



## National Buildings Database Prj\_1821 - Q&A

**21<sup>st</sup> June 2023**

This document answers questions raised during the ITT Q&A window of the National Buildings Database Programme. Questions have been anonymised.

*Any questions coming in after the deadline, which DESNZ deems to be of material significance will be answered and shared with all suppliers.*

Clarification to the ITT Document:

The equation used to calculate the score for each bid against criterion 4 (Price) will be as follows

$((\text{Maximum Budget} - \text{Bid Price}) / (\text{Maximum Budget})) \times 15\%$

1. For surveys that contain 'sector specific features/questions' - how would this be stored in the final database? For example, if the required fields in the final database (as stated on Page 28 of the ITT) are:
  - § Building use
  - § Building ownership
  - § Floor area
  - § Public Sector flag
  - § Connection to gas grid (and location of meter)
  - § Fabric construction
  - § Location
  - § Geometry
  - § Energy consumption
  - § Primary heating system type
  - § Secondary heating system type
  - § Heat distribution system type
  - § Back-up generation system
  - § Grid connection capacity

How do you see the responses for survey questions about 'running hours' or 'end-uses' (WP1 - page 23) being stored in the final database, as they are not directly attributed to the above list of fields but useful for future modelling - especially those that are qualitative ('end-uses') and not quantitative (operating hours).

**A: Bidders can propose additional data sources and inputs but it should be noted that any outcome of this phase of the project will need to integrate the existing work done in phase 1 for the hospitality sector and so must still be compatible or have a linking variable (e.g. UPRNs) even if there are added variables, or new data structures.**

2. Are you able to provide details on the sample size, statistical significance, confidence level and margin of error associated with the surveys distributed during the pilot phase?

**A: The purpose of the surveys and targeted collection of data is to supplement analysis and engineering of existing datasets. Past survey approaches in the non-domestic building**



**stock have run into challenges with representativeness of samples as a result of pursuing specific sample size targets, due to response bias within sectors. Therefore a set sample size and confidence level have not been specified for this project, instead the focus should be on guaranteeing representation of building sub-sectors.**

3. How was the survey distributed to respondents? Do you have a preference for distribution methods and are there any services (e.g., existing partners, stakeholders, contractors) available to DESNZ that we have access to?

**A: In Phase 1 the majority of surveys were carried out remotely (by telephone) with a small number of follow-up on-site 'verification' surveys. It is worth noting that this project was in part influenced by Covid restrictions in place during the planning and roll-out of surveys. We do not have a preference for how surveys are distributed, but also do not have existing services that we could make available for this.**

4. How many onsite surveys were carried out during Phase 1?

**A: Phase 1 carried out 25 onsite 'verification' surveys for the hospitality sector and an additional 246 remote telephone surveys.**

5. Are there any legal obligations that require respondents to fill out the surveys?

**A: No – there is no legal obligation for respondents to fill out the survey.**

6. For the pilot study, what are the 'geometric models' referred to when describing the mathematical estimation of missing floor area on Page 19 of the ITT document under the 'Hospitality Sector' heading, and how were they used to derive a floor area even when not recorded by the VOA?

**A: The method developed by the contractor to derive floor area is outlined below.**

***By bringing all of these datasets together and applying a hierarchical approach (illustrated in Figure 1), for each CaRB3 activity, it is possible to derive floorspace figures for the Hospitality CaRB3 class. The hierarchy essentially funnels towards the best recorded floor space, or calculated floor space, for each premises. The VOA floorspace data (red layer) are generally reliable, with a flag to indicate where this is true. If there is no VOA record of floor area, the algorithm moves to gather non-domestic EPC floor area data (blue), if this is available. If this fails and the model permits it for a few specific activities (such as 'Holiday Home (Self Catering)', the next step is to look for domestic EPC floor area (green). (Note that to avoid inaccurate matching, not all activities are allowed to substitute domestic EPC floor area). If no EPC data have been matched, then the calculated floor area derived from the 3DStock model (gold) is used. If all stages fail, there are no floor area data for the premises (grey).***



**Figure 1 The method of selecting the most appropriate floorspace data for non-domestic premises using CaRB3 activity and available data**

7. Further to 'Mixed Use Buildings' on Page 19, how did the pilot study approach the inability to treat domestic and non-domestic building stocks as discrete entities, and how did they capture this complexity.

**A: The pilot study developed an approach using the 'Self-Contained Unit' approach developed by the contractor. This is described below.**

*Defining the extent of a non-domestic unit is not simple and care is required to avoid including elements that are not contained within buildings or non-domestic activity that does not require a building (like an 'Amazon delivery locker' or a 'Quarry' or a 'Beach Hut' or a market stall pitch). The Valuation Office Agency (VOA) is one of the most reliable sources of non-domestic activity for England and Wales. The VOA is a division of His Majesty's Revenue and Customs (HMRC), responsible for the assignment of a theoretical annual rental value to all non-domestic premises, upon which business rates are then based. The VOA actually deals with 'hereditaments<sup>1</sup>' and for this work, we substitute the more user-friendly word 'premises' for this (NB: the word premises is both singular and plural). Not all premises, such as places of worship<sup>2</sup>, incur business rates and so these non-rateable premises do not appear in VOA data.*

*Whilst the records described in the VOA data (largely) represent space within buildings, they do also include space external to buildings (e.g. land used for storage) and units of valuation that are not area-dependent, such as a 'golf course', or the number of bedrooms in guest houses. The 'external' spaces are excluded from the non-domestic unit calculation within this work, which focuses on activity within buildings. Some 'premises' listed in the VOA are entirely external to buildings such as telecommunication masts and advertising*



*hoardings (although some of these may be attached to a building). These are not included as non-domestic units.*

*It should also be noted that not all premises have records that include data describing physical characteristics of those premises. Whilst shops, offices, factories and warehouses will almost always have data on floor space, floor level(s) or storeys, etc. some important classifications such as pubs, hotels, hospitals, schools and universities, will not have these data on physical characteristics. This situation has an effect upon subsequent analyses of floorspace, storey counts etc. in this work.*

*A common misconception is that premises equate to buildings. This is not always the case and the relationship can be complex. This means that the extent of a non-domestic building might include just one premises, or it might include many premises, as is the case with an office block with multiple tenants, or it may be that a single premises occupies many buildings, as is the case with some schools, hospitals and other activities that take place across a site or campus. A premises can also extend through several adjoining buildings as might be found, for instance, when a retail unit expands from one building to occupy the ground floor of an adjacent building. Added to this, domestic premises may or may not be present within a building that also contains non-domestic activity. This could be flats above a shop through to purpose-built blocks that contain a mixture of domestic and non-domestic units.*

*As a result, buildings are not ideally suited as units for modelling the stock. (Indeed, in many situations it is not easy to define where one building stops and another adjoining building starts.) Non-domestic premises as mentioned can correspond to parts of buildings, whole buildings or groups of buildings on common sites. The activity, floorspace and energy consumption normally relate to the premises, which may or may not be the same as the building. Instead of using 'building' the preferred method is to use the 'SCU' or Self-contained Unit. This concept was devised by Taylor et al. (2014 3) and then adopted by Evans et al. (2017 4).*

*A SCU can be generated by plotting the geospatial locations of VOA premises and then aggregating those that fall within a geometric polygon representing the building footprint. Where possible, the different premises are correctly allocated to the floor(s) that they are located on which builds up a three-dimensional stack of activity within each building polygon. When activities span multiple building footprints, such as when a premises is located across multiple (historic) addresses, such as '7 -9 The High Street' then the SCU boundary expands from a single building footprint to encompass the neighbouring building footprints as required. An example is shown in Figure 1. The final SCU boundary does not divide the premises but instead it wraps a polygon around the addresses, map polygon (building footprint) and premises and treats this as a single unit. The self-contained unit provides a well-defined unit for analysis.*



**Figure 1** The assignment of a self-contained unit (SCU) around a ground floor shop that spans two buildings (e.g. 7 – 9 The High Street) with domestic flats above. The SCU is designed so that it does not divide any premises and treats the enclosed objects as a single unit.

*It is too computationally intensive (in terms of computer processing and storage space) to produce a full SCU-based model of England and Wales within the constraints of this pilot study. Instead, a georeferenced model has been built based on the matching of premises to building footprints (and vice versa for multi-building premises), to provide a pilot model that is geospatial, but does not include the detailed 3D models of each SCU.*

8. Would you like to see the survey results that were saved as a supplementary database during the pilot phase to be merged into the primary database in this project?

**A: If a proposal is made for a means to do this which increases the completeness of the database without compromising representativeness or robustness of data it would be an acceptable approach to propose in the bid.**

9. WP2 (Stakeholder engagement) on Page 24, suggests that prominent sector organisations can help to facilitate recruitment. Would this not introduce bias (based on their access to, and knowledge of, respondents) compared to recruitment of respondents based on random sampling of sufficient sample size of addresses in the specific sector?

**A: Yes it is correct that engaging prominent sector organisations could introduce bias, but for some sectors failure to engage key organisations can act as a barrier to recruitment and introduce a self-selection bias as there is no legal obligation for respondents to complete the survey (see response to question 5), and thus in the past random sampling in some**



**non-domestic buildings sectors has produced heavily biased samples. The bidder will need to consider the balance of these concerns in their approach to stakeholder engagement.**

10. In the final paragraph for WP1 (Data Collection) on Page 24, what 'sample recruitment details' will need to be shared with Department Analysts?

**A: These details would likely be contact details for sites surveyed to facilitate recruitment for any future research the department may decide to undertake.**

11. What format (units) are the following representations required to be in: 'Location', 'Geometry' and 'Energy Consumption'. (Work Package 6 Page 27).

**A: Location information should be provided as both an address and geographical coordinates. The Geometry data used in Phase 1 was derived from the National Polygons Dataset with added floor height information. Information about the National Polygons Dataset can be found here: <https://use-land-property-data.service.gov.uk/datasets/nps/tech-spec/1..> Energy consumption should be provided at a minimum in kWh/year for a premises (and floor areas will enable this to be converted to Energy Use Intensity kWh/year/m<sup>2</sup>).**

12. There is reference to Financial Standing. Apologies if I have missed this in the documentation but with regards to ITT -

"Part 3: Selection Questions

5.4 Where we have specified a minimum level of economic and financial standing and/or a minimum financial threshold within the evaluation criteria for this procurement, please self-certify whether you meet the requirements set out."

Can you confirm if a level of minimum financial standing set out in the evaluation criteria?

**A: There is no minimum level of economic and financial standing and/or a minimum financial threshold within the evaluation criteria. However please note that DESNZ use Dun & Bradstreet or Spotlight, the government's online automated due-diligence tool to assist them with their financial due diligence and will request Dun and Bradstreet/Spotlight to provide comprehensive reports on the preferred bidder/s. DESNZ will review the Dun and Bradstreet/Spotlight report prior to notifying bidders of the result of the competition and may need to check [with bidders] that the information within the report is correct. DESNZ may also request the latest accounts and financial information from the preferred bidder/s. Suppliers assessed with a high financial risk status may not be awarded a contract - at this stage we will revert to the bidder to discuss further.**

13. Which 'Work Package' covers identifying, analysing and linking additional datasets (step 2 in the Pilot methodology) that are not sector specific? What factors will DESNZ consider when agreeing to use datasets suggested by the Supplier?



**A: This depends on whether these are datasets applicable to all sectors or a specific sector. If it pertains to datasets relevant to the whole building stock this would be accounted for in 'Work Package 3', whereas if it sector specific (e.g. dataset on schools) this should be accounted for in Work Package 5'.**

14. Have you identified the stakeholders as described in Work Package 2 or will we need to do this?

**A: No we have not identified the stakeholders described in Work Package 2 – this will be a responsibility of the successful bidder. DESNZ will be able to provide contact with some previously engaged stakeholders but this will not be sufficient for the engagement described in Work Package 2 and the successful bidder will need to identify further stakeholders.**

15. Have you identified the members of the Steering Group or will we be responsible for recruitment?

**A: No we have not identified the members of the External Steering Group – this will be a responsibility of the successful bidder.**

16. Can we choose to not perform on-site building surveys if the remote survey approach proposed is sufficient?

**A: Yes – if a remote survey approach can be shown to be sufficient when combined with other datasets and analysis to fill data gaps and provide validation of the final database.**

17. For Evaluation Question 01a, if we're proposing to use the same methodology as broken down by your suggested work packages, can we link to answers to questions 01b 01c 01e etc for further details regarding survey, handover and updating/sampling the database phases respectively?

**A: Yes this is acceptable where necessary to avoid redundancy, but the key argument/case in response to the each of the questions must be in the answer submitted to the respective question.**

18. What is the likely number of buildings per region that you anticipate will require a physical survey?

**A: Physical surveys do not need to cover all regions if it can be shown that the data to be collected and variables of interest are not likely to vary significantly across regions. For example, standardised 'drive through' fast food outlets are unlikely to vary significantly in layout or construction across the different regions.**

19. Is there a Specification for the building survey report?

**A: No – although findings from sector specific surveys, data collection, and modelling should be reported in the sector deep dives in the main output report detailed on page 29-30.**



20. On page 16 the ITT document states that "Phase 1 also scoped out available input data resources for the wider non-domestic stock and devised a methodology that eliminates the need for routine surveys with the potential for a live database that can regularly update as the data resources it draws upon are updated. All outputs from the now completed pilot study will be made available to the successful bidder for use in this project. The project will roll out the approach piloted in the first phase to all other non-domestic sectors, and include a representation of domestic buildings created from existing available data."

On page 18, the ITT document states that "The methodology for the development of this database will be based on the approach developed in the Non-Domestic Building Survey Pilot study. The successful bidder will be provided with code and methodology for data engineering and linking of existing datasets, which they will have to use."

Question: can the department please provide immediate access to the code and methodology for data engineering and linking of existing datasets, which the successful bidder will have to use? Given the obligation to use this material, a detailed review is necessary for pricing and scheduling the work appropriately. Without access to this (and time to review it), bidders may be placed at a disadvantage.

**A: The information provided in relation to the methodology of phase 1 is considered sufficient to allow prospective bidders to scope and price their proposal to a sufficient level of detail. Some of the code and data is associated with sensitive material which we cannot share with bidders outside of a contractual relationship. We have provided some additional information with these questions for added clarity, namely the remote survey questionnaire, a simplified list of datasets used in the phase 1 database, and a list of key variables in the outputs for the hospitality sector in Phase 1. See additional separate documents.**

21. Page 15 of the ITT states : "The survey sample size should be such that suitable confidence can be had in achieving the aims of the project aims, with the number of surveys determined by bidders in accordance with their approach"

Question:

The purpose of the site surveys is to inform the modelling of values which cannot be obtained from the bulk or sector-specific datasets (validation is an equivalent exercise). To size and cost this effort appropriately requires an understanding of the minimum acceptable model fit for these predicted values. Could you therefore please quantify the meaning of the phrase "suitable confidence" here, in terms of minimum acceptable model fit for each sector? Pursuant to this can you also please set out the number of surveys UCL carried out in phase 1 for the hospitality sector, the fit of the model(s) trained using this data, and whether this was deemed acceptable?

**A: This interpretation is not quite correct. In the case of Phase 1 the survey results were not used to directly achieve a model fit, but rather inform assumptions and verify sub-sector variation in the hospitality sector. 'Suitable confidence' in the context of this database is not about statistical confidence but rather reliable representation of the broad range of sub-sectors within a non-domestic sector (for hospitality this includes pubs, restaurants, cafes, hotels, etc.). This being said should a bidder propose a robust, justified approach to use model fit for sub-sectors with survey and data collection to provide a**





**more quantitative validation of results this would also be a valid means of satisfying the requirement for 'suitable confidence'.**

22. Question: Could you please explain how the 24 page Tender page limit will be calculated given that individual documents for individual questions have been requested via the portal (the online procurement platform, through which we are being asked to submit). The portal requests a separate attachment for each of the categories (e.g Methodology: Database Development Approach and so on). How can we ensure we meet the overall page limit criteria?

**A: The page sum of all the responses to all of these questions must be no more than 24 pages. How this is divided between answers is up to the bidders.**

23. Question: Under Data Protection / Section 13 on page 39 it says: "The Personal Data will be retained by the Contractor for a 6 month retention period, following which the Contractor will provide the Authority with a complete and uncorrupted version of the Personal Data in electronic form (or such other format as reasonably required by the Authority) and erase from any computers, storage devices and storage media that are to be retained by the Contractor after the expiry of the Contract and the Contractor retention period. The Contractor will certify to the Authority that it has completed such deletion." Do you expect to receive the personal data from the surveys and what therefore are the obligations for data controllers within the contractor and the department?

**A: Contact details for buildings/company contacts may be shared with the department. If this is the case and there is indeed personal information requiring appropriate GDPR controls then a suitable data sharing agreement will need to be put in place. It is conceivable that business contact details for surveyed buildings would suffice and would not entail sharing personal data. This will be discussed with the successful bidder at the start of the project and DESNZ will work with the successful bidder to put a suitable, compliant arrangement in place.**

24. Clause 1.3.2.2 of Schedule 36 requires that Supplier Existing IPR is provided to the Buyer 'to use, sub-licence or commercially exploit (including by publication under Open Licence) the New IPR and New IPR Items'. Existing Supplier IPR will be used to process and analyse data, some of which is confidential, and the release of this Existing Supplier IPR in this way will serve no purpose because only the Buyer (and perhaps related public sector bodies) will have access to this confidential data. We therefore request that this clause be modified to remove the requirements for 'sub-licence or commercially exploit (including by publication under Open Licence)' as they relate to the Supplier Existing IPR.

**A: Supplier Existing IPR will only be used where necessary, and we will keep in the sub-licence term in the case that the need arises e.g. other government bodies who we need to sub license to, who also have access to the data.**

25. Is it possible to make available any of the outputs from the pilot phase?



**A: Unfortunately, due to the sensitive nature of some of the information in the outputs we cannot share these in their entirety with bidders outside of a contractual relationship at this time. We have provided some additional detail on the method in response to questions namely the remote survey questionnaire, a simplified list of datasets used in the phase 1 database, and a list of key variables in the outputs for the hospitality sector in Phase 1. See separate, additional documents.**

26. Is it possible to give an indication of the sample size identified by the pilot methodology for each of the CaRB3 classes or in general as a reference point?

**A: Phase 1 did not specify a set sample size for surveys of each of the CaRB3 classes to be used in this phase of the project. Page 21 of the ITT provides an approximate breakdown of CaRB3 class size by premises count.**

27. Can you provide details on the survey sample size used to validate the pilot phase Hospitality CaRB3 Class?

**A: See response to question 4.**

28. Can you provide details of the surveys used to validate the pilot model (physical / remote questionnaire)?

**A: See response to question 4 for survey sample sizes. The survey mainly comprised remote surveys with some additional verification surveys.**

29. Does the pilot process validation methodology include a requirement for physical on-site surveys to take place or is this up to the bidder to define?

**A: The bidder is free to propose the breakdown of surveys and data collection as long as this can be justified as capable of fulfilling the role of completing & validating the information in the database. There is not strict requirement for on-site surveys.**

30. Given the number of buildings that data requires collecting from, we feel that gaining a significant sample size from physical onsite surveys alone will be challenging to achieve within the budget. Please may you advise on the liability of the supplier for the accuracy and completeness of data provide by building owners/occupiers, ie if they do not provide sufficiently reliable or sizeable data to complete the database?

**A: This is a challenge which prospective bidders will need to balance – the successful bidder will be responsible for ensuring a reasonable level of quality and accuracy of data collected, and resulting impact on ability to complete the database. Phase 1 combined remote surveys with a smaller number of onsite surveys to ensure robustness of the sample. Combining different data collection and gathering methods will likely be needed. Suppliers will need to set this out in their bid – this is part of the evaluation criteria.**

31. The ITT docs states the proposal should be no more than 24 pages, however the portal requires a separate attachment for each question. Could you advise if you have a preference



to page limits for the individual questions?

**A: There is no preference on how to split the 24 pages limit across each of the questions.**

32. The ITT references a full three-dimensional model of the building stock. Could you provide details of the technology stack used to produce this output and examples of this output?

**A: The 3DStock Model is based on the approach developed by UCL and has been published in open-access peer-reviewed journals, including this article which outlines the approach: <https://www.tandfonline.com/doi/full/10.1080/09613218.2017.1410424>**

33. With regards the treatment of mixed-use premises, the ITT acknowledges that differentiation between non-domestic and domestic properties was not possible. Please could you clarify how this is envisioned to be dealt with going forward?

**A: The pilot phase implemented an approach (described in response to question 7 above) for identifying mixed-use buildings. This outlines how identification of mixed-use buildings was used in cases where a building was not wholly comprised of domestic or non-domestic buildings but rather a mixture of the two. This is the approach which is to be used for this project.**

34. The ITT references that the total floor area calculated using the pilot methodology (EPC data and geometric models) resulted in a four-fold increase compared to the VOA dataset. This is a very significant difference in the data, could you confirm if similar discrepancy is anticipated in the remaining classes and whether this discrepancy has been validated in the hospitality class data set by the pilot phase surveys?

**A: Discrepancies In floor area between the VOA dataset and the database will vary by sector and could indeed be quite significant for some sectors. The work undertaken in phase 1 found that some discrepancies related to issues with definitions and classifications of floor area for certain property types which may not necessarily match the floor area that needs to be considered when estimating energy consumption.**

35. With reference to the simplified domestic sector data modelling, could you confirm that survey validation of this data set is not within the scope of this commission?

**A: Yes this is correct – this project will not involve any surveys of domestic premises.**

36. Our understanding from the documentation you have provided, indicates that there is an existing data model from the PoV phase that the supplier will need to extend using PostgreSQL. The extension is required to cover the remaining 14 non-domestic sectors. Please can you confirm this understanding correct? How will the supplier access the existing data model? Will each sector have variant extension requirements?

**A: This understanding is correct. The data model is contained in a bundle of SQL and python scripts which would be shared with the successful bidder as a file transfer. There will be some difference from sector to sector (including input data and assumptions), but the basic approach should be extendable across all sectors.**



37. With Reference to Section 2, sub-section 5 of the ITT Publication “The successful bidder will be provided with the code and methodology for data engineering and linking of existing datasets, which they will have to use.”— will the linking of the dataset extend to the 14 non-domestic building OR is it just for the building type in scope of the PoV? (i.e. do you expect the provider to do the linking of the 14 non-domestic buildings)?

**A: The code to be supplied will be applicable to the other 14 non-domestic building sectors although there will be a need to integrate in different sector specific datasets (e.g. data specifically on school buildings for the education sector).**

38. What data model technique e.g. relational, snowflake etc. that was applied during the pilot phase? Please also confirm that this remains the adopted technique for this work?

**A: The pilot phase applied a relational approach and that will remain the approach for this phase of the project too.**

39. What are your expectations in relation to the data lineage for the final database?

**A: We would expect data flows in the final database to be well documented at a minimum and provenance, transformations, and analyses applied to each variable should be clear enough to enable reproducibility by DESNZ analysts.**

40. In the RfP document it was noted that the DESNZ analysts maintain a Front-End to access and update the database. Furthermore one of the deliverables is to provide information on how the database will be updated. Please can you clarify more detail on these two statements?

**A: DESNZ analysts will maintain and update the database using PgAdmin on the department’s internal secure cloud computing facility after handover of all deliverables at the end of the project. The successful bidder will need to provide instructions and code to enable an update to be run by DESNZ analysts on a regular basis.**

41. To enable us to fully and effectively answer all your questions and build a proposed solution, it would be beneficial to have access to the pilot outputs, methodology, and technical architecture to ensure your requirements are being aligned to these. Specifically we would like access to:

- The outputs from the Phase 1 pilot- the reference architecture and detailed architecture documentation for the pilot
- the data model (conceptual, logical and physical) linking all the data sets in scope of the pilot
- user access management model
- delineation of responsibilities

**A: We have provided some additional information on the survey questionnaire (attached pdf) used for the pilot study and a simplified breakdown of data sources and variables.**



**Unfortunately, due to the sensitive nature of some of the information in the outputs we cannot share further outputs with bidders outside of a contractual relationship at this time. The information provided in the ITT has been deemed sufficient for a prospective bidder to scope and price a bid to an appropriate level of detail.**

**In terms of delineation of responsibilities bidders will be responsible for all the tasks laid out in the methodology indicative work packages in the ITT (pages 23-28) except where explicitly stated that DESNZ will share responsibility (as is the case for setting up data sharing agreements in WP3, reviewing outputs, and determining database formatting).**

42. With reference to the Objectives detail of Section 2, sub-section 3 in the ITT Publication, "outputs that are accessible to other government departments and external researchers" - is it intended that the database will be accessible by the public and commercial organisations or for internal/governmental agencies only? Are there any functional or technical requirements detailed for this at this stage?

**A: The methods for making outputs that are accessible to other departments and external researchers must be able to generate samples of the database that can be shared with the public as well as government departments. Part of the research to be undertaken in this project will involve determining the appropriate geographic (or other) level of aggregation for doing this while providing the highest spatial resolution possible.**

43. In addition to the above, Objectives portion of Section 2, sub-section 3 in the ITT Publication, "outputs that are accessible to other government departments and external researchers" - is there a requirement for the provider to design an external portal to access the database or is the database accessible internally only (i.e. the external researchers are behind your firewalls)? Could you clarify whether or not the final database will be accessed/available in the public domain?

**A: No, the successful contractor is not expected to produce a portal for external access to the database. The database itself will be internal due to data sharing limitations, hence the need for a method of generating shareable, or synthetic samples that can be made publicly available.**

44. Has the 18-month timeline - outlined in section 1.A of the ITT Publication - been set to align with the level of internal DESNZ resource available to support this project or are other factors determining this timeline?

**A: The 18-month timeline has been aligned to availability of funding. Internal resource required will be made available as needed to support this project.**

45. There are two versions of the Mid-tier contract, of which v1 says do not use. Can you explain the specific changes that have been made to v2 and on what basis?

**A: Please only refer to v2. The reason for the second addition was that in the first version Schedule 36 – IPR was not visible. This has been subsequently fixed, there are no other changes.**

46. The methodology set out in the ITT is prescriptive in nature. We have identified that certain datasets have not been included which could be used to create a more effective methodology to addressing the objectives, however this would mean using a different



methodology. The ITT suggests that the bidder has the initiative to suggest our own survey methods but not the original data linking methodology. Can you clarify what level of deviation from the methodology would be acceptable?

**A: Bidders can propose modification and improved sources or data structures, but it should be noted that any outcome of this phase of the project will need to integrate the existing work done in phase 1 for the hospitality sector and so must still be compatible or have a linking variable (e.g. UPRNs) even if there are added variables, or new data structures.**

47. The ITT suggests that all code and methodology from phase 1 will be made available to the bidder. Can you confirm that the methodology is robust and adequately detailed to be transferred and adopted by any supplier to mobilise the work effectively and efficiently or are there important off-model details that are the IP of the phase 1 supplier, which would require some level of rework?

**A: The methodology is robust and is based on peer-reviewed academic research. The code is reproducible although will require some work by the successful bidder to update filenames and paths relevant to the successful bidder's IT system. This will depend on the bidder's computing facilities, for example if working on a cloud computing facility such as AWS data storage access (with S3 buckets) will differ compared to using a local server (with a path to a file storage location). Securing access to the datasets will require putting in place data sharing agreements and transferring large amounts of data which may take a period of a few weeks to fully complete.**

48. Is the winning bidder obligated to roll out the exact methodology piloted by UCL in phase 1, or are modifications possible if improved sources or data structures can be identified? Please can you share a greater level of detail on the model generated in phase 1, for example a sample of the output giving an indication of the level and complexity and types of steps involved.

**A: Bidders can propose modification and improved sources or data structures, but it should be noted that any outcome of this phase of the project will need to integrate the existing work done in phase 1 for the hospitality sector and so must still be compatible or have a linking variable (e.g. UPRNs) even if there are added variables, or new data structures. The attached spreadsheet provides a breakdown of data sources in the database produced by the pilot study phase for the hospitality sector.**

49. With regards to Table 2, Key Data Sources (p.22), please could you confirm the full list of datasets utilised in the Phase 1 Pilot Study, and the source(s) of the LIDAR dataset?

**A: For added clarity the full list of datasets and sources used in the Phase 1 database are in the attached spreadsheet which also provides a summary of key variables in the pilot output for the hospitality sector. The LIDAR data used comes from the LIDAR Composite DTM (Digital Terrain Model) – more information can be found here:**  
<https://environment.data.gov.uk/dataset/13787b9a-26a4-4775-8523-806d13af58fc>

50. What identifier scheme is intended to be used for identifying buildings (as opposed to premises/addresses/hereditaments) within the database? Or is this still to be defined?

**A: The method for identifying buildings uses the 'Self-Contained Unit' approach outlined in response to question 7.**



51. In relation to the requirement to "extracting synthetic shareable samples", please could you provide greater clarity on this and whether this means making accessible extracting samples of mock data or otherwise?

**A: This requirement to provide a method for generating synthetic shareable samples will involve researching methods for synthetic data generation at a level of spatial aggregation, and proposing a method suitable to the buildings database for generating sample of synthetic building-level data for a geographic area (e.g. Local Authority) which are representative of the buildings in that area but where no individual instance in the dataset represents a real individual building in that area. This involves losing some information about the exact locations of buildings but will be able to provide granular samples of building level data for an area which by being synthetic, are publicly shareable.**

52. In relation to Work Package 1, please can you share an example of the survey questionnaire utilised in the pilot study, to complement the database?

**A: Yes – we have provided an example of the survey questionnaire as a separate document.**

53. A lesson learnt from the pilot sector survey includes an Energy and Emission Information pack incentive (p.23). Please can you define what was included in this pack?

**A: In Phase 1 an information pack with information and advice for surveyed buildings on their energy and emissions reduction potential was piloted as an incentive to aid survey recruitment. This was not found to be significantly effective at improving recruitment.**

54. Are you open to suppliers bringing , in addition, a fully formed proposed solution for the domestic sector or should it follow the non domestic methodology?

**A: Bidders may propose any solution for the domestic sector they can show meets the requirements laid out in the ITT, even if it differs in part from the non-domestic methodology, so long as this approach still makes use of UPRNs and can be integrated as part of a database with the non-domestic premises.**

55. Will the Authority consider any changes to the terms? Is it possible to bid subject to contract or are bids fully binding?

**A: Any bid is on the assumption that the terms are agreed.**

56. Could we please request an extension of time for questions allowing time to review CQ responses once received.

**A. Any questions coming in after the deadline, which DESNZ deems to be of material significance will be answered and shared with all suppliers.**