMR data against the public sector benchmarks, for 2019.

Caveats:
In the User Input tab, there are a few **important points to note when entering data to avoid errors**:
- The building floor area, in m2, that correlates to the user's building's AMR data **must be entered into Cell B2**. A floor area value must be entered into Cell B2 as this is used to calibrate the background AMR data that creates the profiles in the charts. If the floor area is unknown, either a realistic assumption should be entered or that building's data cannot be used in this profiler.
- The AMR data **must** span from the **00:00 time stamp to the 23:30 time stamp** - i.e. it must be in the same format as the User Input tab.
- The users AMR data must be pasted into the User Input tab **as values**, to maintain the conditional formatting on the cells, the date column must be pasted in as a short date.
- Only one year's worth of data can be input (i.e. 365 rows) and shown on the charts.
- **DO NOT** delete or insert rows in the User Input tab as this will affect the background calculations. Only paste, as values, over existing data (if any), starting from Cell E3 (the 00:00 time stamp).

Description of the User Input tab heat map:
The data behind the public sector AMR profile bandwidth (i.e. typical, minimum and maximum benchmarks; see Charts) has been cleaned, conditioned and statistically analysed into 3 percentiles - the 50th percentile (i.e. 2nd quartile or median, termed "typical"); the 25th percentile (i.e. 1st quartile, termed "minimum"); and the 75th percentile (i.e. 3rd quartile, termed "maximum").

Once the user's AMR data has been entered into the User Input tab (pasted as values), they will be highlighted depending on a set of conditions. Before any data is entered, all the cells will be green. There are 3 conditions in this tab:
- **Red** - cells that are highlighted red are those where the consumption for that half hour is **greater than** the upper limit of the profile bandwidth (see Charts). This means that the consumption is above the 75th percentile, or maximum of the range, of the public sector buildings in the sample.
- **Orange** - cells that are highlighted orange are those where the consumption for that half hour is **greater than** the profile of the profile bandwidth but **less than** its upper limit (see Charts). This means that the consumption is above the 50th percentile, or median of the range, but below the 75th percentile of the public sector buildings in the sample.
- **Green** - cells that are highlighted green are those where the consumption for that half hour is **less than** the profile of the profile bandwidth (see Charts). This means that the consumption is below the 50th percentile, or median of the range, of the public sector buildings in the sample.

Given the amount of data in this tab, it is necessary to zoom out to 10% (depending on your screen size) to get an appreciation of the heat map from an annual perspective. Zooming out will allow you to see "hotspots" in your building's consumption compared to the sample profile.

Description of the Chart tab:
As described in the section above, the charts presented in this tab show three profiles:
- Profile bandwidth (orange area) - this visualises the upper and lower bands of the public sector AMR data for the given dataset, i.e. the upper band is the 75th percentile, and the lower band is the 25th percentile. This represents the benchmark range that was derived from the AMR sample data.
- Profile (dotted orange line) - this plots the 50th percentile, i.e. median, of the public sector AMR data for the given dataset and is considered the typical consumption for the building type.
- User Input (solid blue line) - this plots the total consumption at the 30-minute intervals for the user's building, using the data entered in the User Input tab.

There are 12 charts, broken down into the months of the year, and each chart plots the days of the week so that users can compared their data on a weekly and monthly basis. To create these charts, every weekday for each month is averaged and then plotted. For example, in the first chart (January), the three profiles are derived from the average consumption data of all the Mondays, Tuesdays, Wednesdays, etc., in January for the 75th and 25th percentiles (to create the profile bandwidth), the 50th percentile (to create the typical profile), and the user's building's AMR data (to create the user input). Therefore, as the profile bandwidth and typical profile are derived from the average percentile consumption for the occurrence of that day in a given month, they will present a smoother profile than the user data, but they should be a similar shape and magnitude.

Figure 2: About page of the AMR profile, for the example of open plan offices, gas
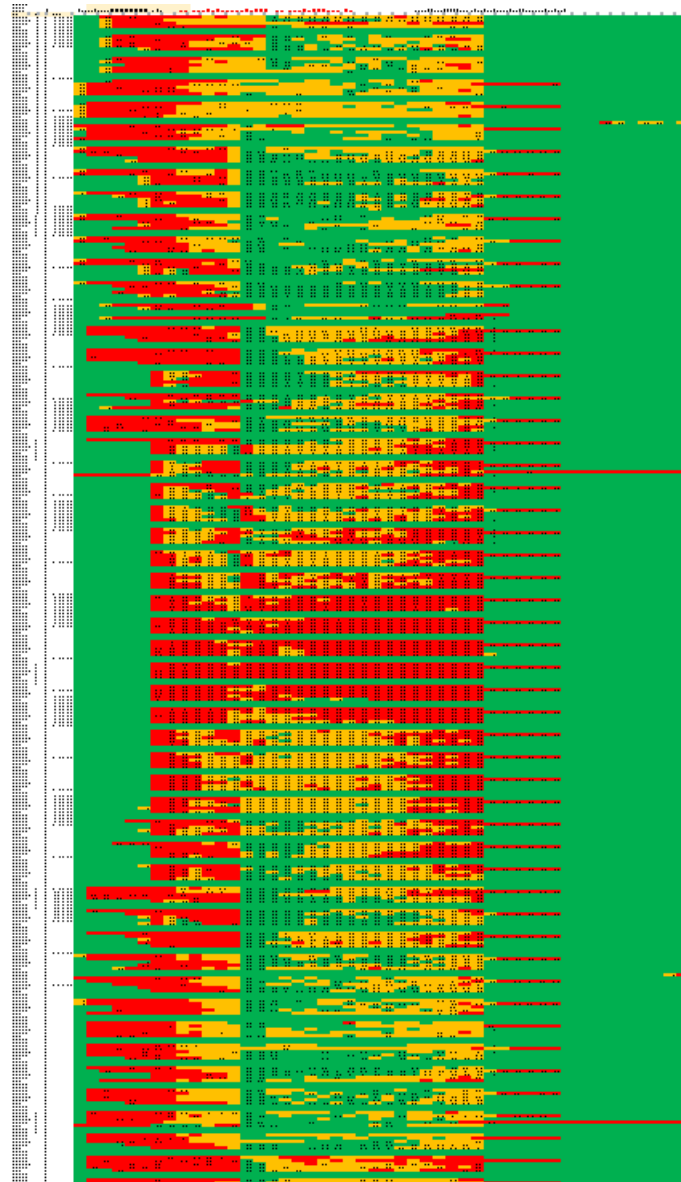
Figure 3: Example the heatmap visualisation of the user's AMR data, in relation to the statistical profile
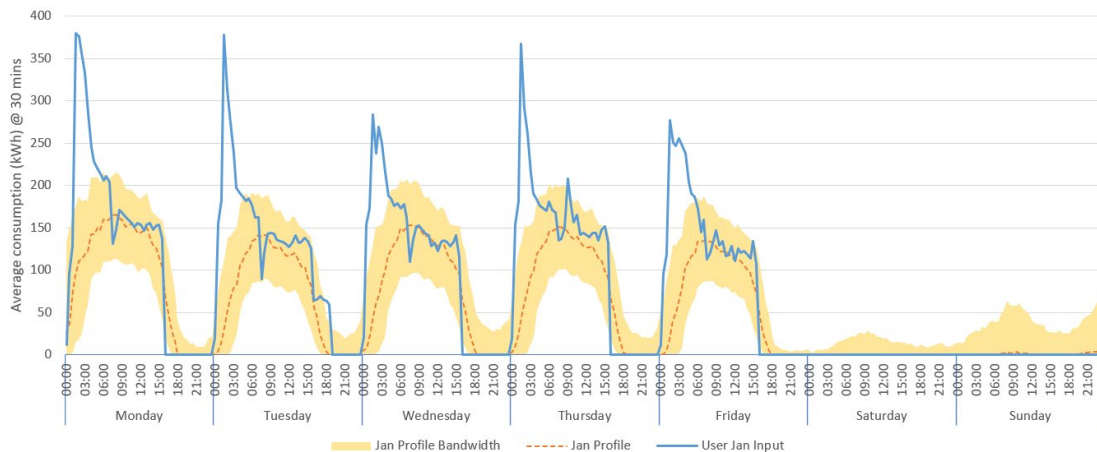
Figure 4: Example of the monthly charts, displaying the user's AMR data compared to a profile bandwidth and 'typical' profile

The SPSEBT 2022 introduces a 'Benchmark Bandwidth'; which is present in both Part 1 ('build you own benchmark' browser platform) and Part 2 (AMR Profiler) of the SPSEBT. The benchmark bandwidth presents the user with 3 values, referred to as minimum, typical, and maximum, expanding on the concept of typical and best practice values. The benchmark bandwidth presents a benchmark range (max to min) and an absolute benchmark value (typical); when comparing a value to this bandwidth, this must be interpreted as:

- Above the profile = the comparing value falls between the typical and maximum benchmark value
- Below the profile (or best practice range) = the comparing value falls between the typical and minimum benchmark value
- Out of profile (excessive) = the comparing values falls above the maximum benchmark value.

At this point, no definition is given to a condition where the comparing values fall below the minimum value. In practice, a building's energy, water or carbon value can fall anywhere within or outwith the benchmark bandwidth, but whilst a comparing value that falls above the maximum benchmark can be referred to as excessive, it may be inappropriate to refer to a building as 'better than best practice' if that building's value falls below the minimum benchmark value. Theoretically, there are instances when a building's consumption and emissions are 'too low' thereby signifying a fault with an energy system or providing an uncomfortable or unhealthy indoor environment. However, if a buildings consumption or emission value falls below the benchmark minimum, and the user understands that the building is operating effectively at the necessary capacity, then this could be interpreted as 'better than best practice'.

2 Methods and Results overview

The SPSEBT 2022, funded by Scottish Government, is the successor of the Scottish Energy Officers Network's (SEONs) Energy Benchmarking Exercise, also referred to as the Public Sector Benchmarking Tool (PSBT) Phase 1. The SPSEBT 2022 supersedes the PSBT Phase 1 by:

- Including a wider range of public sector organisations, beyond local authorities
- Expanding and redefining the Primary Building Categories, into 63 Building Activity Types. These BATs harmonise more closely with the National Calculation Methods (NCM) Simplified Building Energy Model (SBEM) building databases, Energy Performance Certificates building types, and other Government tools, e.g. the Energy Heat Map, EPC non-domestic register, etc.
- Includes benchmark values for operational carbon dioxide emissions
- Ensuring data anonymity for all buildings and organisations used to create the benchmarks, presented in an accessible, secure, online browser-based platform
- Harmonises the benchmarks for heat consumption to include electricity
- Expanding filters to include retrofit, HVAC types, BEMS, and heat distribution types
- Increasing usability and user experience through an optimised set of data visualisation methods
- Providing an AMR profile for utility data relating to electricity (non-heat), gas, and water consumption.

The following 7 sub-sections concisely describe the methodology, results, and relevant information, pertinent to the delivery of Part 1 and Part 2 for the SPSEBT 2022. A library of technical papers, spreadsheets, instructions, and other documentation is available for each of the sub-sections, each providing significantly more detailed information than can be presented here. Due to data security, anonymity, and data protection policies, some of these documents may be inaccessible to the general public.

3 Collecting the data

Collecting the necessary breadth and depth of data was a collaborative exercise, which began with a stakeholder engagement and recruitment event in Summer 2021. This event provided a list of 109 expressions of interest from 156 organisations. From the contact details of 173 interested individuals a target list of 150 public sector organisation was created. These 150 organisations received the first official email survey to build both Part 1 and Part 2 of the SPSEBT. Data capture activities began in October 2021 and continued through to November 2021. Significant support was provided by SEON, Zero Waste Scotland, Sustainable Scotland Network (SSN), NHS Health Facilities Scotland, Environmental Association of Universities and Colleges, and Community Leisure UK for access to specific sectoral groups. Data was captured through an email survey, which contained two parts:

1. Questionnaire
2. Letter of Authority.

The questionnaire took the form of 34 data entry fields built and circulated in the Microsoft Excel spreadsheet. The Excel questionnaire was structured over 5 steps, each step built in its own tab (worksheet), structured as:

- **Step 1_Read me first:** This tab provided an overview of the project, the spreadsheet, and the survey.
- **Step 2_Survey guidance:** The survey's questions were listed here, along with a description for interpreting each question, and our intended use for the values provided.
- **Step 3_Example survey:** Here, 5 examples were provided for context and further guidance to interpreting the questions and how respondents should add their own values in Step 4.
- **Step 4_Your answers:** This is where respondents entered the information about their buildings.

- **Step 5_Final check:** This the last tab, and a place to review a summary of the information entered into Step 4 and to check for errors before returning the survey.

Appendix 2 presents the list of fields (coloured blue) that created the data/information input structure to Step 4. This presented 34 data entry fields, grouped and designed to capture specific information that would underpin the statistical benchmarks and the filtering options in the SPSEBT:

- 3 of these were single entry cells that related to the respondent's entire portfolio of buildings.
- 17 related to each buildings function and form.
- 11 related to the energy and water consumed (and possibly generated) by each building. All 11 were not necessarily applicable to all buildings.
- 2 related to each building's Energy Performance Certificate (EPC) scores.
- 1 allowed respondents to enter any comments they believed were relevant when it came to analysing the information.

Six versions of this questionnaire were built and made available to those who were engaged for participation:

1. A semi restricted version, where dropdown menus provided options for certain entry fields
2. an unrestricted version, with the dropdown menus removed and left as an open input field.

Both of these versions were available for 3 editions of Excel:

1. Built to Excel 365 .xlsx
2. Built to Excel 2007 .xls
3. Built to Excel 97-2003 .xls

These versions of Excel handle some functions differently, i.e. vlookup and dropdown options, but provided the most flexibility for potential participants to use, in relation to their organisations computer operating systems, and edition of Microsoft Office. *Figure 5* shows the layout of Step 1 and Step 5 of the questionnaire spreadsheet.

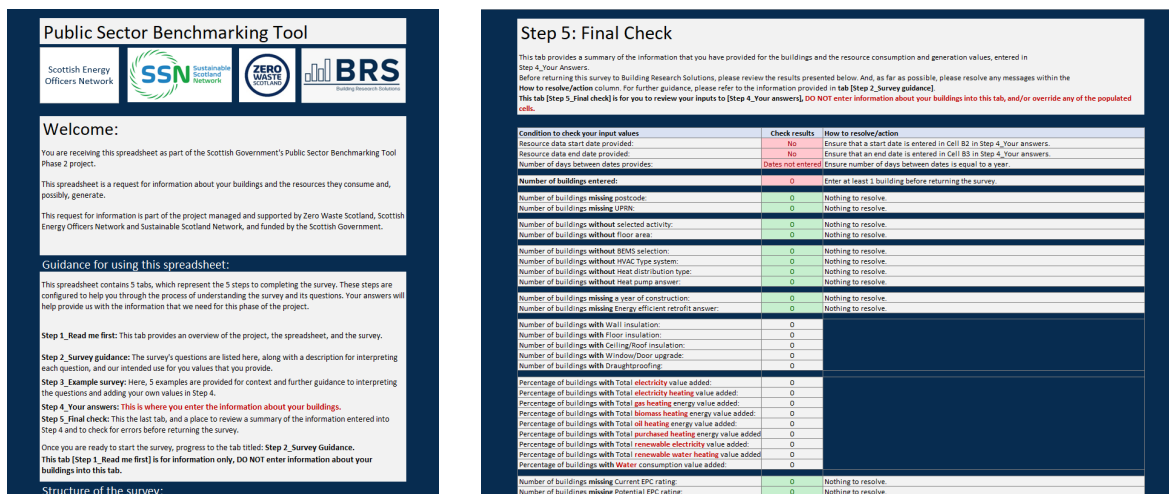**Figure 5:** Screenshots from the participant questionnaire, showing Step 1 and Step 5

Collaboration with NHS Health Facilities Scotland provided access to an existing library of electricity, heating, water, and carbon emissions data, meaning that the questionnaire part of the survey was streamlined to include only entry fields pertaining to a building's physical characteristics, therefore, only these fields were included:

- 17 related to each buildings function and form.
- 2 related to each building's Energy Performance Certificate (EPC) scores.
- 1 allowed respondents to enter any comments they believed were relevant when it came to analysing the information.

The second part of the survey consisted of a predefined letter, containing the necessary instructions for the three utility companies (EDF, Total Energies, and BusinessStream) to provide the project team with AMR data. This Letter of Authority included wording, approved by each Utility company, and required the participate to include their signature and organisation's logo (or office standard header).

Data capture activities concluded in mid-January 2022; questionnaire and LoA return times were relaxed to account for the November mandatory climate change reporting deadline (coordinated by SSN), COP26, COVID, and absence of key staff. The survey was completed and returned by 78 Scottish public sector organisations across 7 sectoral groups, as shown in Figure 6. Information for a total of 7243 sites was returned; after checking, processing, and conditioning these building results, 6331 buildings were suitable to be used to power the SPSEBT.
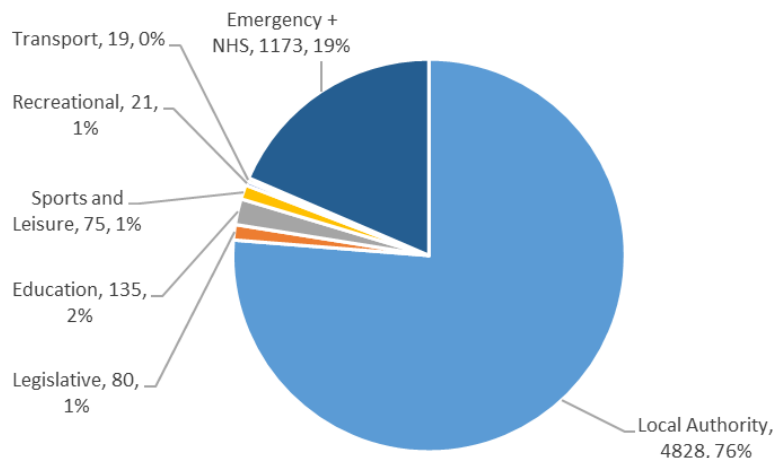
Figure 6: Percentage split, number of buildings per sectoral group

4 Processing the database

During December 2021, the LoA part of the surveys were forwarded to three participating utility companies, in anticipation of receiving the necessary AMR data to deliver Part 2 of the SPSEBT: the AMR Profiler. All completed and returned surveys followed this quality control process:

Stage 1: Validation – using Step 5 of the returned questionnaire to validate that each building (row in the Excel questionnaire) had the correct and necessary building characteristic information to underpin the understanding of the energy and water consumption data.

Stage 2: Verification – reviewing each value in each cell for each building, checking that the information was accurate. This was a three-step process:

1. Searching the sheet for comments, comment balloons, hidden cells, hidden tabs, etc.
2. For any values that provided any uncertainty, direct interaction with the participant clarified and/or corrected these values, e.g. presence of a heat pump, EPC score, data year, etc.
3. Removal or correction of anomalous entries and values that would disrupt Stage 3.

Stage 3: Conditioning – combining all validated and verified questionnaires into one database. The database was processed to ensure that all values were reporting comparable information, both per building and per data energy field. This included comparing the building's naming convention to the selected Building Activity Type, creating new BATs, and removing unused BATs, etc.

Stage 4: Checking variables – checking variables were created to ensure that a building's entry would not skew or disrupt the benchmark values for its designated BAT, this mostly focused on buildings with energy values only for electricity, and others with certain HVAC types.

Stage 5: Calculations for benchmarking – additional fields were added to the database which allowed for the calculation of climate change impact scores (carbon dioxide emissions), and normalised values.

During these quality control checks, the database was rebuilt an additional 4 times, and enhanced between 3 and 8 times per build, where the final version of the database is referred to as Build05v05. This process created two databases:

1. Master Database, which carries the buildings to be used in the Benchmarking and AMR Profiler Tools, see Appendix 2 for the full list of fields (column headers) used in the Master Database.
2. Reject list, which carries the buildings that were excluded from further analysis.

The Rejects list consisted of buildings with either:
- Unusual consumption values (n=57), where the building presented no or unusually low normalised consumption values for energy, water, and carbon
- Unusual floor area (n=14), where the building presented with floor areas that could not be verified by the participant or satellite map measurements, which resulted in excessive energy consumption
- No floor area (n=731), where the participant or satellite map measurements could not define a definitive floor area
- This reject list would later be expanded to include those buildings that would significantly skew the percentile ranges during the benchmarking analysis (see Section 2.3).

The initial list of Sub-BATs was expanded from 58 to 67, which would later become the final 63. These Sub BATs can be clustered into 10 Parent BATs and Figure 7 presents the percentage split and number of buildings which were clustered into each Parent BAT. Number of buildings per Sub-BAT are not presented here, however, the average (mean) number of buildings per Sub BAT was 100, 19 Sub-BATs had more than 100 buildings, 44-Sub BATs had less than 100 buildings. The Primary School Sub-BAT has the most buildings (n=1431), National Health Laundries buildings had the least (n=2). Figure 8 presents the distribution for the number of Sub BATs with a similar number of buildings (bin size 20 buildings).
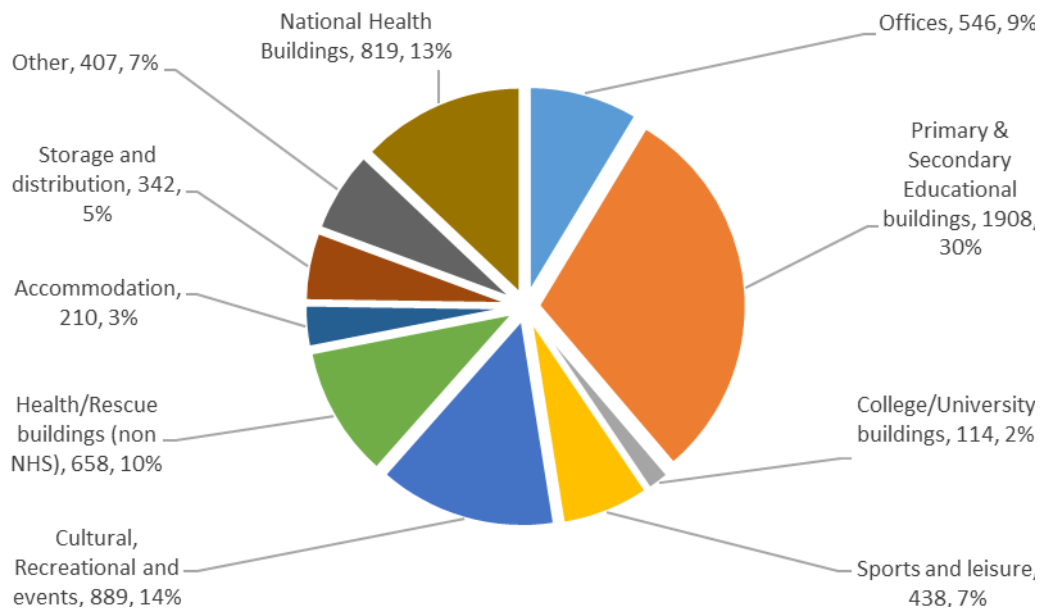
Figure 7: Percentage distribution of buildings, grouped into Parent BATs
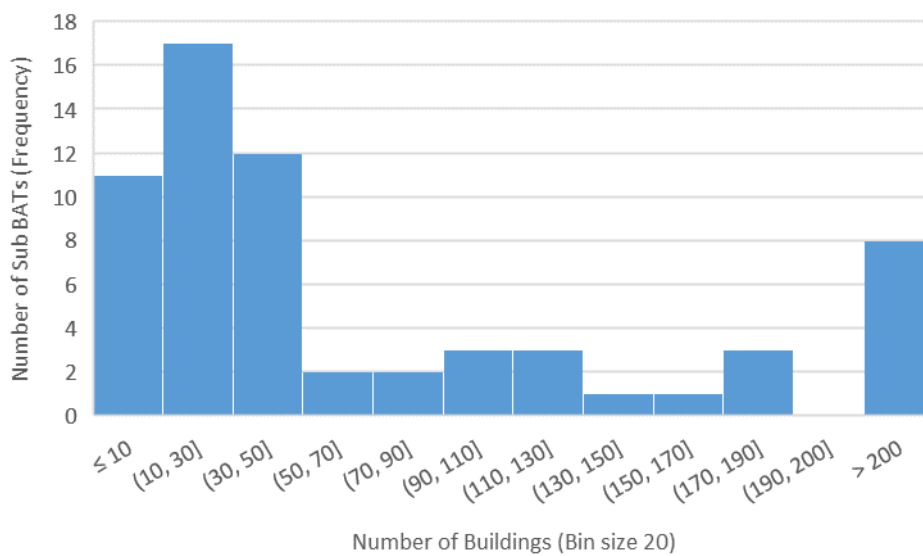
Figure 8: Distribution of number of Sub-BATs with a similar number of buildings

4.1 Creating benchmarks

Upon completion of the master database, each of the Sub-BATs were isolated and analysed independently. In keeping with other Scottish Government decision support tools, e.g. Scottish Index of Multiple Deprivation, Scottish Heat Map, etc., Part 1 of the SPSEBT (the online dashboard) is based on the statistical principles of data distribution and percentile distributions. This principle was applied to the SPSEBT values over a series of stages:

Stage 1: benchmark range. Within each Sub-BAT, buildings with skewing and potentially skewing values were removed using a novel method that used the calculated normalised carbon dioxide emission values, which were generated from the climate impact from the building's electricity, heating fuel type, and water consumption values, to define the benchmark range. This range was reviewed

using values for skewness and kurtosis, which are numerical values to describe the shape of data distribution within a given set of numbers. Tables, at 1 percentile increments, were generated for the normalised electricity (non-heat), heating, water, and carbon dioxide emissions data ranges. At this stage, where a building presented a value or values that significantly skewed the range, in either direction, they were removed from the benchmark analysis and added to the reject list (n=73), this was the final point within the whole analysis procedure when a building could be removed and added to the reject list. Whilst site visits and deeper investigation of a buildings consumption or emissions was beyond the scope of the SPSEBT 2022 methodology, buildings rejected here were done so for the betterment of the benchmark values.

Stage 2: Exclusion percentile ranges were created to disaggregate the benchmark range. Analysis showed that, consistently, buildings whose values fell within the top 5 percentile (96th to 100th) were identified as potentially up-skewing the benchmark values for a Sub-BAT. Furthermore, where the percentile range and data distribution showed a skew towards the lower (smaller) values, the first 20 percentiles (0 to 19th) were identified as potentially down-skewing the benchmark values for a Sub BAT. The same was identified for Sub-BATs where their percentile range and data distribution showed a skew towards the upper (large) values, in this case the first 50 percentiles (0 to 49th) where identified as potentially down-skewing the benchmark values for a Sub- BAT.

However, as individual building surveys were outwith the scope, it is unclear as to why that building's consumption or emission values were so high or low, therefore they were not rejected. Rather, those buildings that fell outwith the exclusion range (20th or 50th to 95th percentiles) were grouped with a skewness potential value (SPV) of 1, and others within the exclusion range (0-19th, or 0-49th, and 96th to 100th) were grouped with a SPV of 2.

For Part 1 of the SPSEBT (online dashboard) and the definitive benchmark table (available in Appendix 1 of this document or as an export from the SPSEBT platform), this SPV is applied in three ways:

1. **On Level 1 of the SPSEBT dashboard (Main level) -** Level 1 shows the benchmark values calculated from those buildings grouped with a SPV of 1, for the 10 parent BATs.
2. **On Level 2 of the SPSEBT dashboard (Detailed data)** - Level 2 shows the benchmark values calculated from buildings with a SPV of 1 and 2 (all buildings in the range), for the 63 Sub-BATs. SPV 1 and 2 are included in the Detailed data because the buildings with a SPV of 2 may provide valuable data for a range under certain conditions. Meaning, when enough filters are applied on Level 2, the wider range of buildings presents the opportunity for the benchmark results to better represent the parameters of the real-world building - whose value(s) are being compared to the benchmark.
3. **Definitive Sub-BAT benchmarks table** – (available in Appendix 1 of this document or as an export from the SPSEBT platform) the table shows benchmark values for buildings with a SPV of 1, for the 63 Sub-BATs. Furthermore, the values here are calculated from the statistically generated percentile tables. As no filters can be applied to this static table, presenting a value calculated from buildings with an SPV of 1 provides the most robust, statistical benchmark bandwidth range.

The difference in benchmark values between Level 2 of the SPSEBT platform and the definitive table are small. However, when referencing a Sub-BAT benchmark value with no filters applied, **you must report those from the definitive table.**

Stage 3: Benchmark Bandwidth. The SPSEBT provides a benchmark for electricity (non-heat), heating, water, and associated carbon dioxide emissions, for 10 Parent BATs and 63 Sub-BATs. The SPSEBT 2022 expands on the PSBT Phase 1 with the introduction of a 'Benchmark Bandwidth', which presents the user with 3 values, referred to as minimum, typical, and maximum, expanding on the PSBT Phase 1's typical and best practice values:

- **Typical** – is the 50th percentile of a given data range – i.e. the median or second quartile
- **Minimum [min]** – is the 25th percentile of a given data range – i.e. the first quartile. This is not the same as the 'range minimum', where the smallest value from a range would be presented

- **Maximum [max]** – is the 75th percentile of a given data range – i.e. the third quartile. This is not the same as the 'range maximum, where the largest value from a range would be presented.

The benchmark bandwidth presents a benchmark range (max to min) and an absolute benchmark value (typical) somewhere within that range. When comparing a value to this bandwidth, this must be interpreted as:

- Above the profile = the comparing value falls between the typical and maximum benchmark value
- Below the profile (or best practice range) = the comparing value falls between the typical and minimum benchmark value
- Out of profile (excessive) = the comparing values falls above the maximum benchmark value
- At this time, no definition is given for a building's performance where their value falls below the minimum benchmark value.

4.2 Developing the User Interface

A user-centred design method underpins the front-end user interface (UI) of the SPSEBT (Part 1 dashboard). This means that, from the initial design through to the implementation of the UI, a representative of the target end user group was involved with reviewing each iteration of the UI's evolution. This holistic and scientifically verified approach ensured that the SPSEBT Part 1's UI fulfilled the needs of the end users, and that the UI is accessible to a wider range of users (skills, intentions, expectations), upholding the principles of logical design and perspicuity. The development of the SPSEBT Part 1's UI took a 2-phase approach:

- Phase 1 – UI design, with 7 stages
- Phase 2 – UI implementation, with 12 design elements (10 dashboard pages and 2 design features).

Phase 1

Stage 1: PACT Analysis – This provided the SPSEBT user experience (UX) designers with a wide spectrum specification list, constructed to ensure that the variety of requirements of end users are centric throughout the design stage, through to full implementation and handover. The analysis focuses on four main aspects of any complex, interactive system: People, Activities, Context, and Technology (PACT).

Stage 2: Heuristic evaluation – A top-down and bottom-up review of the PSBT Phase 1 benchmarking tool, plus a courtesy interaction with other Scottish Government tools provided the definitive list of features and functions that the design of the SPSEBT online platform could employ to enhance the user's experience.

Stage 3: Personas and Scenarios – 5 personas and 15 scenarios were created, designed around the interaction, requirements, and expectations of the primary, secondary, and tertiary users. This provided the guiding principles to enable the delivery of the SPSEBT UI.

Stage 4: Paper prototyping – This delivered an initial outline sketch, including the critical information that the UI should display, without the confines of a digital environment, now referred to as the dashboard. This stage set the visual structure and layout of necessary information on the dashboard, including intended navigation pathways, and methods to reduce user misinterpretation of the presented information.

Stage 5: Mid-fidelity prototyping – This stage transferred the paper dashboard onto an interactive digital platform, but not the final designation of the dashboard. This delivered an optimised layout, ensuring that sizes and positioning of each item is calibrated to streamline the user's journey from arriving on the dashboard, to gathering the information that they came to find.

Stage 6: Usability testing and focus groups – This stage captured feedback on the designs from a group representing the target end users of the SPSBET platform. This acted to reinforce and validate earlier design decisions and provided a discrete list of modifications to optimise the experience for certain user types.

Stage 7: User testing and final feedback sessions – At strategic points during Phase 2: UI implementation (described below), a semi or fully functional version of the dashboard (connected to one of the builds and versions of the master database) was tested with a group representing the target end users of the SPSBET platform. At this stage, a 'near-complete' version of the dashboard was used. This was done using the scenarios defined in Phase 1: Stage 3. Here, the User Experience Questionnaire showed that the SPSEBT dashboard design scored positively across all 6 standardised tests. The dashboard scored above average on Perspicuity (i.e. how easy it is to learn to use), Efficiency (i.e. the tasks the users need to carry out can be done without unnecessary effort), and Stimulation (i.e. how exciting and motivating it is to use the product). The SPSEBT scored even better on Dependability (i.e. do the users feel in control of the interaction) and Novelty (i.e. how innovative and creative the tool is and how interesting it is to the users). Finally, the SPSEBT scored the highest on Attractiveness (i.e. the overall impression and if the users like or dislike it), putting it in the top 10% of interactive products for overall satisfaction.

Phase 2

Design elements 1 to 12: Implementation – Microsoft's PowerBi data visualisation and data dashboard was selected as the most appropriate ecosystem to contain and carry Part 1 of the SPSEBT. This allowed for:

- Enhanced interoperability with Excel, the ecosystem that contains the master database
- Improved capacity to update and publish changes to the benchmarks
- Optimised handover between project collaborators, and those who would host the final version
- Removal of the requirement for the end user to install additional software, enable Macros, install Plug-ins, etc, bypassing the potential restrictions caused by IT security policies, employee permissions, and firewalls for some organisations
- Securing the raw data from access by the end user.

Within other supporting documentation, detailed design implementation information is provided for each of the charts, tables, and other visualisations for each of the SPSEBT Part 1's 10 pages:

- Level 1, and 2: Home page
- Level 1, and 2: Carbon page
- Level 1, and 2: Electricity page
- Level 1, and 2: Heating page
- Level 1, and 2: Water page

Including descriptions of the selected colour palette and the other visualisation functions common across all areas of the SPSEBT platform, e.g. filters, pop-up 'Tooltip' information balloons.

Finally, the SPSEBT has been tested on different laptops and desktop computers with different screen sizes, different operating systems (i.e. Windows 7, Windows 10, Windows 11, and macOS Monterey (v12.2.1)), and different browsers (i.e. Microsoft Edge (v99.0.1150.39), Opera (v84.0.4316.31), Google Chrome (v99.0.4844.51), Mozilla Firefox (v98.0.2), and Safari (v15.3)). The SPSEBT Part 1 broswer platform is not optimised for, nor was it intended to display effectively on a mobile device – i.e. smart phone, or smart tablet device.

4.3 Integration of the Database into PowerBi

Build05 version05 (v05) of the master database provides all necessary filtering fields (see Appendix 2) and raw data needed to power the SPSEBT dashboard, within the Microsoft PowerBi ecosystem. The process of connecting the database to the dashboard (referred to here as the front-end) required interventions to optimise the SPSEBT's performance and usability. These interventions are referred to as 'back-end' processes. Back-end processes focused primarily on:

Anonymising the data: Ensuring data security and anonymity was one of the top priorities for the SPSEBT platform. This was achieved by removing any column or field from the master database that could be used to directly identify any of the buildings used to generate any results; resulting in 8 of the first 10 fields in Appendix 2 being removed from public access. For the purpose of database management and running another phase of data capture/data updating, each building has a unique SPSEBT ID, allowing the database manager to update values in the SPSEBT platform without compromising the location of any given building.

Maps and locations: Any field that can directly identify any building has been removed from the database that powers PowerBi. The National Record for Scotland's Scottish Postcode Directory (SPD), areas, districts, and council area codes are used to power the map visualisation (Level 1 on the dashboard), see Figure 9. The map in Figure 9 is a copy of the map on the Home page of Level 1; the gradation of blue represents the median normalised carbon dioxide emissions ($kgCO_2e/m^2/year$) from the buildings in that area.

Figure 9: Interactive map of Scotland included in the dashboard, divided into geographical areas

Interoperability: To reduce loading times and streamline dashboard management, many of the charts and other visualisations are built using standard or default Microsoft components. Due to the statistical nature of creating the benchmarks (see Section 2.3), additional back-end formulae were created to provide the necessary values for charts and tables that show max, typical, and min values.

Export report: The Export Report button, bottom left-hand corner of the SPSEBT dashboard, delivers a Microsoft Excel table, see Figure 10. This table is referred to as the definitive Sub-BAT benchmarks table, also available in Appendix 1 of this document. This is the table that should be referenced if taking benchmark values from any of the unfiltered 63 Sub-BATs. At this time, the SPSEBT platform is unable to export an Excel sheet with data specific to the filters applied by the user. The current downloadable spreadsheet is independent of the master database that powers the SPSEBT dashboard's visualisations. For clarity, the values in this table are calculated from the buildings grouped with the skewness potential value of 1 and calculated from statistically generated percentile tables.

Parent Building Activity Type	Sub Building Activity Type	Carbon dioxide emissons (all energy + water) kgCO2e/year/m2			Electricity consumption (non-heat) kWh/year/m2			Heat consumption kWh/year/m2			Water consumption m3/year		
		Min	Typical	Max	Min	Typical	Max	Min	Typical	Max	Min	Typical	Max
Offices	Offices - generic	34	45	64	47	65	99	103	137	207	147	382	1031
Offices	Offices open plan	34	47	60	60	76	109	104	144	189	310	845	1947
Offices	Offices cellular	42	50	62	44	57	71	146	177	207	120	245	549
Primary & Secondary Educational buildings	Primary & Secondary Educational buildings - generic	29	34	44	37	51	69	63	113	151	576	1560	3191
Primary & Secondary Educational buildings	Kindergarten/Nursery	43	53	67	46	64	93	141	190	243	353	473	715
Primary & Secondary Educational buildings	Primary school	37	45	55	41	52	68	123	149	184	851	1397	2266
Primary & Secondary Educational buildings	Primary school with Pool	47	55	82	55	57	57	182	215	257	1692	2560	2727
Primary & Secondary Educational buildings	Secondary school	32	38	44	43	49	59	105	130	155	2968	4171	6553
Primary & Secondary Educational buildings	Secondary school with pool	40	52	62	54	59	71	152	194	226	4647	6035	8140
Primary & Secondary Educational buildings	School special	45	57	72	57	65	75	140	173	253	1181	1872	2647
College/University buildings	College/University buildings - generic	39	50	61	63	77	108	124	160	200	1974	4011	8065
College/University buildings	Teaching College/University (class rooms, lecturing spaces, arts)	43	62	76	25	52	89	173	193	297	136	1350	3119
College/University buildings	Special facilities College/University - high consumption, labs, etc.	164	186	195	166	256	279	490	644	701	1087	1468	2664
Sports and leisure	Sports and leisure - generic	45	61	83	94	121	166	192	234	325	751	1364	2598
Sports and leisure	Leisure centre - Indoor dry sports	59	65	76	93	106	133	212	247	288	1297	1630	3072
Sports and leisure	Leisure centre - Indoor dry sports and swimming pool	142	189	237	152	177	236	492	648	899	5568	8080	11216
Sports and leisure	Leisure centre - Indoor and external dry sports with external lighting	71	90	105	87	99	119	239	316	372	303	1235	2659
Sports and leisure	Sports and leisure - golf course	53	61	62	87	92	100	192	224	277	1305	1851	2734
Sports and leisure	Sports ground changing facility >500 CO2	2144	2264	2440	3476	3929	4439	6314	6987	7722	1098	1478	1498
Sports and leisure	Sports ground changing facility	27	36	54	75	90	131	134	172	291	369	520	878
Cultural, Recreational and events	Recreational and events - generic	42	56	76	65	84	157	101	167	254	635	1082	3058
Cultural, Recreational and events	Libraries	44	56	71	49	62	88	145	195	269	39	88	279

Figure 10: Excerpt from the benchmark table that can be exported from the SPSEBT dashboard, using the Export button

4.4 AMR Conditioning and processing

Sections 2.6 and 2.7 relate to Part 2 of the SPSEBT – the AMR Profilers, which were created in the Microsoft Excel ecosystem. Each utility has its own Profiler and, for this proof-of-concept stage, there is a Profiler for specific Sub-BATs (described below).

Signed and returned Letters of Authority (LoAs) from each participating organisation were forwarded to the three engaged utility companies (EDF, Total Energies, and BusinessStream). The LoAs provided access to automated meter reading (AMR) data, with the following limitations:

- BusinessStream were unable to provide data before January 2021
- Only Platinum Total Energies accounts provided half hourly (HH) AMR data
- Organisations with a larger portfolio could only be accessed through EDF's MyBusiness portal. Access to this portal was not granted during the life of the project to contribute to the AMR Profiler
- Some of the participating organisations were not supplied by one or more of these three utility companies (not on the framework).

The format of the AMR data differed across the three utilities, thus it had to be handled, conditioned, and processed before it could be used to develop the Profilers (discussed in Section 2.7). This process followed the same 5 stages for each utility:

Stage 1: Creating an AMR file per organisation – the AMR dataset arrived as 1 spreadsheet file per building, per organisation. The handling stage checked each dataset for high level anomalies, i.e. date range, daily profile values, and excessive 0s or no data during peak operating times. Those that passed the validation check were added to a master AMR database, created per organisation.

Stage 2: Aligning meter numbers with addresses – the utility company provided a meter number and building name or address for that meter number. For instances where these were in a separate codex or available in each dataset, they were combined as a contents page for each organisation's master AMR database.

Stage 3: Aligning AMR addresses with the unique ID in the master database – each building within an organisation's content page, for their master AMR database, was aligned to the master database described in Section 2.3. This resulted in each AMR dataset being assigned one of the unique IDs from the master database. This was the conditioning stage that removed AMR datasets that could not be aligned with a building in the master database. Every reasonably practicable attempt was made to align an AMR dataset with a unique ID, this included, but was not limited to:

- Search by building name, street name, and postcode
- Using alternative, abbreviated, and/or incorrect (commonly misspelt) words
- Using acronyms
- Using English and Gaelic spellings
- Using district and area level postcodes
- Using alternative postcodes and street names acquired from web searches

Unaligned AMR datasets were rejected only after multiple attempts were made to align it with a unique ID. The reason for aligning with the master database was to ensure that the AMR data offered accurate and realistic consumption profiles – through this alignment, the annual consumption (of electricity (non-heat), gas, and water) reported by the respondents of the survey was compared against the annual consumption from AMR to make sure they were within reasonable limits of each other.

Stage 4: Filtering out AMR datasets – those AMR datasets that achieved positive alignment with a unique master database ID were furnished with additional building details, brought in from the master dataset; this included verification checks, such as the name of the owning organisation, Sub-BAT, and the reported annual energy or water consumption. The annual energy consumption, for 2019 for energy and 2021 for water, was calculated from each AMR dataset, and compared to the reported annual energy consumption value in the master database. A tolerance of +/-25% difference was accepted, meaning that AMR datasets with an annual consumption in 2019 greater than 25% or less than -25% of the participants' reported value were rejected from proceeding into the AMR Profiler. This percentage difference verification check was not completed for the water AMR data as 2019 data was being compared against 2021 data, and there was no guarantee that the differences weren't due to the differing years.

Stage 5: Creating a Sub-BAT AMR library – the AMR Profiler was built per Sub-BAT, many of the organisations' master AMR databases carried a variety of AMR datasets for a range of different Sub-BATs. At this point, a Sub-BAT AMR library was created for Primary Schools and Open Plan Offices, due to the greatest availability of data and difference in function. Primary Schools has an AMR library for electricity, gas and water, while Open Plan Offices has an AMR library for electricity and gas only. Table 1 shows the number of usable AMR datasets available for each Sub-BAT, for each utility. However, due to the scope of the project and the proof-of-concept nature of this AMR work, it was not possible to create an AMR Profiler for every Sub-BAT. Note that NHS Health Facilities Scotland also provided AMR data for gas but given the time constraints, we were unable to condition it, thus NHS buildings are excluded from Table 1.

Table 1: List of usable AMR datasets available for each Sub-BAT, for each utility

Sub-BAT	Total Energies	EDF	BusinessStream
Primary school	377	105	120
Secondary school	72	54	40
Community centre	64	14	6
Care homes	45	24	12
Libraries	39	11	2
Day centre	32	5	6
Offices - generic	28	37	9
Offices open plan	26	26	6
Offices cellular	25	12	0
Sheltered housing	20	1	2
Kindergarten/Day centre/Nursery	18	3	13
Secondary school with pool	17	8	8
Leisure centre - Indoor dry sports and swimming pool	15	25	15
Depot	12	14	9
School special	11	1	1
Sports and leisure - generic	11	4	15
Venue - Conference centre, Concert Hall and Theatre	11	5	0
Museums and Galleries	8	6	1
College/University buildings - generic	7	15	6
Leisure centre - Indoor dry sports	6	6	1
Sports ground changing facility	6	5	5
Hotels/hostels	4	3	1
Primary school with Pool	3	3	0
Special facilities	3	2	0
Visitor/information centre	3	2	0
Data Centre/Servers	2	2	0
Primary & Secondary Educational buildings - generic	2	9	3
Accommodation - generic	1	1	0
Leisure centre - Indoor and external dry sports with external lighting	1	6	3
Outdoor events, parks and bothy	1	1	0
Passenger terminals	1	2	0
Public toilet	1	1	2
Recycling, waste, landfill	1	6	1
Teaching (classrooms, lecturing spaces, arts)	1	4	3
Car/Vehicle Parking	0	1	0
Cemetery and religious	0	1	10
Cemetery and religious >100 CO_2	0	1	1
Food and drink >100 CO_2	0	1	1
Industrial-type activities	0	3	1
Industrial-type activities >100 CO_2	0	2	0
Storage (conditioned)	0	2	1
Storage (unconditioned)	0	1	0
Student accommodation (catered)	0	1	0
Student accommodation (self-catered)	0	2	0

4.5 Creating the AMR Profiles

After conditioning the AMR data so that all the utilities are in a standardised format, the AMR Profilers could be created using the Sub-BAT libraries created for gas and electricity for the Open Plan Offices Sub-BAT and gas, electricity, and water for the Primary School Sub-BAT. The AMR Profilers were designed and operate within Microsoft Excel, it relies on standard (no Macro or VBA) cell formulas to reduce compatibility issues and reduced processing time and bugs. The AMR Profiler method is a direct follow on from the stages described in Section 2.6, the stages described below are largely consistent for each utility type and Sub-BAT, albeit water AMR is in 15-minute increments while gas and electricity is half-hourly. The stages below briefly outline how the conditioned data was used to generate the AMR Profilers, to deliver Part 2 of the SPSEBT:

Stage 1: Filtering and checks - Using the libraries created in Stage 5 of the AMR conditioning process, the AMR data for each building within the Sub-BAT had to be merged into a single tab (worksheet) to enable statistical analyses. The data then had to be filtered to make sure that no anomalous buildings (over and above the ±25% check done in Section 2.5) were included in the final profiles. The checks done here included:

- Verifying that the normalised annual AMR consumption is in alignment with the normalised consumption in the master database – for gas data, if this was not the case, then heating demand was being supplemented by a source other than gas, therefore the gas AMR dataset would not be fully representative of the building's heating demand. These buildings were rejected as they would skew the profile. For electricity data, this verification was 1 of 2 checks to exclude electrically heated buildings from the profile.
- Alignment with the master database to exclude buildings grouped with the skewness potential value (SPV) of 2 (defined in Section 2.3) from the entering profile.
- For electricity, the second of 2 checks excluded buildings that were known, from the master database, to be electrically heated.

These checks resulted in the final filtered group of buildings to be included in the AMR Profile.

Stage 2: Percentiles and benchmark bandwidth - Minimum, maximum benchmark bandwidth and the typical benchmark were generated using the same 25th, 50th and 75th percentiles principles as used for SPSEBT Part 1. For optimal heat map visualisation these values were calculated for each day of the week and each month of the year, for each time stamp. For optimal chart visualisation these values were averaged for the same day within a given month, to generate an average weekly profile per month, e.g. all the Monday readings in January (for each time stamp) were averaged. The 25th and 75th percentiles become the benchmark bandwidth – the range where data falling between these points is "in profile" – and the 50th percentile represents a 'typical' profile (see Figure 4).

Stage 3: Normalisation and scaling – AMR profiling required standardisation to calibrate the profile's shape to the magnitude of the user's energy or water demand. Floor area was selected from a range of standardisation methods, reducing the complexity of the Profiler. The annualised AMR profile was normalised by the average (mean) floor area of the buildings used to create the profile, allowing each of the benchmark values (25th, 50th and 75th) to be standardised for any building of the same Sub-BAT. Direct user input provides the floor area and annual AMR data for the building to be compared to the benchmark. The user's building's floor area provides the scaling value to scale the standardised benchmark profile to be representative of the user's building.

Stage 4: AMR Profiler visualisations – The AMR Profiler visualises the user's data in two formats:

1. Overlayed onto the averaged-weekly charts, per month for 12 months – the averaged-weekly scaled percentiles are plotted, per month, to generate 12 benchmark bandwidths and typical profile lines. The user's AMR data is processed to calculate a comparative averaged-weekly line, per month and overlain on the monthly benchmark profiles (see Figure 4).

2. Colour coded heat map - Cells within the user's AMR data are conditioned to colour the value compared to where it falls in relation to the profile's benchmark bandwidth. These conditioned cells generate a heat map (see Figure 3). Red cells represent consumption that is greater than the 75th percentile and outwith the benchmark bandwidth (i.e. excessive); orange cells represent consumption that is in the upper range of the bandwidth, between the 50th and 75th percentiles; green cells represent consumption that is below the typical profile (i.e. 50th percentile). Note that, in the green cells, there is no distinction between data that falls within the best practice range (i.e. between the 25th and 50th percentiles) and data that is below the 25th percentile and it may be inappropriate to refer to a building as 'better than best practice'.

5 Lessons Learnt

The methods employed to collect, condition, process, and analyse the data, to create Part 1 and Part 2 of the Scottish Public Sector Energy Benchmarking Tool, can be optimised to enhance the research and development of the next iteration of the SPSEBT. These are detailed in the supporting technical papers, here are the main lessons learnt:

1. **Standardisation of the data collection** – Data requested during the PSBT Phase 1 and the SPSEBT 2022 were somewhat different. The data capture questionnaire for SPSEBT 2022 was highly optimised to collect the minimal required information to produce robust, meaningful, and insightful benchmarks, enhancing the usability of the SPSEBT Part 1 platform. However, requests for certain information within the SPSEBT 2022 questionnaire presented an obstacle to many organisations, especially those with decentralised building information, i.e. utility data held separately to building characteristic information (e.g. building age, floor area, HVAC systems, retrofit level, etc). However, building characteristic information is vital to provide necessary context for the creation of benchmark values, and to enhance filtering options for the evolution of the SPSEBT. At this stage, at least for those organisations that participated in SPSEBT 2022 data collection, public sector organisations now have this information held centrally. Therefore, to streamline future data collection, two approaches are recommended:

   a. The exact same questionnaire survey (used for SPSEBT 2022) is reused for future iterations of the SPSEBT Part 1, or
   b. The building, energy, and water data currently held within the SPSEBT 2022 master database is disaggregated, sorted per organisation, and returned to the participating organisation. Future requests for building and utility data would be matter of reviewing, editing, adding, removing and/or otherwise updating an existing list of buildings, rather than recreating the list.

2. **AMR data capture** – The procedures to handle, condition, process, and analyse AMR data can be resource heavy. It was evident that the framework utility providers required additional support during the AMR data collection phase, which served to restrict the flow of data and decrease necessary analysis time to explore the creation of a wider range of AMR Profilers. For future iterations of the SPSEBT Part 2, either the utility provider must be stakeholders within the project, whereby the quality and delivery of the AMR data is either guaranteed by a certain date or processed and conditioned by their in-house staff. Or a data sharing agreement is setup, between the research team and the utility providers in advance of any data analysis phase, this agreement could take the form of automated data sharing on a weekly or monthly basis. Three items are paramount for future iterations of the SPSEBT Part 2:

   a. The framework providers must have access to data for the same reporting period as requested, from the organisations on the survey questionnaire,
   b. The AMR data must be in a standard time by date format,
   c. Request for AMR held by the organisation should be made to the organisation during the next LoA survey. This would serve to streamline the AMR collection process, reduce dependency on utility companies, and potentially increase the number of AMR datasets, to include organisations who are either not on the same framework as the utility provider, and/or who do not have HH or 15-miinute AMR accounts for some or all buildings.

3. **Harmonising public sector reporting** – For many participating organisations, and those who declared an inability to participate, overlapping reporting commitments was the primary reason provided for delays, partial returns, and drop-outs. It was observed that the Public Bodies Climate Change Duties Annual Reporting Requirements deadline fell over the same period as the SPSEBT

2022 data collection surveys were issued. Future public sector benchmarking data capture activities would benefit from either:

a. Aligning with SSN data capture calendar, where a unified questionnaire survey could be optimised to reduce pressure on organisations. It should be made clear that Climate Change Duty Reporting Requirements are at site level, whereas benchmarking reporting requirements are at building level, or
b. A public sector/public body reporting review is carried out to identify the optimal period to request information from target organisations to, as far reasonably practicable, avoid times of multiple reporting.

6 Next steps

To provide Scotland's public sector organisations, stakeholders, and policy makers with the most robust decision-making tools available, the SPSEBT Part 1 and Part 2 should and must evolve. To enable this evolutionary process, these are a few of the main next steps for Part 1 and Part 2 of the SPSEBT:

1. Wider database of buildings and improved integration with Public Bodies Climate Change Duties mandatory reporting – Participation of organisations in underrepresented sectors and building activity types. Current estimations suggest that Scotland has approximately 20,000 public sector buildings. Data capture in Phase 2 secured 7243 sites (36% of the total public sector building stock); of these, 6,331 were suitable for statistical analysis (32% of the total public sector building stock and 87% of the initial sample size); see Technical Paper 02. It should be noted that, some sub-BATs contain the entire population of buildings that define that activity type (e.g. the National Health Buildings), for other sub-BATs, e.g. Primary Schools, they will contain a sample size of the total population – e.g. because 6 local authorities were unable to participate in this phase of data capture. However, as definitive population sizes do not exist for some or all sub-BATs, a percentage breakdown cannot be created. Further investigation into how the Benchmarking Tool can be used to improve annual mandatory climate change reporting (coordinated on behalf of Scottish Government by SSN) and how mandatory reporting could be used to drive on-going data gathering to aid the creation of future benchmarks while avoiding duplication of effort across the public sector.

2. A SPSEBT central portal – A centrally managed, yet secure database of information held on currently captured buildings, with remote, organisation level specific login features, i.e. a database portal. Thereby streamlining the data capture process to remove the need for the, somewhat limited, Excel questionnaire surveys, and eliminating the transfer of potentially sensitive data via email. This portal could be designed to hold building information for buildings already collected, whereby future information requests would be a matter of an organisation checking and updating their information rather than recreating it.

3. 'Design year' AMR Profiler – the SPSEBT Part 2 AMR Profiler is built using AMR data from 2019. Future iterations of this Part should seek to secure at least 10 years of AMR data from the participating utility companies. In a similar fashion to dynamic simulation modelling (DSM) for energy demand calculations of new build and retrofit projects, the AMR Profiler would benefit from being built from a 'design year'. Meaning that the profiles are made from monthly energy or water data captured from across 10 or more years, i.e. each week or month of the profile year is statistically selected from a difference period over a 10 + year span. Whilst not intended to be used for forecasting or predictive projections, a design year profile would increase the confidence of the bandwidth and reduce its dependency on local factors that have affected 1 particular month in one particular year.

4. AMR Profiler add-in – the next evolutionary stage for the SPSEBT Part 2 AMR profiler is to move calculation process from a method that requires the user to paste the AMR data into a specific file, to allowing the user to do a profile comparison within their own Excel sheet, i.e. not having to paste data from one sheet into another. This can be achieved by using the Microsoft Excel add-ins ecosystem, similar to add-ins like Energy Lens – Energy Management "Software", or Kutools™ , and other Excel add-ins that are available. Whilst this would optimise the process for users to gain the insights from benchmark profiling comparisons, install the add-in may still require certain user level permissions due to their organisation's IT security policies. However, an Excel add-in allows more complex formulas to be applied without being affected by the computer's processing power, as is the cause with cell level formulas; allowing the user to remain within the familiar Excel interface and bypassing

the known usability issues brought about by individual Macro buttons, and bespoke software that requires installation on a computer's local memory.

Appendix 1 – Benchmark Table, 63 Building Activity types

Parent BAT	Sub-BAT	Carbon dioxide emissions (all energy + water) kgCO$_2$e/year/m^2			Electricity consumption (non-heat) kWh/year/m^2			Heat consumption kWh/year/m^2			Water consumption m^3/year		
		Min	Typical	Max	Min	Typical	Max	Min	Typical	Max	Min	Typical	Max
Offices	**Offices - generic**	34	45	64	47	65	99	103	137	207	147	382	1031
	Offices open plan	34	47	60	60	76	109	104	144	189	310	845	1947
	Offices cellular	42	50	62	44	57	71	146	177	207	120	245	549
Primary & Secondary Educational buildings	**Primary & Secondary Educational buildings - generic**	29	34	44	37	51	69	63	113	151	576	1560	3191
	Kindergarten/Nursery	43	53	67	46	64	93	141	190	243	353	473	715
	Primary school	37	45	55	41	52	68	123	149	184	851	1397	2266
	Primary school with Pool	47	55	82	55	57	57	182	215	257	1692	2560	2727
	Secondary school	32	38	44	43	49	59	105	130	155	2968	4171	6553
	Secondary school with pool	40	52	62	54	59	71	152	194	226	4647	6035	8140
	School special	45	57	72	57	65	75	140	173	253	1181	1872	2647
College/University buildings	**College/University buildings - generic**	39	50	61	63	77	108	124	160	200	1974	4011	8065
	Teaching College/University (classrooms, lecturing spaces, arts)	43	62	76	25	52	89	173	193	297	136	1350	3119
	Special facilities College/University - high consumption, labs, etc.	164	186	195	166	256	279	490	644	701	1087	1468	2664
Sports and leisure	**Sports and leisure - generic**	45	61	83	94	121	166	192	234	325	751	1364	2598
	Leisure centre - Indoor dry sports	59	65	76	93	106	133	212	247	288	1297	1630	3072

	Leisure centre - Indoor dry sports and swimming pool	142	189	237	152	177	236	492	648	899	5568	8080	11216
	Leisure centre - Indoor and external dry sports with external lighting	71	90	105	87	99	119	239	316	372	303	1235	2659
	Sports and leisure - golf course	53	61	62	87	92	100	192	224	277	1305	1851	2734
	Sports ground changing facility >500 CO_2	2144	2264	2440	3476	3929	4439	6314	6987	7722	1098	1478	1498
	Sports ground changing facility	27	36	54	75	90	131	134	172	291	369	520	878
Cultural, Recreational and events	**Recreational and events - generic**	42	56	76	65	84	157	101	167	254	635	1082	3058
	Libraries	44	56	71	48	62	88	145	195	269	38	88	278
	Museums and Galleries	35	47	63	43	77	120	123	172	213	154	550	1656
	Visitor/information centre	31	49	67	53	135	204	180	200	259	154	439	965
	Community centre	43	56	73	33	46	68	161	215	292	166	4567	30999
	Venue - Conference centre, Concert Hall and Theatre	43	53	73	47	60	100	163	221	252	820	1472	2794
	Outdoor events, parks and bothy	37	53	80	29	78	140	119	204	263	655	929	1290
	Food and drink >100 CO_2	134	149	152	388	404	421	191	231	270	334	1140	3077
	Food and drink	22	41	43	35	48	64	77	116	166	37	136	359
Health/Rescue buildings (non NHS)	Care homes	62	80	100	82	101	120	227	313	365	839	2066	3028
	Day centre	51	66	79	46	62	77	199	255	306	377	530	896
	Fire and rescue	45	56	74	56	72	98	205	249	290	46	140	562
Accommodation	**Accommodation - generic**	75	97	114	88	96	121	327	386	431	434	466	875
	Hotels/hostels	56	71	80	55	66	83	213	265	329	648	1119	1414
	sheltered housing	44	55	69	36	50	121	137	178	263	415	1150	1290
	Student accommodation (catered)	45	54	59	54	59	61	180	199	222	7482	10280	17020

Category	Subcategory												
	Student accommodation (self-catered)	14	23	35	36	39	41	61	61	125	1738	3835	4271
	Travelling persons site	233	312	369	1097	1097	1097	438	438	438	1405	2042	3001
Storage and distribution	Depot	21	32	56	54	73	106	66	102	187	327	597	1234
	Recycling, waste, amenity, landfill >100 CO_2	186	281	524	1360	1623	2072	544	649	829	63	127	223
	Recycling, waste, amenity, landfill	15	24	52	59	99	181	25	36	63	1	1	1
	Storage (conditioned)	41	51	72	35	49	87	117	155	289	35	85	154
	Storage (unconditioned)	24	31	66	59	73	81	35	44	61	24	57	177
	Data Centre/Servers	224	342	369	729	933	966	182	278	626	573	626	1310
Other	Crematorium	295	317	358	95	124	163	1191	1397	1493	349	380	840
	Cemetery and religious > 100 CO_2	170	280	360	26	81	206	527	644	919	300	561	839
	Cemetery and religious	24	38	50	44	75	123	38	44	70	43	87	179
	Public toilet	24	47	81	71	99	178	297	408	430	238	458	771
	Car/Vehicle Parking	60	97	116	302	401	561	242	315	376	596	1315	2135
	Passenger terminals	75	81	102	191	215	257	203	280	413	328	834	1368
	Industrial-type activities >100 CO_2	324	439	582	885	1019	1096	342	408	439	1011	1529	2069
	Industrial-type activities	28	39	55	59	63	68	79	121	175	185	426	816
National Health Buildings	Teaching Hospital	71	73	75	69	77	84	255	255	255	43337	63119	82900
	Large, Children and Maternity Hospitals	105	111	125	119	149	169	367	388	440	66945	106052	144077
	General, Dental, Community, Clinic Hospitals	84	96	110	87	99	106	332	371	420	3018	3636	5558
	Acute, Day, Small, Psychiatric, Diagnostic Hospital	70	76	86	65	76	94	242	280	319	957	5529	21772
	Mental, Learning Disabilities Hospital	51	63	77	46	56	69	158	235	280	368	634	880

Dwellings and Accommodation	38	44	69	16	32	39	172	232	305	188	272	2650
Clinic, GP, Dental, Health centres	46	56	68	47	57	71	172	211	258	245	11831	55894
Laundries	643	650	657	262	272	282	2690	2711	2732	62647	68337	74028
Offices, Workshops and Resource centres	41	49	62	57	68	82	153	175	226	557	1166	1998
Special facilities	71	89	114	112	133	177	250	302	418	366	514	745
Storage	44	64	81	40	59	85	149	214	310	318	465	613

Appendix 2 – SPSEBT database fields

Blue cells denote fields that were included within the questionnaire, the data/information/value was provided directly from the participating organisation; this list excludes two data date checking fields and the 'comments' field. Green cells denote fields that were calculated, extracted, or otherwise derived, from one or more of the blue coloured variables.

- Blue cells denote fields/information provided by participating organisations.
- Green cells denote calculated, extracted, or otherwise derived fields/information.

Purpose	Master database column name:	Unit
Identifier	SPSEBT Unique ID:	Number
	Initial grouping:	Number
	Organisation's name	Character
	Site's name:	Character
	Building's name:	Character
	Postcode:	Character
	Outward part of postcode:	Character
	Postcode area:	Character
	Coded Postcode area:	Character
	UPRN :	Character
Class	Sub-Building Activity Type:	Character
	Parent-Building Activity Type:	Character
	Carbon benchmark grouping:	Number
	Electricity benchmark grouping:	Number
	Heat benchmark grouping:	Number
	Water benchmark grouping:	Number
Building Characteristic	Floor area	m^2
	Total floor area (grouped):	Range
	BEMS:	Yes/No
	HVAC Type system:	Character
	Heat distribution type:	Character
	Filter, include in electricity (non-heat) data:	Character
	Heat pump:	Yes/No
	Year of construction:	Number
	Year (age) grouped:	Range
	Energy efficient retrofit (EER):	Yes/No
	Any energy retrofit:	Yes/No
	Retrofit measure - wall insulation:	Yes/No
	Retrofit measure - floor insulation:	Yes/No
	Retrofit measure - roof/ceiling insulation:	Yes/No
	Retrofit measure - window/door upgrade:	Yes/No
	Retrofit measure - draughtproofing:	Yes/No
Resource use	Total electricity:	kWh/year
	Total electricity heating:	kWh/year
	Total gas heating energy:	kWh/year
	MPRN Meter Point Reference Number:	Number

	Total LPG heating energy:	kWh/year
	Total biomass heating energy:	kWh/year
	Total oil heating energy:	kWh/year
	Total purchased heating energy:	kWh/year
	Total renewable electricity generated:	kWh/year
	Total renewable electricity exported:	kWh/year
	Total renewable water heating:	kWh/year
	Total water consumption:	m^3/year
EPC	Current EPC rating:	Number
	Potential EPC rating:	Number
Carbon dioxide emissions	Climate change impact – electricity:	kgCO$_2$e/year
	Climate change impact - electricity heating:	kgCO$_2$e/year
	Climate change impact – gas:	kgCO$_2$e/year
	Climate change impact – LPG:	kgCO$_2$e/year
	Climate change impact – biomass:	kgCO$_2$e/year
	Climate change impact – oil:	kgCO$_2$e/year
	Climate change impact - purchased heating:	kgCO$_2$e/year
	Climate change impact – water:	kgCO$_2$e/year
	Climate change impact – total:	kgCO$_2$e/year
For normalisation	Total Electricity (non-heat)(elec + gen):	kWh/year
	Total Heat (all fuel types):	kWh/year
	Total Energy (all energy types):	kWh/year
	Total renewable generation (electricity [gen + export] + water heating):	kWh/year
Normalised resource use	Normalised Carbon (all):	kgCO$_2$e/m^2/year
	Normalised Electricity (non-heat)(elec + gen):	kWh/m^2/year
	Normalised Heat (all fuel types):	kWh/m^2/year
	Normalised Water:	m^3/m^2/year
Normalised carbon for each utility	Normalised Carbon (Electricity):	kgCO$_2$e/m^2/year
	Normalised Carbon (All heating):	kgCO$_2$e/m^2/year
	Normalised Carbon (Water):	kgCO$_2$e/m^2/year
Normalised checks	Total Energy (all energy types):	kWh/m^2/year
	Normalised Total renewable generation (electricity [gen + export] + water heating):	kWh/m^2/year

BRS
Building Research Solutions