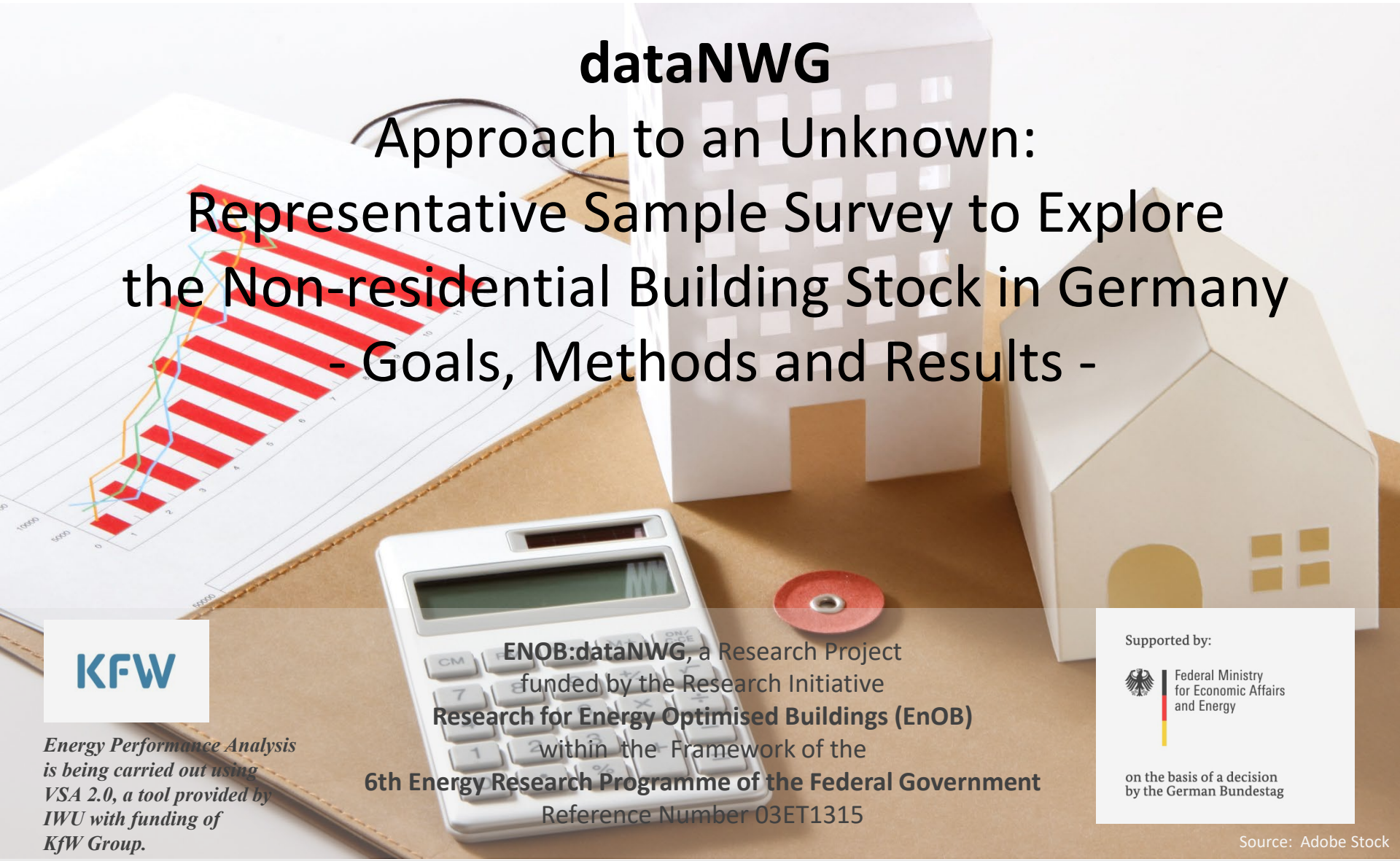
dataNWG

Approach to an Unknown: Representative Sample Survey to Explore the Non-residential Building Stock in Germany - Goals, Methods and Results -



ENOB:dataNWG, a Research Project
funded by the Research Initiative
Research for Energy Optimised Buildings (EnOB)
within the Framework of the
6th Energy Research Programme of the Federal Government
Reference Number 03ET1315

Supported by:
Federal Ministry for Economic Affairs and Energy
on the basis of a decision by the German Bundestag

Energy Performance Analysis is being carried out using VSA 2.0, a tool provided by IWU with funding of KfW Group.

Source: Adobe Stock

1 dataNWG Research Goals

- Representative Building Stock Database of Monitoring Variables
  - **Structural parameters** of the non-residential building stock (NRBS) like spatial distribution, building types, total number, total area, building envelope areas etc.
  - **Energy-related characteristics** of building envelopes and technical installations of relevant non-residential buildings in the stock.
  - **Measured energy consumption** and real usage parameters.
- Building Stock Model and Scenarios
  - **Model-input variables** like renovation progress and annual rates of building parts and technical installations, U-values, heat generators' expenditure factors, calculated energy demand
  - Development of a reduced order **Energy Performance Simulation Tool (DIBS,** available on GitHub**) for building stock modelling** operating with only 35 of these model-input variables.
  - **Calibration** of energy demand calculations by measured consumption data
- Geo-spatial Data Analysis:
  - Calibration of **geoinformatic recognition algorithms** of non-residential buildings combined with image processing and machine learning to recognize NRBs and their building function from building polygons and 3D building models

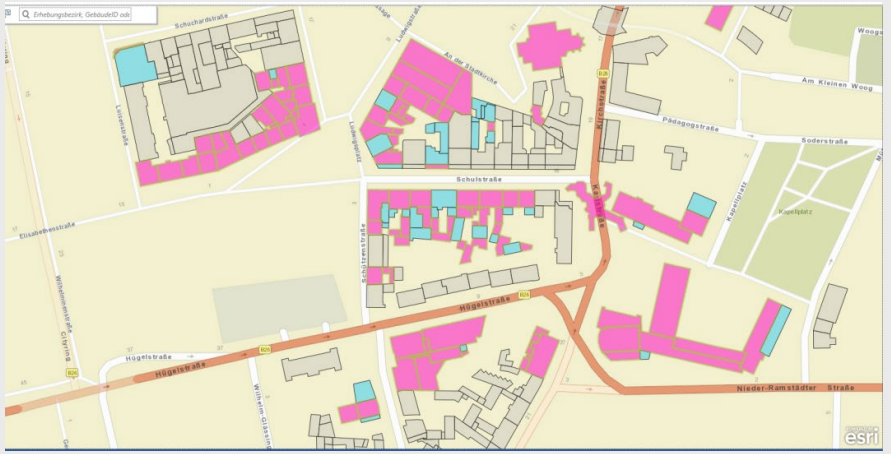
2 Key Findings

- How many are there?
  - 21.124 ± 0.445 Mio. non-residential buildings (NRB)
  - 1.981 ± 0.152 Mio. BEA(1)-relevant NRBs on 3,507 ± 399 Mio. m² GFA(2), thereof 1.146 ± 0.110 Mio. old buildings(3)
  - NRB-Typology differentiated according to building category and construction age classes
- Energy-related building renovation progress and rate in old buildings
  - weighted by the area fractions g_i of the building envelope components(4) : renovation progress 30.3% ± 7.1% of component surface; corresponding to an average yearly renovation rate in 2010 – 2019 of 0.7%/a ± 0.2%/a
- Missed opportunities
  - Repainting or plaster renewal of outer walls without applying thermal insulation at the same time: 3.2%/a ± 0.5%/a (of all buildings built before 2010).
  - Single measures preferred in 70% ± 5% and 4 or 5 measures (deep renovation?) in less than 10% ± 4% of the renovation cases in 2010 – 2014.
- Heat supply in old buildings
  - Modernisation rate of heat generators: 2.5%/a ± 0.3%/a
  - No fuel switch in 81% ± 6% of gas heating renewals

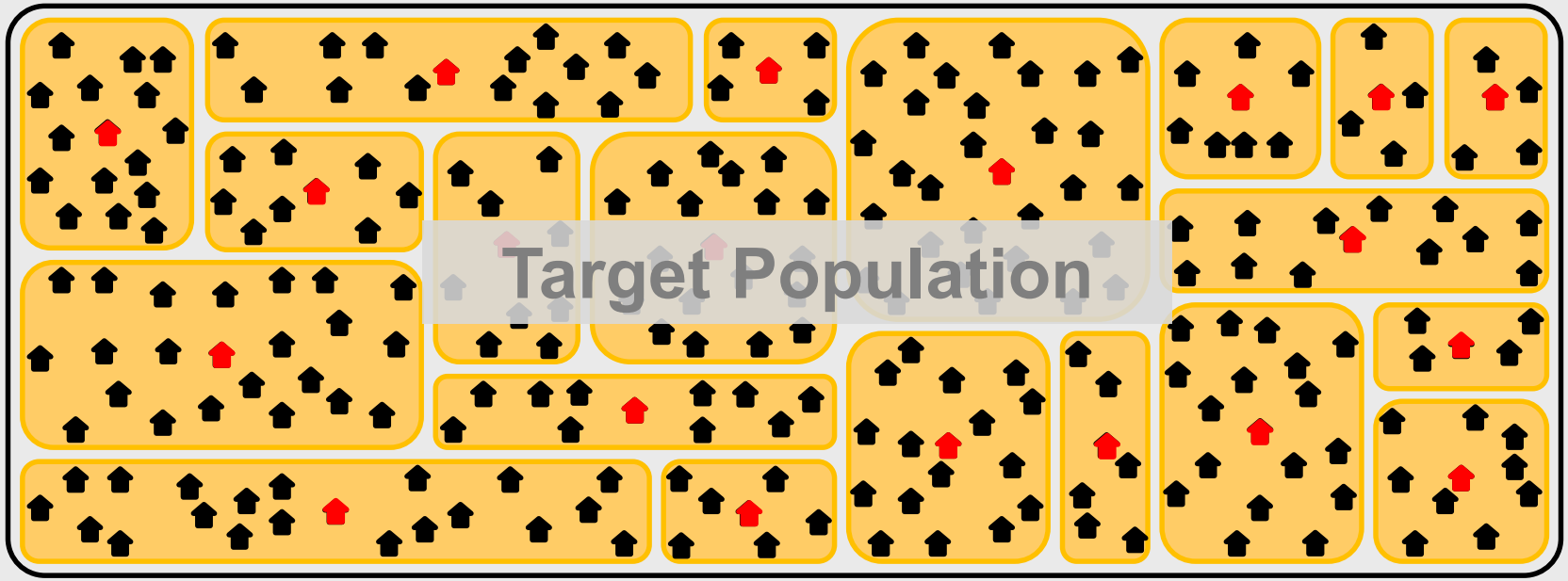
(1) BEA-relevant: thermally conditioned and relevant under the German Building Energy Act (Gebäudeenergiegesetz (GEG))
(2) GFA: Gross floor area
(3) Old buildings: built before 1978
(4) Components: outer walls, windows, roofs, floor slabs

3 What‘s so special about these data?

- This type of data can only be obtained through interviews with owners. In order to keep effort and costs within limits, this needs to be carried out in **sample surveys**. However, a suitable sampling frame has been lacking so far.
- **Authoritative geospatial data**, i.e. Building Polygons (HU-DE) and 3D Building Models (LoD1-DE) for the entire Germany from the national Geodata Infrastructure (GDI) became available only in 2015. Geospatial data digitally represent the **real estate cadastre** and are updated annually.
- These geospatial data comprise all existing buildings (with a certain time delay). They can serve as **sampling frame** and allow for the drawing of a **representative sample** in any building stock. The standard error can be specified, thereby **quantifying the reliability** of the information.
- Geospatial data allow for a geometric definition of the **renovation rate** as the ratio of **renovated over total component surface area** in the stock.
- Geospatial data were not originally intended for this application. In particular, current address data of property owners is not well-maintained and is generally not accessible for research purposes due to privacy concerns. This made **extensive data processing** and an **on-site screening** necessary.



4 What is *representativeness*?



1. If pure chance decides on the inclusion of a sampling unit into the **sample s**,
2. (1st order) inclusion probability of all sampling units (SU) in the sample is known and
3. every population unit (RO) has a chance, i.e. a positive inclusion probability, to be included in the sample

then **unbiased result estimates of parameters of the population U from the sample s** with a quantifiable and lowest possible sampling-related uncertainty.

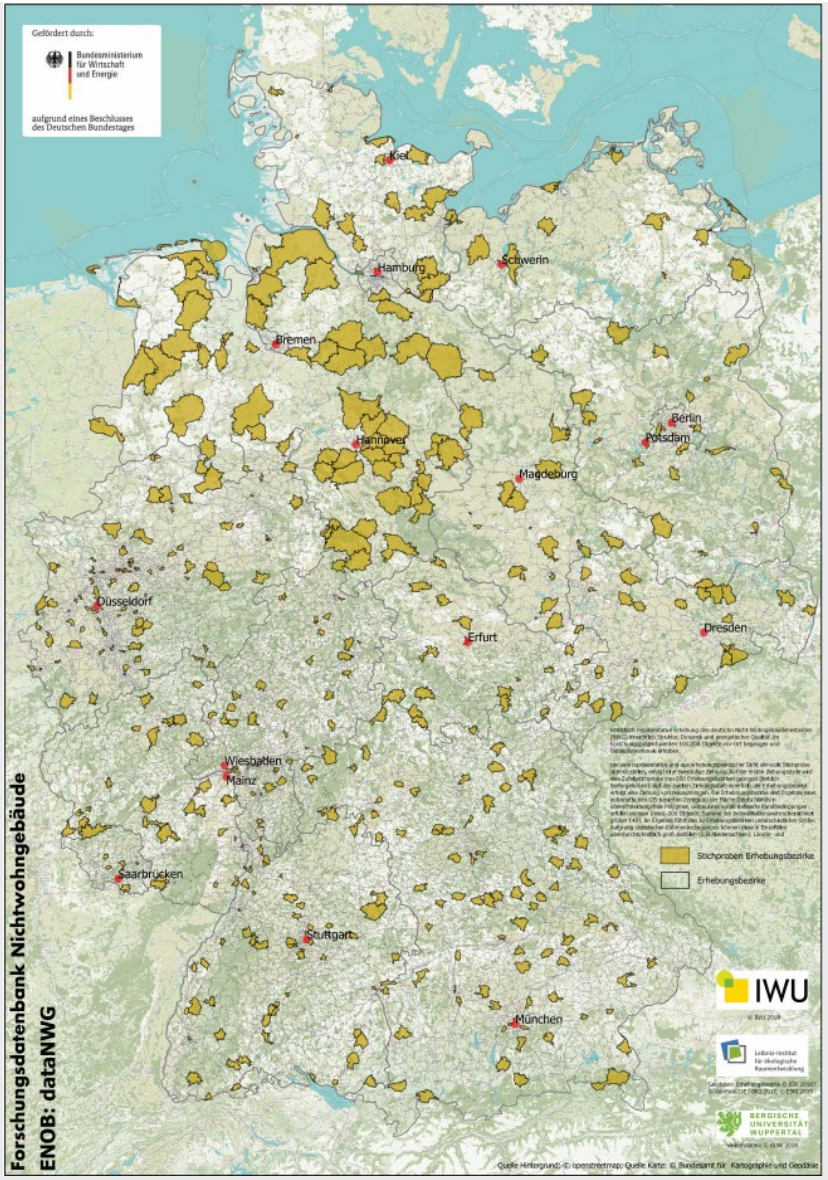
This is what we call a representative sample survey in inductive statistics. Geospatial data products, like HU-DE / LoD1 in Germany, as sampling frame provide all necessary features.

5 Are there no other data sources?

- Unlike residential buildings, non-residential buildings are not included in the **census**.
- **Official statistics** in Germany (destatis) and Europe (eurostat) cover only parts of the NRBS. **Building registers** exist only in a few EU MSs, like Austria, Finnland and Denmark.
- If at all, then some **structural data** on the non-residential building stock is provided, like the number of building permits per year. **Energy-related characteristics** are not available. Owner addresses are subject to data protection.
- According to the EPBD, EPCs are to be issued for new buildings and in case of resale or reletting only. Renovation is not necessarily a trigger for the issuance of a new EPC. EPCs are valid for 10 years, regardless of whether an energy-related renovation has taken place or not. Hence, **EPC databases are a highly biased sample and therefore not "representative"**.

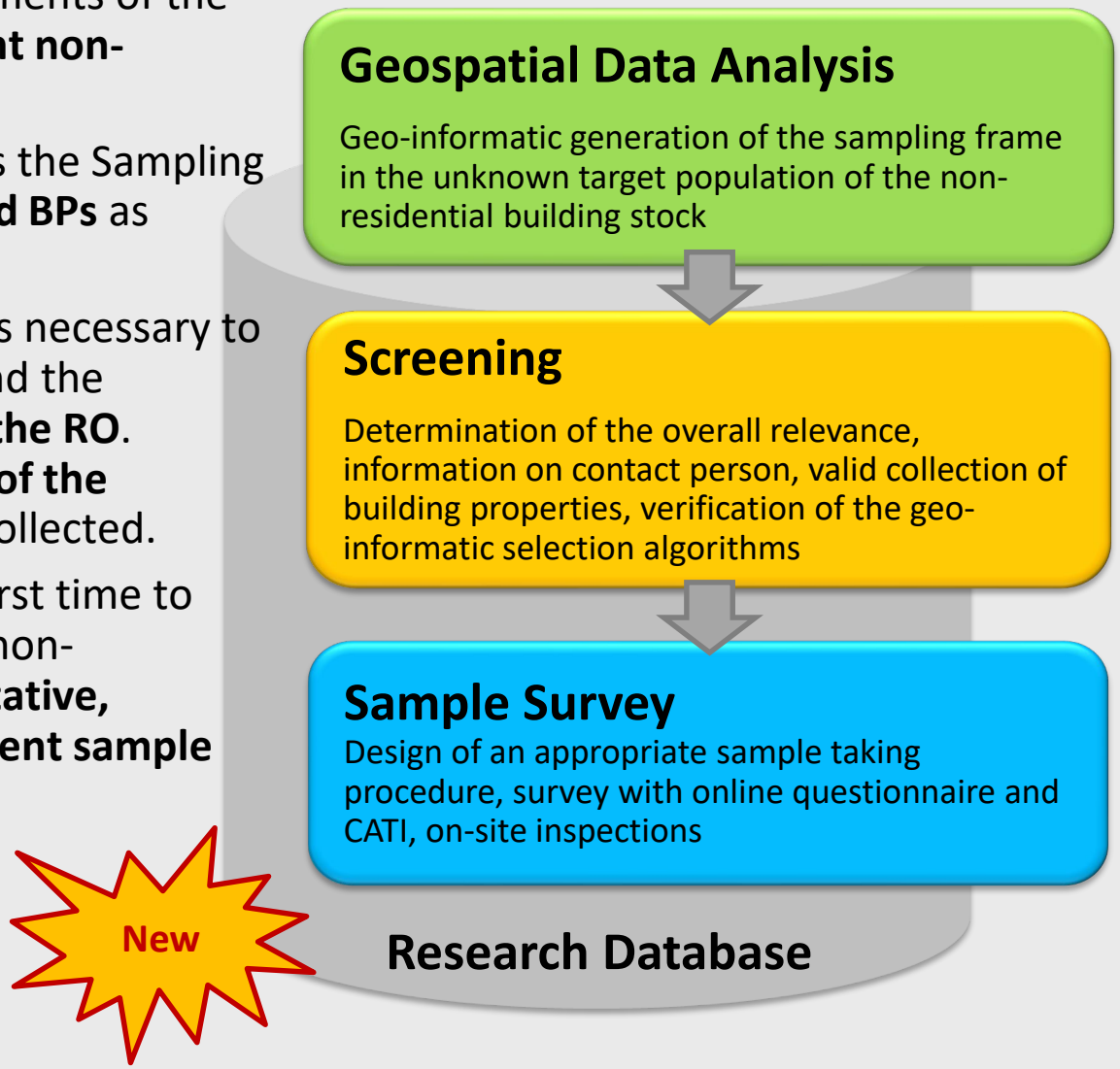
6 Sample surveys: Aren't they too inaccurate?

- The accuracy can be influenced by various factors, including the **size** and **representativeness** of the sample design, the **sampling method** and the quality of the **data collection**.
- **dataNWG: Two-stage stratified sample taking** with 500 districts proportionally stratified by states and planning regions (Raumordnungsregionen) and 200 BPs per district, disproportionally stratified by 5 relevance probability classes
- If the methods of inductive statistics are strictly adhered to, the **statistical error can be quantified** and the reliability of the results made assessable.
- Carefully designed questionnaires and good quality assurance as an integral part of the survey at all stages **minimise non-statistical errors**.



Project Design

- **Research Objects (RO)** are the elements of the Target Population, i.e. **BEA-relevant non-residential buildings**.
- Geospatial Data Analysis generates the Sampling Frame consisting of **geo-referenced BPs** as **Sampling Units (SU)**
- **Screening** of the buildings on site is necessary to identify the **relevance of the SU** and the **relationship between the SU and the RO**. Furthermore **contact information of the building owner** or user has to be collected.
- This approach enables us for the first time to explore the sector of the German non-residential buildings in a **representative, regionally balanced and cost efficient sample survey.**



8 From Sampling Units to Research Objects

1. Geodata Analysis

- Processing of 52 Mio. BPs (elimination of micro polygons)
- Adding LoD1 Building Function and 40 geometric and semantic attributes

2. Sample Taking

- Two-stage disproport. stratified sample
- 500 Districts per 200 BPs

3. Screening

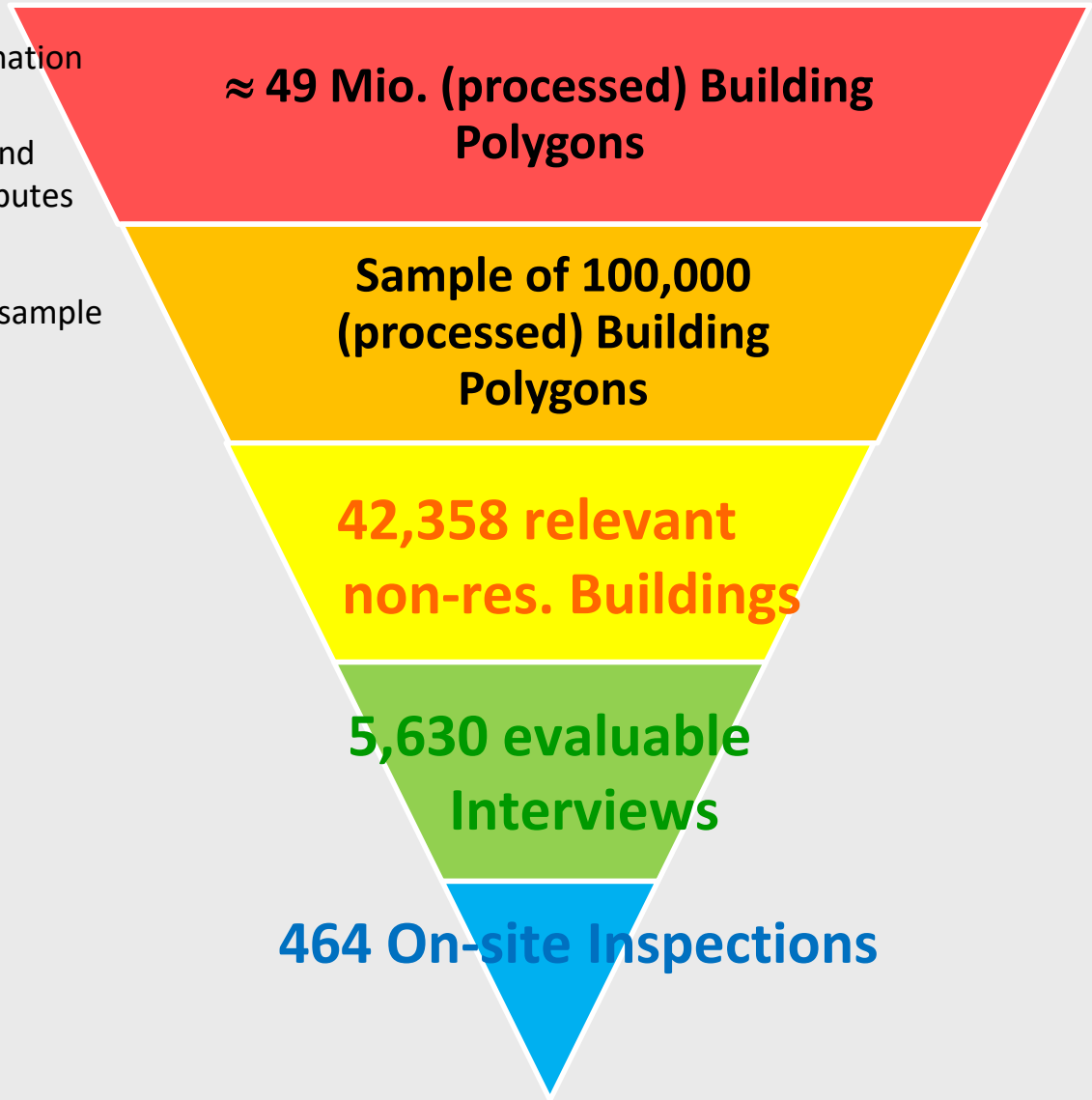
- Relevance
- Relation polygons - buildings
- Owners‘ adresses
- Basic building characteristics

4. Interviews

- Structural attributes
- Energy-related attributes,
- Owner characteristics
- Facility Management

5. On-site Inspections

- Measured consumption
- Calculated demand
- Usage parameters



9 Summary: The Sampling Design works!

- Geospatial data constitute a **suitable sampling frame** for a representative sample survey to explore the formerly unknown stock of non-residential buildings in Germany.
- **Screening** on site is a necessary step to relate BPs as SUs to buildings as ROs.
- **Two-stage, disproportionately stratified Sampling** with 500 Survey Districts as PSUs with 200 Building Polygons as Secondary Sampling Units (SSU) each make a good sample.
- **Response rates** in the interview phase turn out to be **sufficient** in order to do meaningful statistics with reasonable sample sizes (and standard errors)
- About 50% of the respondents in the interviews are interested in an on-site inspection.
- **The Sampling Design works!**

10 Transfer to other EU MSs and the BSO

- Most EU MSs also lack data on NRBSs (and often on RBSs also), but **cadastral data** are presumed to be available everywhere and form the basis of digital geospatial data. With the end of the implementation phase of the EU **INSPIRE Directive** by the end of 2021, all EU MSs should have their geospatial data infrastructures cross-border compatible and shareable.
- In principle, so it is conceivable to **transfer the survey methodology** described here to other EU MSs in order to carry out representative sample surveys in building stocks with reasonable effort and sufficient accuracy.
- We propose to conduct a **preliminary study** to elicit the availability and quality of geospatial data in the EU MSs and, if necessary, to describe the steps needed to make it usable for regular data collection to close the data gap in building stocks.
- A **Building Stock Observatory (BSO)** with representative data and regular updates on the structures and energy-related properties of national building stocks would be the ideal foundation for successful climate protection policies in the EU building sector.
- In addition to its function in the census, the future **Building Register** should also be able to serve as a sampling frame for sample surveys (-> Enable owner adress transmission from municipal property tax offices, for example!).

Research Database Non-residential Buildings

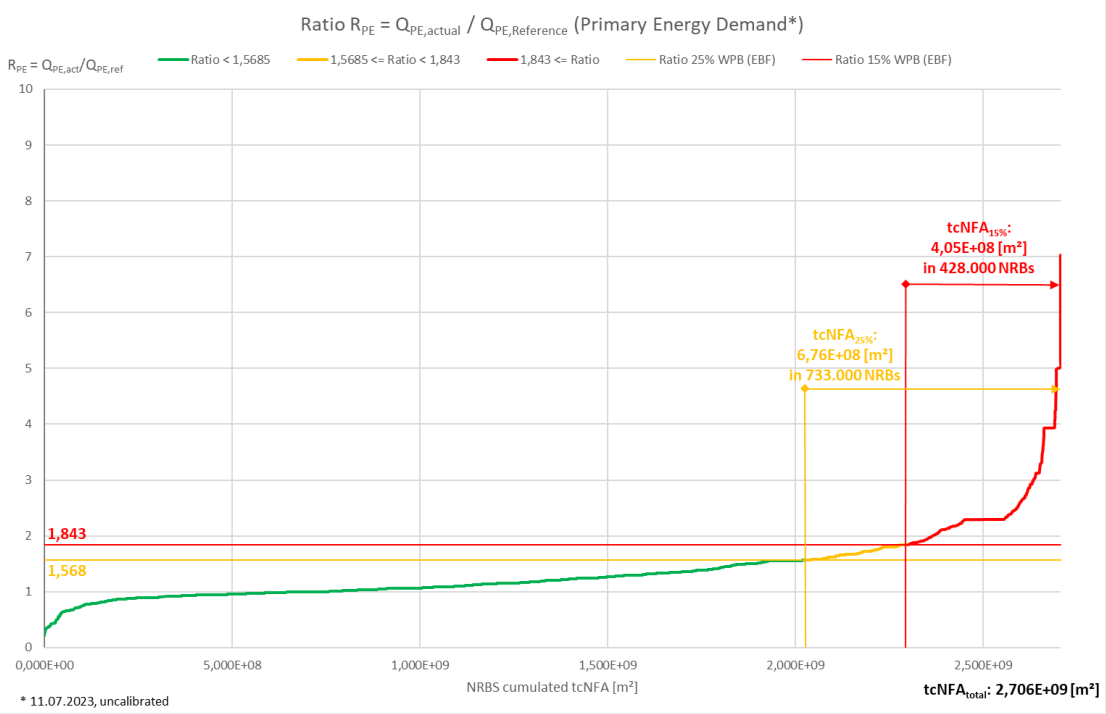
(www.dataNWG.de)

Exploring an Unknown

Source: AdobeStock

11 MEPS for NRBs

- Definition of efficiency classes
  - Metric definition: useful, final or primary energy demand depending on the intended effect taking into account the wide range of building functions (reference building approach)
  - Metric calculation by reduced order energy performance simulation tool DIBS
  - Figure 1: 15% / 25% percentile, area weighted -> metric threshold value
- Proof of fulfilling the obligation
  - Germany: In a short period, several hundred thousand pieces of evidence must be provided.
  - A calculation method with low effort and a small number of model input variables is essential.



7a Definition of „single“ building

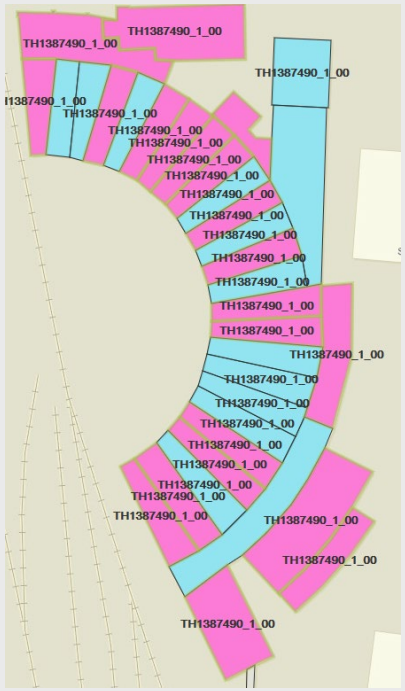
Non-residential Buildings are dedicated to non-residential uses on more than 50% of their useful floor area.

Single Buildings are detached buildings as well as those that consist of building parts having been built based upon an integrated architectural concept at the same time plus retrofitted parts that are to be assigned regarding access and function, because they cannot be used independently.

All building parts must be structurally connected on the ground.

In case of doubt, salability is another criterion for determining which parts belong to a single building.

Relevance: Thermally conditioned, i.e. heated and/or cooled, non-residential buildings that are subject to the Building Energy Act (BEA).



4a What is *representativeness*?

- In the sense of sampling theory, this means that the sampling model has to create the basis for (at least approximately) ***unbiased*** (in German: *unverzerrt* or *erwartungstreu*, respectively) **result estimates of parameters of the population U** from the sample *s* with a quantifiable and lowest possible sampling-related uncertainty.
- Three conditions must be met:
  - *All NRBs k in the population U must have positive inclusion probabilities, $\pi_k > 0, \forall k \in U$.*
  - *Sampling must be carried out at random.*
  - *The inclusion probabilities π_k of every NRB $k, \forall k \in s$ included in the sample s must be known.*
- In the case of identical inclusion probabilities $\pi_k = \pi = \frac{n}{N}, \forall k \in U$, with n the sample size and N the population size.
- Unbiased estimates of the standard errors: The inclusion probabilities for any two NRBs to be included together in the sample must be positive and known.
- **HU-DE / LoD1 as sampling frame provide all necessary features.**