

National Energy & Environment Data Specification

B. Data Resolution, Data Accuracy, Data Type

July 2026

Version 001

This report has sponsored by the **Local Net Zero Hubs** and has been published by the **North West Net Zero Hub** supported with funding from the **Department of Energy Security and Net Zero**.

The report is provided for the purposes of knowledge transfer across the public sector. It aims to raise awareness of good practice approaches to delivering projects that support the net zero agenda. The **North West Net Zero Hub** is not liable for any decisions adopted by organisations based on the recommendations contained in this report.

Funded by:

Commissioned by:



Delivered by:



The logo for BURO HAPPOLD, featuring the text "BURO HAPPOLD" in a bold, black, sans-serif font, centered between two horizontal yellow lines.



Table of Contents

Introduction.....	5
Data Resolution	6
Data Accuracy	7
Data Type	9
Qualitative vs Quantitative Data	11
Case Studies and Examples	11

Acronym	Description
DFES	Distributed Future Energy Scenarios
DSA	Data Sharing Agreement
EV	Electric Vehicle
GDPR	General Data Protection Regulation
GIS	Geospatial Information System
ITT	Invitation to Tender
LAEP	Local Area Energy Plan
LiDAR	Light Detection and Ranging
LSOA	Lower layer Super Output Area
MPAN	Meter Point Administration Number
MSOA	Middle layer Super Output Area
NEEDS	National Energy Environment Data Specification
OA	Output Area
OS	Ordnance Survey
PSGA	Public Sector Geospatial Agreement
SSSI	Site of Special Scientific Interest

Introduction

Digital planning is inevitably based on a “data heavy” process including utilising multiple sets of input data, modelling and producing output and meta data in the form of a digital plan. A robust data strategy is therefore critical to successful planning and presents one of the most significant risks in quality issues and barriers to dissemination or breaches of governance including GDPR.

This subsection covers in brief the key considerations for implementing a data strategy and ensuring this is properly communicated in the specification.

- 1. Data inputs:** the ITT should establish resolution requirements from the start by specifying spatial and temporal resolution of input data (e.g. property level, postcode level, hourly, annual). This helps practitioners understand the level of granularity which is needed and ensures that data sources are appropriate and consistent with the plan’s objective.
- 2. Modelling requirements:** the ITT should outline that the resolution is preserved through the analysis. Practitioners should define how resolution choices affect the accuracy and relevance of the outputs.
- 3. Data outputs:** the ITT should define the expected resolution of outputs to ensure they are usable by the organisation and stakeholders, ensuring that the outputs retain the necessary detail for local decision making and support and integrate with other spatial and energy plans.
- 4. Reporting and dissemination:** the ITT should outline expected level of resolution for reporting (e.g. maps, tables). For Local Area Energy Plans (LAEPs) this is commonly aggregated to Lower layer Super Output Area (LSOA), but the local context may mean that another level of aggregation is desirable.
- 5. Permissions and governance:** should be considered and compliance with license and statutory.

The topics of Data Resolution, Data Access, Data Type (categorisation) are explored further below.

Data Resolution

The scale and resolution of data should be carefully considered when undertaking the baselining and then aggregating for modelling and determining the scale and resolution for dissemination of outputs.

Figure 1 Summarises the different resolution layers of data typically used for LAEPs.

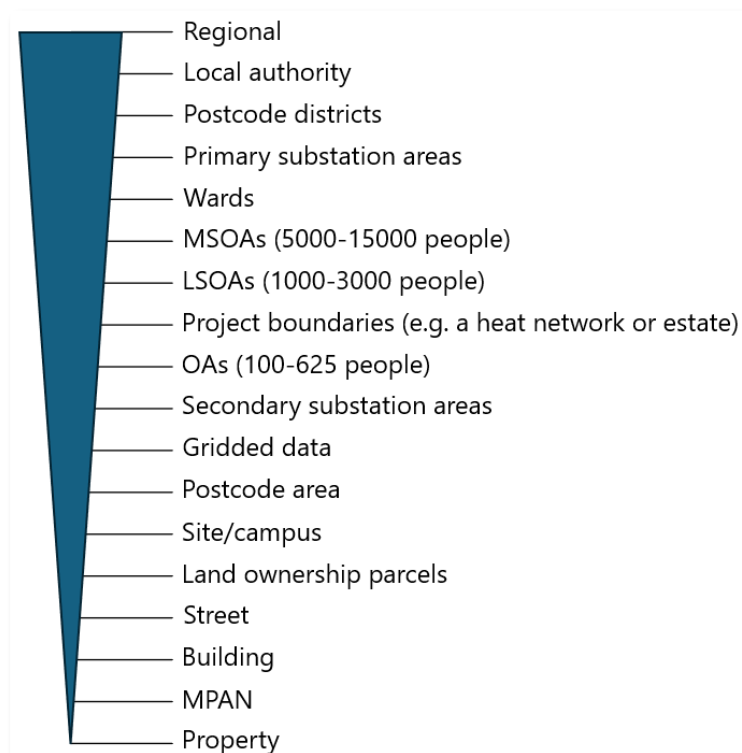


FIGURE 1: EXAMPLE DATA RESOLUTIONS – WITH LARGEST BEING REGIONAL LEVEL AND SMALLEST PROPERTY (MSOA – MIDDLE LAYER SUPER OUTPUT AREA, OA – OUTPUT AREA, MPAN – METER POINT ADMINISTRATION NUMBER)

A key consideration is to understand what the most detailed resolution possible is then determine a strategy for aggregation. It is easier to aggregate high resolution data, whilst potentially not possible to bring resolution to aggregated data. Thus, having **base data in the most detailed resolution possible** is normally preferential.

Third party data sources that provide additional context and clarity should be considered. For example, Ordnance Survey (OS) Mastermap data is often a key source of high-resolution data, meaning an OS Public Sector Geospatial Agreement (PSGA) is a standard requirement¹. OS is also expanding to a more detailed building stock model, the National Geographic Database (NGD). This contains fields relating to building age, use type, and building material – etc.

¹ The OS PSGA is a key underlying requirement for nearly all non-open-source spatial data in the UK. It is explored further in the Governance guidance document.

Measured vs Modelled

Whilst Mastermap can be considered a source of actual measured data (being the standard behind many legal items such as planning documents), NGD includes many modelled fields. This consideration applies to differentiation between modelled and measured data is important to keep in mind throughout the commissioning as measured data is often more accurate but may not cover all areas required. Different technical and commercial areas of the plan may require varying levels of granularity.

Example: UK heat network funds require street level summaries, whilst to feed outputs of LAEPs into Distributed Future Energy Scenarios (DFES), LSOA level data is preferred. These two resolutions cannot be moved between easily but aggregating property level data to these resolutions is straightforward.

Aggregation to spot errors and calibrate data

Being able to aggregate can be used to **identify likely errors and validate/calibrate data**. Many national level datasets exist at LSOA or Middle layer Super Output Area (MSOA) relating to items such as building stock, car ownership, socio demographics, and energy consumption. Aggregation of finer resolution property level information to this resolution is a key way to identify any potential errors in data, which can then be addressed to improve accuracy and confidence.

Governance and data management

Items to be aware of with high-resolution data, such as property level information, include ensuring the appropriate governance and agreements are in place. Therefore, special consideration should be given to the following:

- Data Sharing Agreements (DSAs)
- Commercial sensitives/ Terms of use
- General Data Protection Regulation (GDPR).

The emergence of accurate data including personal data sets

Data sources such as Smart Meter data are increasingly recognised as a key data source to better inform energy transition based on actual consumption. Organisations such as DNOs are increasingly using this and publishing aggregated outputs. Sourcing directly is possible but again special attention is needed when considering permissions.

Data Accuracy

Accuracy and resolution of data are interrelated. Accuracy in data can vary significantly, and so can the confidence associated with it. To ensure reliability, the ITT should request that practitioners validate key datasets, or at least assess the confidence in them. Items which impact accuracy and thus confidence in data including:

- **Completeness** – evaluating the range of coverage and any gaps. *Example - non-domestic EPCs cannot be relied upon to capture all non-domestic properties, as the data has a low coverage of UK non-domestic buildings.*
- **Consistency** – a dataset can contain inconsistencies. *Example - EPC data (which forms the basis of many building models) can vary considerably, this may simply be down to the assessor in a particular area.*
- **Precision and proximity** – of co-located sites may mean data is not accurate. *Example – a doctors surgery located on a hospital complex may be incorrectly allocated the demand of the hospital.*
- **Timeliness** – all analysis is going to rely on data which is not current. The age of the data and how often it is refreshed will inevitably impact analysis. *Example – commodity-based cost data can vary considerably in an unstable market.*

Appropriate allocation of budget and resource should be considered in the commission to allow for data cleansing and aggregation to mitigate the items above and ensure a robust plan is developed based on a clearly understood data set.

Figure 2 summarises the relationship between accuracy at scale and project development as it matures.

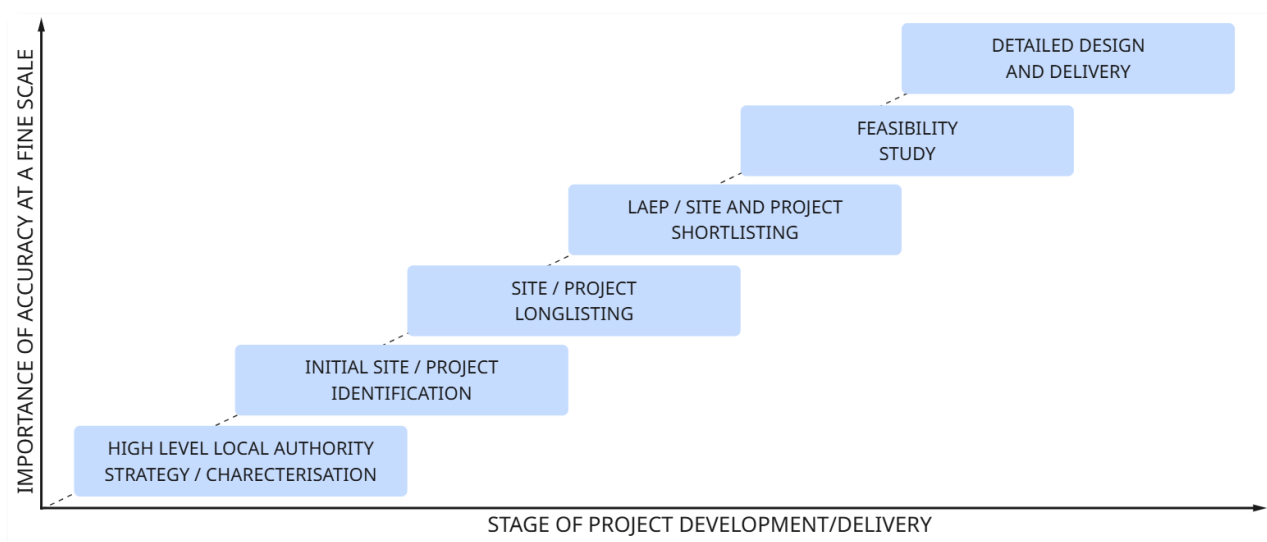


FIGURE 2: IMPORTANCE OF ACCURACY AT A FINE SCALE

Some example data types are outlined below in Table 1

TABLE 1: EXAMPLES OF HOW DATA ACCURACY MAY DIFFER ACROSS DATA SETS

Type	Description and examples	Implications and impact
Modelled / benchmarked data	These are often used when there is a gap in existing data or where future projections are modelled. Examples include building stock models, Electric Vehicle (EV) uptake, Future Energy scenarios and Distribution Future Energy Scenarios. Data should be validated to increase confidence.	The accuracy can vary significantly depending on input data, assumptions, and validation approach.
Modelled / benchmarked with local insights	Enhanced modelled/benchmarked data by applying local knowledge and insights.	Significantly strengthens the data, often both in terms of accuracy and precision.
Measured and recorded data	Often required for detailed applications. Examples include gas and electricity data (e.g. annual, peak demand, hourly and half hourly data), building stock	Higher accuracy and precision
Precise accurately defined areas	Often key spatial boundaries related to a specific characteristic or can be linked with other datasets. Examples include census areas, land ownership parcels, historic areas (e.g. conservation areas).	No need for validation, can sometimes require enhancement. Often have legal status. Generally considered accurate and precise.
Areas defined by models or proxies	These are defined areas by modelling. The accuracy varies on methods and base data. Data can include emissions, soil type (which is important for renewable assessments), land use, brownfield land, areas served by specific electricity substations.	Accuracy and precision should be considered.
Live / recent data	Examples include real time data (e.g. Building energy management System [BEMS] data), smart data, sensor data, weather data, satellite data.	Accuracy and precision are generally good; however, this will depend on measuring approach. Permissions need careful consideration.
Historic data	Examples include historic EPC, gas and electricity usage, weather data, Light Detection and Ranging (LiDAR).	Accuracy and precision will vary and can be out of date but can be useful for assessing trends.
Input data for modelling	This could be benchmarks if input is used as a proxy to infer what you need (e.g. emissions for energy).	Accuracy and precision will vary and can be useful as an initial estimate.

Data Type

Data types define the structure and format of information, influencing how it can be processed, interpreted, and integrated. As such, the format of data-provided significantly affects its usability. Well-structured, clean data is far more likely to be successfully integrated at all development stages of a project. The following items are worth considering:

- **Spatial data:** Ideally this will be a standard Geospatial Information system (GIS) format (e.g. shapefile, GeoJSON, or GeoPackage), in a known projection and with common spatial identifiers included.
- **Images:** can be useful to bring context but consideration is needed if they are to be used in modelling as they can be time consuming to integrate into projects and can also lead to inaccuracy.
- **Tabulated:** most data are likely provided in Excel or CSV format. If the data is integrated into a spatial format, **ensure that this tabular data has:**
 - **A spatial identifier** (e.g. Unique Property Reference Number (UPRN)², TOID (topographic ID from OS which will link to a building)
 - x/y coordinates
 - Postcode
 - Lower layer Super Output Area (LSOA) code).
- **Text based/report data:** It is very difficult to extract data from documents in PDF form, particularly large volumes of spatial data. The relevance and importance of this data must be considered before extraction.

Budgeting and time considerations

Determining the available data types, formats and sources within a local authority can be time consuming so special consideration is needed to either pre-prepare this and/or request in the specification that the delivery party provides clearly defined guidelines on their required sources and formats (assuming they will have this knowledge). Alongside the data types there are other important factors to consider, including:

- **Data assumptions:** if the commissioning organisation have predefined data assumptions (e.g. cost, benchmarks, performance metrics, or reference values) or would like these to be created during the process, these should be outlined in the ITT. This should also specify the format the commissioner and practitioner require the data to be presented in to ensure consistency.
- **Pre-processing and data gathering:** Large volumes of data that require processing or high input from the council, must be considered a key risk by practitioners responding to an ITT. Suitable timelines should be agreed by all parties to allow for this.
- **Governance:** a separate guidance note is provided for governance³. Correct agreements must be in place before any data is shared. The local authority may hold data they think can be used but this may have costs associated with passing the data to a third party.

² This refers to child UPRNs in some instances there will be a parent UPRN (for example a block of flats). Child UPRN is the more important data point, due to the ability to aggregate individual properties but not disaggregate them. However, parent UPRNs are also of use and can prove more consistent for modelling than TOIDs which are used widely in many datasets and national models.

³ C_Governance guidance

Qualitative vs Quantitative Data

Whilst data formats have been considered above, it is important to understand the data fields themselves and the attributes they contain. These are outlined in Figure 3.

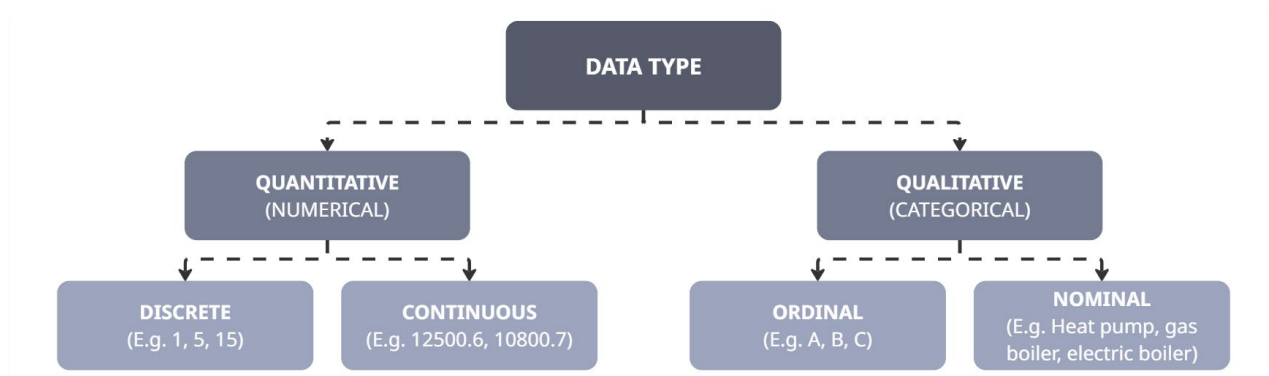


FIGURE 3: DATA ATTRIBUTE TYPES

Quantitative (numerical) data consists of numerical values that support mathematical operations. It can be categorised into the following:

- **Discrete data:** values that represent fixed numbers or counts (e.g. number of occupants – 1, 4, 5)
- **Continuous data:** represented values that can be expressed as fractions and do not need to be presented as whole numbers (e.g. coordinates: 53.4009, -2.9961)

Qualitative (categorical) data describes characteristics or categories and is split into the following:

- **Ordinal data:** categorical data with a meaningful scale (e.g. energy efficiency ratings: A, B, C)
- **Nominal data:** categorical data with no meaningful scale (e.g. types of heating systems: heat pump, gas boiler, electric boiler)

Considering these attribute types in data is important for compatibility and the type of modelling that can be undertaken. Further details and examples of data fields and attributes are provided in the accompanying *NEEDS data tool*.

Case Studies and Examples

EPC data is often the basis for property level information. There are multiple suppliers of EPC derived data, with infilling processes if a property does not have an EPC to create a complete picture. However, this precision is not always accurate, with errors in the infilling process being observed (particularly in non-homogenous building stock such as non-domestic properties or rural areas) as well as inaccuracy in base EPC measured data. Being able to analyse this fine resolution and then aggregate to help mitigate the impact of errors. A study can look to enhance these data, through various crosschecks (e.g. consumption data from network operators or larger aggregated datasets).

Heat Network Zoning data is building level. However, the focus is on heat demand so it can not currently be used to inform retrofit. Data resolution in relation to different fields and parameters is important to consider when understanding what is required of a commission. The accompanying Data Mapping Tool helps to break down the required parameters for different outputs.

Aggregated data is often open access and an important element of modelling in many instances (for example understanding target demographics). It can be a low cost or often free solution to understand an area but if at a later stage individual building data is used it will often give different results. One example of an aggregated data tool using open access data is <https://www.nongasmap.org.uk/>.

Smart meter data is a valuable resource for local authority projects, from energy planning to supporting vulnerable households. Granular datasets, such as household-level readings, offer the most accurate insights and enable robust modelling, but also bring additional requirements around consent, cost, and governance. These obligations vary depending on the access route and should be considered early in project planning. Where such detail is not essential, aggregated data may be preferable, as it often reduces cost and simplifies compliance while still meeting most strategic objectives.