Proposal for a Data Warehouse

Supporting the RetroMeter solution through the storage and assessment of Metered Energy Savings and contextual information

20/02/2024

Connor Enright
Consultant,
ep Consultancy

+44 (0) 7984 349109

Connor.Enright@energyproltd.com
epgroup.com

Contents

Introduction – the relevance and role for a Data Warehouse

The goal within the RetroMeter project is to determine how best to de-risk and deploy value streams within a metered energy savings (MES) retrofit structure, where measurement and verification approaches are used to "meter" the impact of retrofit. By doing so, the aim to catalyse an MES-enabled approach which can be upscaled and adapted across the UK. Upscaling of MES will be dependent on the engagement of external stakeholders and industry actors, and so ep has taken a twofold approach to motivate this engagement.

The first is to take a holistic view of the available value streams, with the aim of converting these, where possible, into robust and de-risked revenues which will enable financing of MES-enabled approaches at scale. The risk and revenues approach assists with the financial assessment of MES-enabled schemes, however wider evidence is needed, particularly to provide actuarial evidence. The formation of this ongoing evidence base is our secondary task within the twofold approach, leading to the proposal of a "data warehouse" to store actuarial data on MES-enabled retrofit projects and their financial & technical performance.

This report aims to identify how a data warehouse could support the goals of de-risking the development of MES-enabled project, and providing an ongoing evidence base for this de-risking impact at each stage of the project or scheme lifecycle.

Literature and Market Review

When examining the value of, and approach to setting up a data warehouse, there are a number of use cases and data modelling approaches that are relevant. This can be seen in the UK market, where examples of data repositories or data warehouses exist, though these pre-existing examples are not specialised or generally suitable for use in an MES approach.

This is because an MES approach considers multiple factors, whilst these pre-existing examples may only provide sufficient data coverage for a single aspect. These multiple factors or use cases include:

- Function 1: Storage of primary data for iterative analysis or use in comparison-based methodologies
- Function 2: Assessment and storage of project performance and financial indicators, such as % over/underestimation of energy savings
- Function 3: Development and description of exemplar cases and scheme designs to provide an evidence base for ongoing upscaling and replication of success programmes and retrofit designs.
- Function 4: Providing consumer-facing data access to ensure households have a transparent view of their data. This will motivate additional engagement by building trust and transparency, as well as potentially providing consumer-level analysis to track energy and cost savings, highlight non-routine events, or highlight the availability and activation of potential flexibility services or revenues.

Having set out the typical functions and services that a data warehouse can provide, we can map these functions onto pre-existing examples in the UK and beyond, as shown below in Table 1.

Exemplar Data Repository	Key Function(s)	Alignment with a RetroMeter solution	Necessary adaptations for a RetroMeter solution.
Smart Energy Research Laboratory (Established by UKRI Engineering and Physical Sciences Research Council - (SERL, 2023))	1 & 2	• Large scale, secure data collation • Use cases serve the wider market and public good	• Enabling commercial access whilst maintaining data privacy • Associating changes in energy consumption with retrofit design (intervention data collected focuses on what assets were installed, not the installation and design process)
De-risking Energy Efficiency Platform (DEEP)[1] (EEFIG, 2016) and EN-TRACK[2] (European Union, 2024)	2 & 3	• Large scale, secure data collation • Focus on associating project aspects such as presence and methodology of verification with project performance	• Additional capture and storage of contextual data – relating to the project delivery route and customer context. • Accessible storage of primary data
Smart Energy Data Repository (Established by Hildebrand and others through the Smart Meter Energy Data Repository Programme [Phase 1]) (Department for Energy Security and Net Zero, 2023)	1	• A centralised energy data repository enabling privacy preserving analysis of smart meter data • Analysis and use cases foreground network services which might include understanding the impact of energy suppliers' flexibility events, conducting carbon accounting, or aggregating demand to ensure grid stability	• Consumer access protocols required • Additional capture and storage of contextual data and project performance against initial estimations. • Association of changes in energy consumption with specific scheme designs or retrofit projects.
The Energy Technology List (Department for Energy Security and Net Zero, 2020)	3	• This example is unique as it provides nominal performance data for specific technologies and models, which has been measured and verified through standardised approaches upon inclusion in the energy technology list. This allows exemplar cases to be developed, specified and replicated.	• This example does not use live data from real projects, but relies on standardised exemplar cases. Adapting the model to provide performance indicators or attributes from real projects would be required, preferably associated with the pre-retrofit performance estimations. • Consumer access would allow customers to compare the performance of specific assets in their home to performance in other projects/portfolios

[1] DEEP was established by the European Commission's DG ENER as part of the Energy Efficiency Financial Institutions Group (EEFIG) De-risking Project 2016-2017 https://deep.ec.europa.eu

[2] EN-TRACK was built as a European Commission Horizon 2020 funded project, project number 885395 https://en-track.eu

Hildebrand Bright (supporting the GLOW Consumer Access Device (CAD) line) (Hildebrand, 2024)	4	• Live data connections enabled through the Data Communication Company (DCC) or Hildebrand's proprietary GLOW CADs • Households can view their own data and simple analytics comparing trends and time periods. • Data is transferred via secure encryption and the data owner may request deletion at any time.	• This platform does not track energy improvements or comparative performance / benchmarking and therefore does not enable assessment of project performance • The platform does not enable sharing of data with third parties for iterative analysis or comparison-based methods
Octopus "my energy insights" data viewer plus associated applications (Octopus Watch / Octopus Energy Watchdog) (Octopus Energy, 2024)	1 & 4	• This platform's primary use case is to enable consumer access and a transparent view of a household's smart meter data. • The platform provides consumers with the ability to download their own primary data for external analysis or benchmarking. • There are specific applications associated with this platform for use on consumer's smartphones or smart watches. Highlights for flexibility	• Additional tracking of interventions and retrofits required • The system does not contain any internal analysis tools currently.
Measurabl (Measurabl, 2023)	1 & 4; Function 2 limited to ESG indicators	• Collection & sufficiency /anomaly testing of meter-level electricity and fuel data • Benchmarking services enable internal and external portfolio comparison • Automated APIs and data imports/exports	• Normalised building performance can be tracked, but there is no input of- and direct association with energy improvement projects • Scheme attributes are not tracked or publicly disclosed • System is focused on Environmental, Social and Governance (ESG) aspects rather than financial indicators
Deepki (Deepki, 2021)	1 & 4; Function 2 limited to ESG indicators	• Automated data connections and APIs to collect live and highly granular project data • Data is structured at the fund, asset and portfolio level, with automated filing • Data coverage monitoring to improve data sufficiency • Asset and portfolio benchmarking	• Building performance can be tracked, but there is no input of- and direct association with energy improvement projects • It is unclear how fund / portfolio data structures capture the attributes and design of the underlying improvement scheme • System is focused on ESG aspects rather than financial indicators

Table 1 above shows the examples currently existing in UK markets and beyond. These examples have been categorised by their general format or market use case below:

- Access controlled research repository: SERL
- One way data input with actuarial and benchmarking outputs: EN-TRACK / DEEP
- Project development case studies and market summaries: the Energy Technology List
- Commercial data repositories - Smart Energy Data Repository (Hildebrand)
- Consumer-facing Energy Data repository: Hildebrand Bright / Octopus data viewer
- ESG data platforms for building and asset portfolios: Deepki / Measurabl

Considering these categories and the key functions, it can be seen that the RetroMeter data solution does not align neatly with any single data platform category, and must be considered as a novel approach to integrate the four key functions described underleaf.

Based on this insight, this report will propose a composite approach to determine the key features and functions of the above exemplar data platforms which could be implemented within the RetroMeter Data Warehouse. These can then form the specification for the final solution, enabling the production of key documentation for the proposed implementation, including but not limited to a proposed development timeline, key attributes of the data model, key data connections and access protocols and core use cases for the platform's sustainability and funding model. Figure 1, overleaf, shows these key features for each exemplar case, with "core" functions highlighted in dark blue, and ancillary functions colour coded in light teal. These ancillary functions will likely be implemented as ongoing developments for the platform.

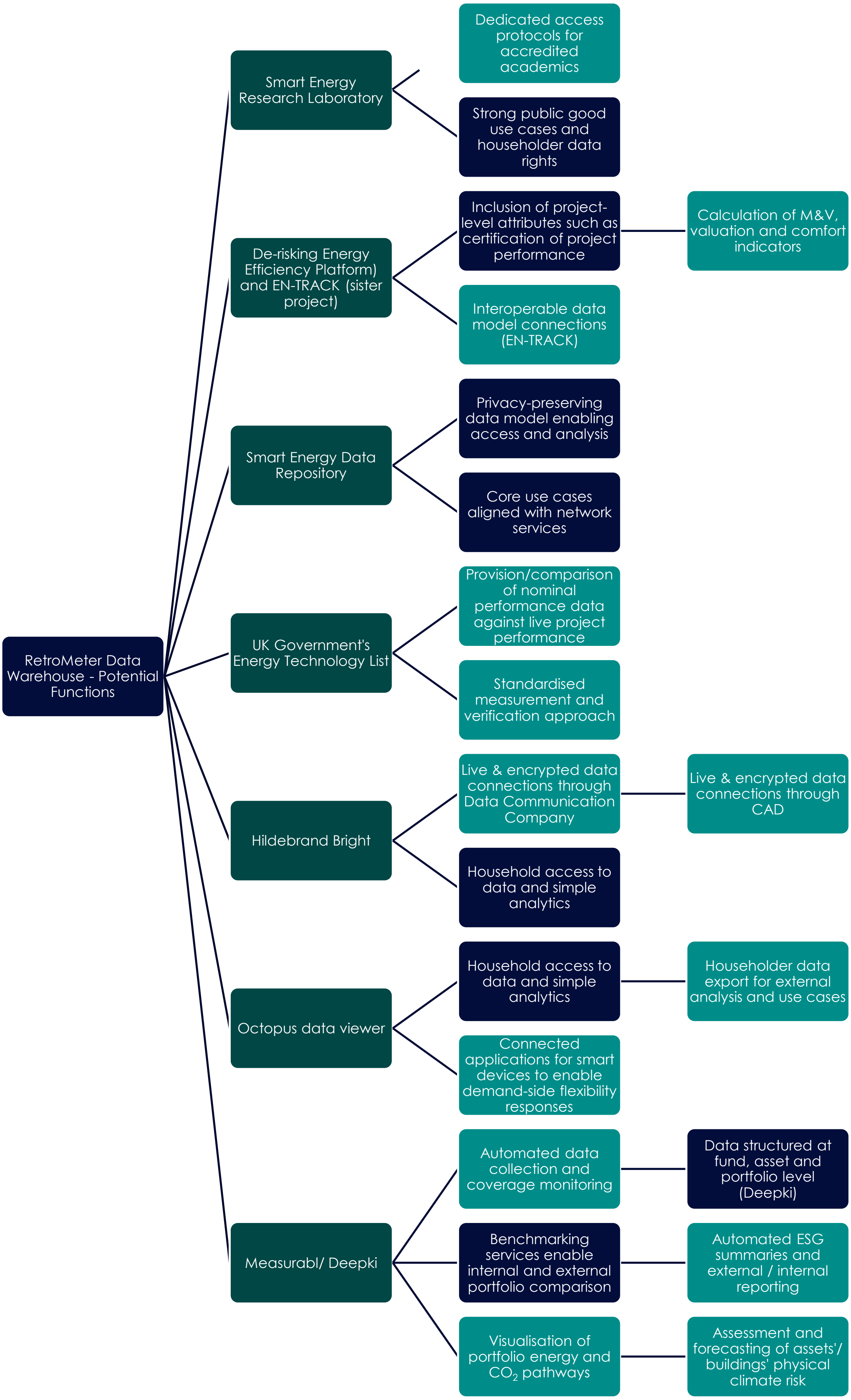


Figure 1: Key features for exemplar case studies of existing data bases, with "core" functions highlighted in dark blue, and ancillary functions colour coded in light teal.

Financial and Actuarial Use Cases

Alongside the functions described above, conversations with financial institutions led to the identification of additional use cases & functionality relevant to the data warehouse:

- **Fund pre-qualification:** The data infrastructure required to verify ancillary benefits and revenue streams (particularly those providing societal value), is a key factor when connecting pay-for-performance and metered energy savings methods with wider societal benefit and societal value streams. Discussions with financial institutions led to the conclusion that investors are far more likely to invest in an indirect financial instrument such as a fund than to internally establish and maintain all the data infrastructure required for project-level investments. This is because of the high transaction costs of structuring a bespoke financing structure for each project.

  Typically a fund is a limited liability partnership structure with Limited Partners ('LPs') who are the investors, and a General Partner ('GP') which manages the fund and is responsible for origination, deal structuring, risk assessment and on-going monitoring on behalf of the fund's investors. Through this mechanism, the investors can be indifferent to the specific projects within the fund, as the GP is responsible for ensuring the LP's risk and return requirements are met.

  Part of this resourcing should include the qualification criteria that enables a project to enter the fund's investment portfolio. The data warehouse could play a key role here by gathering and transferring relevant data on project attributes, ESG impacts and forecast risk/returns profile prior to the transaction. This will enable iterative and reflexive risk assessments and lower the transaction costs of pre-qualifying projects for investment through the fund, and thus upscale the delivery of private finance into metered energy savings retrofit schemes. Examples of this approach include the Tallarna platform (Tallarna Ltd, 2022).

- **Tracking the decarbonisation of investment portfolios:** Many financial institutions are working to decarbonise their portfolios to respond to regulatory and social pressures and their progress on this increasingly has to be verified. Typically, investors will not have the resources, or expertise, to individually assess each property subject to a retrofit to ensure results are being delivered. The RetroMeter solution (Energy Networks Association, 2023), and the resultant data warehouse could assist on two key aspects of verification.

  Firstly, lending can be associated with a "measurable" change in load profiles to verify that the money has been spent on an energy/carbon-saving asset, as described and intended. Secondly, this "measurable" change in load profiles will reveal the performance of the asset and provide the input data required to quantify the carbon abatement resulting from this lending. A granular approach of the latter point would also reveal at what time of day the emissions abatement occurred, which is important as carbon intensity of the grid varies through the day. This work could support the assessment of the risk of asset stranding, i.e. from failure to comply with MEES regulations. An automated approach to access and summarise this data would represent a valuable product for financiers looking to decarbonise their portfolios and demonstrate performance.

- **Actuarial project performance data & lower costs of capital**: The use case described above comments on how a "measurable" change in energy consumption will reveal the performance of the funded asset and provide the input data required to quantify the resultant carbon abatement. However, this change in energy consumption can also provide energy bill savings to the retrofit recipient, and therefore strengthen their ability to repay any retrofit finance. Many lenders already provide beneficial interest rates for building retrofit. These reduced rates can come about from a variety of sources including: margin reduction in order to meet Corporate Social Responsibility or Environmental, Social, Governance objectives; reduced risks of default; or reduced cost of capital for green lending.

  In order to upscale these offers there is a need to provide additional evidence that building retrofits provide a net reduction in energy-related costs, and verify that these lower ongoing costs provide a reduction in default rates for retrofit loans. This evidence can then be used to quantify the reduction in the cost of capital for green financiers, and must be collected at actuarial scales, preferably associated with the attributes of effective retrofit schemes. The RetroMeter data warehouse could be uniquely situated to provide this actuarial evidence to financiers. This approach is exemplified by the case study solution piloted by Sealed and New York Green Bank (Johnson & Molta, 2017).

Synthesis and Extrapolation of Repository Functions

This section will map the core functions described above onto key design aspects (see table below). Some discussion of the financier/aggregator-focused functions are also presented, along with identifying synergies and use cases to motivate the development of ancillary functions.

We have identified a set of core and ancillary functions that the data warehouse could perform. These can now be developed further to understand their implications for the proposed data warehouse design. The core functions are further developed in Appendix 1. We should note that the function titled "Strong public good use cases and householder data rights" is implicit and revealed through the discussion of use cases and audiences, and overlaps with other core functions such as "Privacy-preserving data model enabling access and analysis":

1. Privacy-preserving data model enabling access and analysis
2. Inclusion of project-level attributes such as certification of project performance
3. Core use cases aligned with network services
4. Household access to data and simple analytics (to enable project performance issues, non-routine events and demand response opportunities to be highlighted)
5. Data structured at fund, asset and portfolio level
6. Benchmarking services enable internal and external portfolio comparison

Referencing the exploration of the above functions (in the Appendices), we can begin to draw together overlaps in the required data modelling and implementation approach. These key overlaps include:

- Secure storage of input data and anonymisation of private identifiers, with the ability for householders and data owners to access their data through a secure, likely password-protected data connection.
- Ability to store assessment summaries or pre-analysed/pre-aggregated data, and to segregate access to this data and downstream dependencies at the project/asset, portfolio and fund level.
- The ability to input normalising variables and normalise/adjust energy consumption models, or store pre-processed models, either for individual homes or bundles of 5+ homes.
- The ability to produce simple visualisations such as energy consumption trend curves plotted against various granularities of time, normalising variables and comparable home profiles. This visualisation engine should be able to produce line graphs and curves or histograms, however the ability to produce normal probability plots of residuals or plots of residuals against fitted values would also be highly valuable.
- The ability to input tariff cost and carbon profiles to identify implicit flexibility opportunities and cost/environmental implications. This will be supported by the ability to model discrepancies or non-routine consumption, such as that which arises during a demand response event or non-routine event (i.e. equipment fault).

Developing a data warehouse to satisfy all these function requirements and implementation options is highly complex. The upcoming sections discuss key decisions and recommendations for the establishment of a data warehouse. However, these can be overlaid on a structural data modelling and software development approach to ease and distribute this complexity, as shown below in Figure 2.

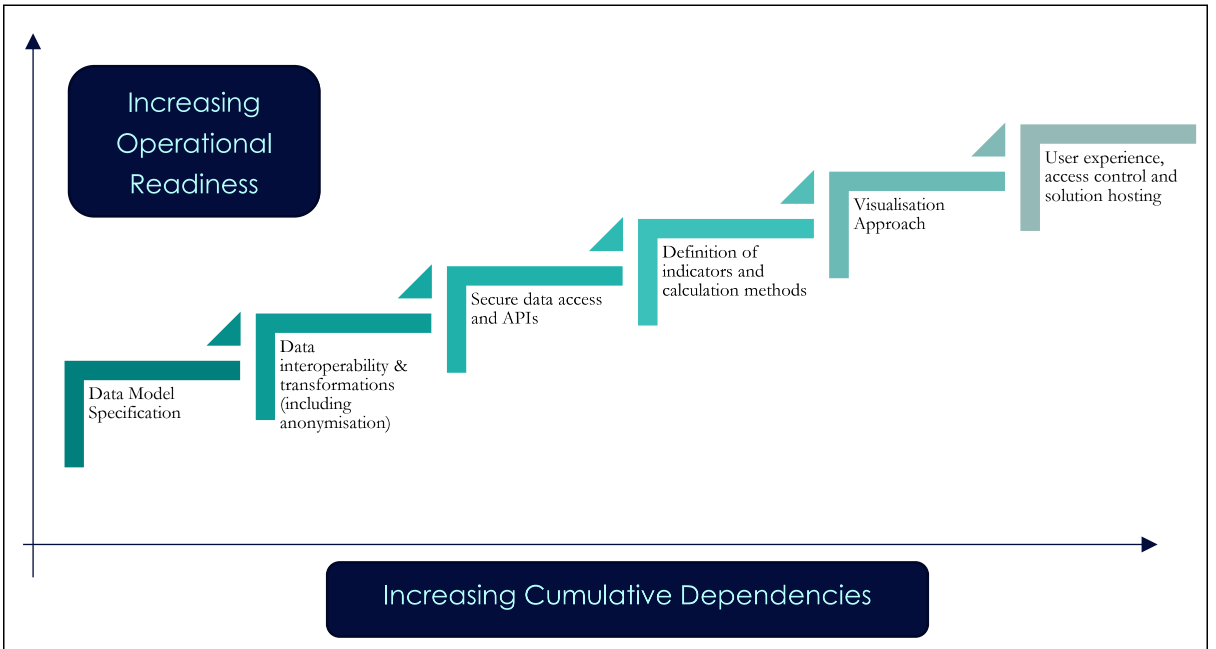


Figure 2: Software development approach showing the key development steps against their increasing cumulative dependency and contribution to operational readiness.

Key Decisions for Establishing a Data Warehouse

This section will discuss the implementation options for the functions above, revealing any key trade-offs, mutually exclusive implementation options, persistent barriers, and commercial decisions. This will also reveal the key decisions required to qualify and develop further / ancillary functions, namely those highlighted in Teal in the figure above. Each key consideration is outlined under a numbered heading in the list below:

1. Data Ownership

Data Ownership will be of key concern, in part because of the varied users of this proposed data solution. It is anticipated that the majority of data will be sourced from, and owned by, retrofit recipients and occupant households, leading to significant data privacy/security requirements and the need to comply with legislation such as GDPR regulations. However, the retrofit recipient will not have a full view of the project-level data, particularly for project/scheme attributes and technical details, such as the presence of measurement and verification processes, project development specifications, or funding breakdowns.

As such, some data will need to be procured from retrofit providers or participating funders to track how these attributes interact with project performance. There will be key decisions around the ownership of this data, and the ongoing privacy liabilities that may arise from its association with private identifiers, or identifiable energy consumption data from specific households.

Often it is the combination of these two data sources that enable modelling of retrofit impacts and links to key reporting and disclosure frameworks that may drive institutional funders (such as NHS Trusts or large private sector landlords). As such, data ownership will need to be reconciled. One approach may be to store these data inputs at different levels, with a secure firewall around the private identifiers that can be used to correlate and connect these data sources. Further discussion of data ownership is discussed under consideration 4 – "Commercial Use Cases and Funding Streams". Further discussion of data partitioning approaches occurs under the consideration below "Storage and Partitioning of Data".

2. Storage and Partitioning of Data

Data Storage will be a key consideration for ensuring the security and privacy of user data. A range of anonymisation approaches have been proposed, including data masking; pseudonymisation; data swapping; data perturbation and the production of synthetic data. Each of these will have benefits in terms of the strength of security, likelihood and options for re-identification, and impacts on data utility (i.e. the generalisation of energy consumption data may preclude granular modelling approaches).

In order to enable different use cases and value to various users, it is proposed that data may be partitioned into various interconnected and interoperable data structures, namely at the asset/project, portfolio and fund level. This would allow households/occupants and their representatives to access their own data at a granular, identifiable level, improving transparency and reducing information asymmetry regarding their project and its performance. Meanwhile, large-scale retrofit providers could access more generalised data regarding project performance, non-routine events and portfolio impacts, likely through project-level financial/environmental indicators or aggregated project consumption profiles. Equally, financiers could track their investments and impact across multiple project portfolios, assisting with their decarbonisation tracking and accounting, and providing significant value regarding ESG information, which could help to fund the solution moving forward, as discussed under consideration 4 – "Commercial Use Cases and Funding Streams".

3. End users: Exports and APIs

The data collected by this proposed data warehouse would have wide-ranging use cases and value to external actors such as insurance providers. There is a key question around whether it is appropriate to enable data export or external data connections to realise this value. Firstly, these use cases would have to be defined and accepted by households within their Data Rights Agreements, and the acceptability of external use cases has not been tested in detail as part of this project.

Secondly, there is the issue of interoperability. Data is often only useful when the appropriate data connections can be put in place, enable the data to be connected to- or transformed into a data format and structure that aligns with external organisation's ongoing approach and services. There are existing examples of interoperable data bases, such as DEEP or EN-TRACK, but these databases have a small number of external connections and were often built upon open source or interoperable data models that have been pre-agreed or co-developed. Ensuring export and interoperability with numerous external entities could result in a "moving feast", where the complexity of aligning many different approaches hampers a coherent implementation.

There is also an interaction between export and privacy concerns. On one hand, there may be a requirement to store private identifiers required to enable users to access their own data, and therefore there will be a need associating data with the occupant. However, this association, or related private identifiers should not be accessible or exportable for external actors, but may be required to filter and select the appropriate data. There may be a need therefore to include a function within the data partition that enables secure selection, filtration and aggregation of identifiable data, without direct access through the API.

4. Commercial use cases and funding streams

Alongside the identification of an appropriate funding mechanism for the up-front capital investment required to build the platform, the data warehouse solution we propose will need to be financially self-sustaining in order to provide value to the market and satisfy the requirements of specific functions, particularly those that track project performance over the long-term and produce actuarial assessments and data summaries. The initial investment may need to be developed at risk or using grant-funding, which must be considered as part of further funding stream discussions.

In order to ensure a financially self-sustaining model, there is a need to consider and trade-off between commercial use cases and public-interest funding streams. There are several commercial audiences that may be interested in funding or purchasing services through the proposed data warehouse, as described in the sub-section titled "Financial and Actuarial Use Cases", along with use cases specialised to Retrofit Providers and their products.

- Financial Audiences: There are two key drivers to drive. commercial services and revenues for financial audiences. The first is to understand how existing or upcoming investments on their loan book can be decarbonised, leading to reputational benefits, supporting risk management, the accounting of environmental benefits and reporting mechanisms. These benefits can lower the cost of capital where a financier is already undertaking steps to achieve these goals, by providing a low transaction cost, auditable framework to verify these benefits. The verification that investment has supported the ongoing longevity and creditworthiness of the recipient could also work to reduce the cost of capital by demonstrating reduced default rates.

  A further reduction in the cost of capital arises from the second driver, which is to assist financiers in the identification of high quality projects with a positive impact. These positive outcomes can leverage in additional funding, thereby lowering the financial exposure to lenders, but the RetroMeter solution and data warehouse could also lower the transaction cost of identifying high quality projects, and pre-qualifying these projects for inclusion in a fund, thereby reducing the cost of capital and allowing financiers to reclaim some margin which may be given away within previous "green" or "impact" finance approaches.

- Actuarial Audiences: Insurers may be willing to provide ongoing revenues and funding for services that allow them to measure the performance of a funded asset and understand the risks to project performance which they are asked to underwrite. Collecting information on the default rate of retrofit investments will help evidence these risk profiles, as well as assisting with the identification of project attributes which may mitigate these risks, such as the use of accredited installers or measurement and verification procedures.

- Retrofit Providers - Modelling target market segments and comparison groups: Retrofit Providers may value the data gathered by the data warehouse as they seek to understand how retrofits can be deployed across varied housing and project typologies to achieve the greatest impact and best risk-return profile available within a given geography. This is particularly the case for area based schemes that will blend a wide ranging project and housing archetypes, and therefore may not have a proprietary evidence base to deploy, upscale or replicate a successful retrofit programme.

  In addition, the RetroMeter solution may rely on comparison-based methodologies to accurately model household energy consumption in a changing energy market, impacted by the cost of living crisis, the COVID-19 pandemic and many external factors. To do so, a comparison group must be identified to find representative homes where a retrofit has not occurred, either through the assessment of like load profiles or building attributes. The data warehouse would be an ideal platform from which to source these comparison groups at scale.

- Retrofit Providers and Public Sector Evaluation - Methodology testing / Retrofit Scheme Evaluation: As retrofit upscales across the UK, and our housing stock transitions towards Net Zero, there is a need to pilot and evaluate retrofit technologies, designs and deployment approaches at larger scales and lower transaction costs. In order for such evaluations to be comparable and useful across different schemes and evaluation periods, there is a need for a standardised methodology and data structure to test and capture project performance. The RetroMeter methodology and proposed data warehouse would be well situated to provide and aggregate such evaluation services.

  Furthermore, this service would be delivered on behalf of public sector bodies, foregrounding the delivery of public services and support for retrofit, as well as assuring value-for-money to the taxpayer. This may make this revenue stream and associated prerequisite data rights more acceptable to participating households and retrofit recipients, as it does not leverage their personal data for solely private benefit.

5. Platform Management/Governance approach

Any large-scale data platform will require ongoing management, and this management should be underpinned by an equitable, transparent governance structure. Whilst the final selection of stakeholders will need to be determined, potential actors within this governance structure could include, inter alia, UK Government's Department of Energy Security and Net Zero, the Data Communications Company, the Coalition for Energy Efficiency Buildings or the Green Finance Institute.

One would also need to consider the governance mechanisms through which stakeholders and managers act. One option would be to establish a board of directors, including those who must satisfy "independent requirements", who in turn will oversee executive and technical committees. This approach is similar to that established by the Data Communications Company. Further detailed consideration of precise articles of association, licensing requirements and the definition of roles and responsibilities will also be required.

6. Data sources

Whilst this report has discussed how data may be stored and assessed, and the issues surrounding this processing, there has been minimal comment on how data will be sourced. This is in part because a large-scale data warehouse would need to be flexible to integrate data from a range of input sources, particularly if providing an open-source approach or industry standard solution. Specific options for gathering data include:

- Via the Data Communications Company: The Data Communications Company or DCC operate Britain's secure smart meter network, and therefore would be responsible for establishing any APIs or data connections through their DCC gateway, which currently provides access to UK smart meter data for energy suppliers, network operators and authorised third parties. The proposed data warehouse would likely apply under this third category. This approach is advantageous as it can operate at large scale with minimal transaction costs, but involves an expensive set-up process and ongoing regulatory compliance.

- Consumer Access Devices: Consumer Access Devices (CADs), also known as "in-home devices" are cloud-connected secure meter gateway devices that enable access to real-time energy data from smart meters, sending such data to a designated cloud service. These CADs could feed data into an intermediary platform or API, or directly into the data warehouse. This approach is advantageous as it has a lower technical complexity than establishing a secure national API, but adds additional cost to individual retrofits.
- User Input: Users could input their data manually via an online portal. This approach is useful for addressing data gaps or telecommunications drop-outs retrospectively, but is dependent on timely engagement from households. This approach may give households a feeling of greater control over their data sharing, but these households must also be motivated to engage to much greater extents over much longer periods. Whilst the set-up costs of this option is lower than the prior options, the portal must also be secure, and the cost of ongoing household engagement may be much higher. This is not considered a viable option.

Recommendation and Proposed Solution / Implementation

This section will propose a high-level development timeline, recommendations for developing specific functions and how they relate to the remaining decisions that will be considered as part of the establishment/commercialisation of the data warehouse solution. Our recommended development process would cover the following key steps:

1. Establish a steering committee or similar to conduct on-going stakeholder engagement and governance conversations. This will ensure that the final solution is equitably co-produced in line with the needs and trade-offs that each stakeholder and end user is concerned with.

2. Through the steering committee or preliminary governance structure, begin discussions to confirm the data model specification, categorise core and future functions/services & discuss a self-sustaining funding solution.

3. Identify a funding mechanism that can be used to underpin the significant up-front investment required to build a data warehouse platform

4. Identify stakeholders to assist with the piloting and development of the data warehouse.

5. Implement or adapt the following development pathway (see overleaf):

6. Assess and prioritise the following ancillary functions, engaging the steering committee or stakeholders to assist with the construction of an ongoing development pathway and related services.
    a. Dedicated access protocols for accredited academics
    b. Calculation of M&V, valuation and comfort indicators
    c. Interoperable data model connections (EN-TRACK)

d. Provision/comparison of nominal performance data against live project performance
e. Standardised measurement and verification approach
f. Live & encrypted data connections through Data Communication Company
g. Live & encrypted data connections through CAD
h. Householder data export for external analysis and use cases
i. Connected applications for smart devices to enable demand-side flexibility responses
j. Automated data collection and coverage monitoring
k. Automated ESG summaries and external / internal reporting
l. Visualisation of portfolio energy and CO_2 pathways
m. Assessment and forecasting of assets'/ buildings' physical climate risk

Appendix 1: Proposed Core Functionality

Core Function: Privacy-preserving data model enabling access and analysis
Use cases and Audience: • Enabling the input of primary data and export/visualisation of analysis summaries (all audiences) • Enabling privacy-preserving access to accredited researchers and data partners (accredited researchers and data partners)
Supporting Functionality: • Differential privacy, perturbation/noise-adding mechanisms, data masking, pseudonymisation or other privacy-preserving anonymisation techniques
Implementation – Data Inputs (Data needs and sufficiency): • Input raw data (storage may or may not occur) • Description of anonymisation or re-identification procedures. The latter may not be feasible or appropriate, depending on the data rights granted and the anonymisation approach.
Implementation – Data Storage (Privacy and Security): • Any correlated private identifiers and pseudonyms / anonymised identifier should be stored securely. • Data connections between the private identifier (i.e. via consumer log in) and the anonymised data should occur through secure APIs only, likely password-locked. • There may be a need for regular penetration testing to validate the privacy-preserving implementation and design of the data model.
Implementation – Data Analysis (general approach, aggregation and outputs): Potential implementations include: • Data Masking – mirror database and implement alteration strategies such as character shuffling, encryption, term, or character substitution • Pseudonymisation – Swapping private identifiers with pseudonyms • Generalisation – excluding or generalising data (i.e. first 4 letters of postcode or nearest weather station instead of full postcode) • Data swapping – rearrange dataset attribute values so they do not fit the input information. • Data perturbation – addition of random noise or application of round-numbering methods • Synthetic data – construction of synthetic data sets that model the original data.

Core Function: Inclusion of project-level attributes such as certification of project performance (likely referencing the nuanced definitions developed by prior public platforms such as EN-TRACK, DEEP and the Building Performance Database)
Use cases and Audience: • Formation of an actuarial evidence base to validate the improvement in project performance and default rates that may arise from MES methodologies (financiers and insurers) • Post-project assessment of scheme designs, leading to the development and description of exemplar cases and scheme designs (retrofit providers) • Consumer tracking of project performance and non-routine events (household)
Supporting Functionality: • Standardised M&V approaches • Standardised project performance indicators and calculation methods (such as normalised energy savings, percentage over/underperformance, outcomes-based KPIs, etc.)

Implementation – Data Inputs (Data needs and sufficiency):

- Input energy consumption data or energy consumption models covering the baseline and reporting period
- Normalising variables such as localised heating degree days
- Thresholds and indicators for comfort-takeback rating
- Project performance estimations / specifications

Implementation – Data Storage (Privacy and Security):

- Storage and utilisation of pre-processed energy consumption models would remove some privacy concerns
- Input energy consumption data should be stored behind a secure firewall if not pre-anonymised for analysis and visualisation
- Project performance summaries should be aggregated to a minimum number of homes to avoid identification of specific homes and their project performance

Implementation – Data Analysis (general approach, aggregation and outputs):

- One key decision will be whether project performance indicators will be calculated and stored at a set period (i.e. 1 year after project completion), or will be live and re-calculated at the point of access.
- The integration of model adjustments may be complex and require input not only from households but also from technical staff from the retrofit provider.
- Decision to be make whether to bundle performance data for representative portfolios or neighbourhood projects, presenting an aggregated view to mitigate privacy concerns and statistical noise.
- Consideration of the minimum period over which to present project performance (i.e. one full year or heating season).
- Storage of meta data such as model accuracy and input data sufficiency

Core Function: Core use cases aligned with network services

Use cases and Audience:

- Measurement and verification of load or peak load reduction (the latter of which may need to be added later due to complexity) enabling deferred network reinforcement (network partners)
- Measurement and verification of contracted (explicit) flexibility services (network partners)
- Measurement and verification of non-contracted (implicit) flexibility services (households, energy suppliers and commercial actors)

Supporting Functionality:

- Normalised/adjusted energy consumption models pre- and post-retrofit
- Storage of- and assessment against flexibility service specifications

Implementation – Data Inputs (Data needs and sufficiency):

- Flexibility contracts and service specifications
- Sufficient raw data or pre-processed models to enable confidence within statistical measurement and verification
- Tariff-aligned load profiles as needed for implicit flexibility

Implementation – Data Storage (Privacy and Security):

- Secure or anonymised storage of private identifiers in flexibility contracts or service specifications

- Aggregated flexibility profiles of 5 or more homes could be input/output from DNO data

Implementation – Data Analysis (general approach, aggregation and outputs):

- Data analysis should be aligned with the International Performance Measurement and Verification Protocol (IPMVP) best practice
- Aggregation could occur on tranches of 5 or more homes that were developed and assets installed as a "bundle".
- Output the volume and confidence of flexibility provided.

Core Function: Household access to data and simple analytics

Use cases and Audience:

- Data Transparency and trust building (household/retrofit provider)
- Flagging non-routine events
- Highlighting and tracking flexibility events and responses

Supporting Functionality:

- Secure access control
- Basic visualisation and regression tools
- Non-routine event detection (significant discrepancies)
- Notification solutions for flexibility event responses (email, push notification, smart device APIs etc.)

Implementation – Data Inputs (Data needs and sufficiency):

- Data Connection / API to input and identify flexibility events
- Explicit flexibility availability periods and dispatch thresholds
- Live / Smart meter data to be visualised and accessible to householder

Implementation – Data Storage (Privacy and Security):

- Input energy consumption data should be stored behind a secure firewall if not pre-anonymised for analysis and visualisation
- Secure or anonymised storage of private identifiers in flexibility contracts or service specifications
- Secure access of private identifiers and log in details for the visualisation and export portal.

Implementation – Data Analysis (general approach, aggregation and outputs):

- Visualisations of energy consumption plotted against:
    - Time (Sub-daily, daily, monthly, annual)
    - Normalising variables (heating degree days or external temperature)
    - The modelled/aggregated consumption of comparable homes
- Non-routine event detection will require identification of discrepancies in the normalised energy consumption models, alongside thresholds at which these should be flagged as non-routine events
- Connections to energy tariffs could convert measured flexibility responses into expected bill savings, incentive payments or carbon impacts.

Core Function: Data structured at fund, asset/project and portfolio level

Use cases and Audience:

Segregation/partitioning of the data structure enables different use cases at different levels:

- Fund: Fund pre-qualification (i.e. based on attributes and performance of prior comparable projects); Tracking the decarbonisation of investment portfolios (validating the installation and real performance of financed assets, particularly through unsecured finance); Assembly of an evidence base of actuarial project performance data (hopefully demonstrating a de-risked MES method can lower default rates and therefore costs of capital)
- Asset/Project: Assessment of project performance (i.e. savings and financial returns); provision of secure but transparent data access for householder; assessment and improvements to asset performance (i.e. highlighting flexibility or non-routine events); Benchmarking services
- Portfolio: Portfolio benchmarking services; Tracking and improving portfolio performance; Identifying new sites and projects that align with portfolios and their development specification.

Supporting Functionality:

- Segregated/partitioned data model and data connections

Implementation – Data Inputs (Data needs and sufficiency):

- User access rights / register at the fund, asset/project and portfolio levels
- Data translation specifications, for either internal calculations or external interoperability

Implementation – Data Storage (Privacy and Security):

- Data should not be accessible across the different levels without specialised use case and data permissions. This will enable more commercial use cases with minimum impact on data privacy and security
- Data rights agreements will need to cover allowable data use at all levels
- Storage of pre-analysed or pre-aggregated data at portfolio and fund levels will help preserve the privacy and security of raw or identifiable input data

Implementation – Data Analysis (general approach, aggregation and outputs):

- Simple visualisations could use the visualisation engine developed within the "Household access to data and simple analytics" use case.
- Complex visualisations and data models may reference existing solutions, such as those produced by the EN-TRACK project (see Appendix 2)

Core Function: Benchmarking services enable internal and external portfolio comparison

Use cases and Audience:

- Comparison of project performance within internal portfolios, enabling the development of improved household targeting, rapid redress of performance issues and evaluation of scheme design (retrofit providers)
- External comparison between different portfolios and scheme designs to reveal household and retrofit design attributes that improve or hamper project performance (financiers and insurers)

Supporting Functionality:

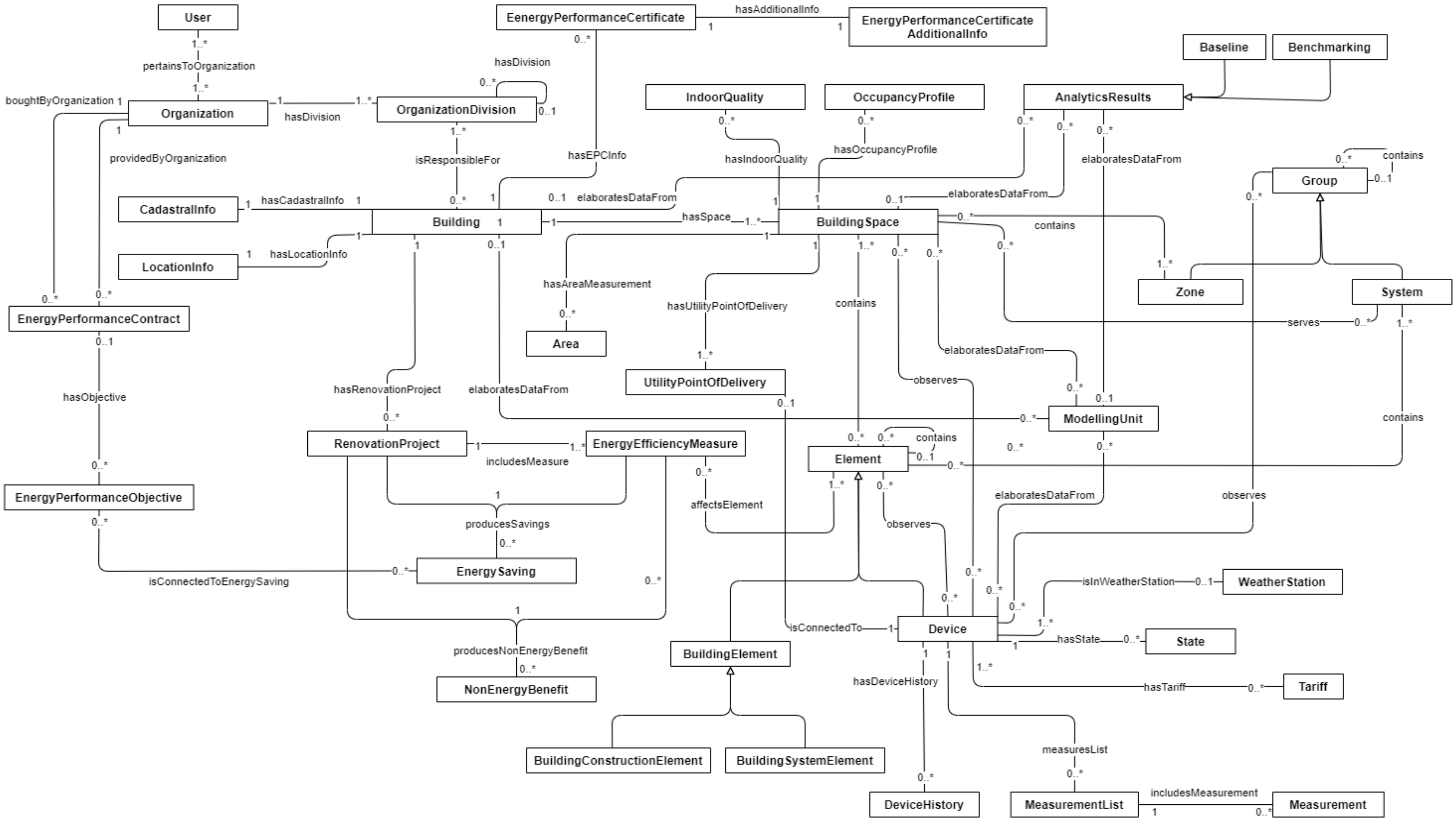
- Visualisation engine showing distributions of project performance as histograms, normal probability plots of residuals or plots of residuals against fitted values
- A partitioned and interoperable data model and transformation/translation methods between data levels and syntaxes.

Implementation – Data Inputs (Data needs and sufficiency):

- Access to raw data or data summaries for visualisation.

• Available forms of benchmarking visualisation and minimum data requirements (i.e. minimum of 5 pre-aggregated sites to preserve data privacy)
Implementation – Data Storage (Privacy and Security): • Portfolio assessment could occur at regular timeframes to ensure the minimum number of homes and months of heating data are present. These assessments could then be stored at the portfolio level in a summarise or pre-aggregated format to enable rapid visualisation whilst partitioning data at the fund, portfolio and project level.
Implementation – Data Analysis (general approach, aggregation and outputs): • Simple visualisations could use the visualisation engine developed within the "Household access to data and simple analytics" use case. • Complex visualisations and data models may reference existing solutions, such as those produced by the EN-TRACK project (see Appendix 2)

Appendix 2: EN-TRACK Data Model



References

- *De-risking Energy Efficiency Platform* (2016) *DEEP*. Available at: https://deep.ec.europa.eu/ (Accessed: 20 February 2024).
- *Deepki Solutions* (2021) *Deepki*. Available at: https://www.deepki.com/solutions/ (Accessed: 20 February 2024).
- Department for Energy Security and Net Zero (2023) *Smart meter energy data repository programme: Phase 1 projects*, *GOV.UK*. Available at: https://www.gov.uk/government/publications/smart-meter-energy-data-repository-programme-successful-projects/smart-meter-energy-data-repository-programme-phase-1-projects (Accessed: 20 February 2024).
- *EN-TRACK Homepage* (2024) *EN-TRACK*. Available at: https://en-track.eu/ (Accessed: 20 February 2024).
- *Energy Technology List (ETL)* (2020) *GOV.UK*. Available at: https://www.gov.uk/guidance/energy-technology-list (Accessed: 20 February 2024).
- Johnson, L. and Molta, P. (2017) *Case Study: Home-Efficiency Financing with Sealed And NY Green Bank* , *Yale Center for Business and the Environment*. Available at: https://cbey.yale.edu/sites/default/files/2019-08/NYGB.pdf (Accessed: 20 February 2024).
- *Octopus energy homepage* (2024) *Octopus Energy*. Available at: https://octopus.energy/ (Accessed: 20 February 2024).
- *Our products* (2024) *Hildebrand*. Available at: https://www.hildebrand.co.uk/our-products/ (Accessed: 20 February 2024).
- *Product features - ESG Software* (2023) *Measurabl*. Available at: https://www.measurabl.com/esg-software/ (Accessed: 20 February 2024).
- *Retrometer* (no date) *ENA Innovation Portal*. Available at: https://smarter.energynetworks.org/projects/10055401/ (Accessed: 28 February 2024).
- *Tallarna homepage - A climate tech company for the built environment* (2022) *Tallarna*. Available at: https://tallarna.com/ (Accessed: 20 February 2024).
- *Welcome to the Smart Energy Research Lab* (2023) *Smart Energy Research Lab*. Available at: https://serl.ac.uk/ (Accessed: 20 February 2024).

Company details

Company name: EnergyPro Ltd

Registered address: 7 Bell Yard, London WC2A 2JR

Company no.: 08236665

VAT no.: 179-5115-83