Data Quality Audit.

"Quality of Addresses - EPC address format does not seem to follow the PAF standard format - The new contract should provide EPB register addresses in a standard format. There can be multiple address and UPRNs for the same property. For Open Data and Stats purposes that this can then be matched with other standard address level data sets. No currently available dataset deals with the need for manual addresses requiring to be added and Non Domestic properties can be carved up for the production of EPCs and DECs in a way that is not reflected in any address dataset" -Landmark / Valtech Discovery. - can this be improved by making changes in the entry form? - data quality and audit issues are much wider than simply the address format.

Outside our Remit (these need to be implemented by the schemes in the front end):

1. Controls on what can be added in each field when an assessor is making a lodgement - charts and stats - data should be of the same order of magnitude.

2. Controls on which fields need to be filled - how does the certificate work? - which fields are crucial and which are optional?

The quality of data could be enhanced by implementing validity checks as part of the link between the assessor's lodgment screen and the register - these could be implemented by the MHCLG EPC project without the need to change the input pages developed by/for the schemes.

see CSVs - brisdesc, hartdesc, and carddesc - covering Bristol, Hartlepool, and Cafdiff.

in brisdesc, the maximum for ENERGY\_CONSUMPTION\_CURRENT, ENERGY\_CONSUMPTION\_POTENTIAL, CO2\_EMISSIONS\_CURRENT, and CO2\_EMISSIONS\_POTENTIAL are all orders of magnitude higher than the 75% quartile value. The same pattern is seen for Cardiff and Hartlepool.

Further, for these fields, the minimums are negative values in all three cities.

The high maximums appear to be accounted for by large homes - the maximum floor area in Hartlepool is 40429.48; max. habitable rooms is 21, and max heated rooms is 19.

Meeting with Phil Beschizza - suggested that the majority of data auditing is currently done by the schems rather than by MHCLG EPC team - though ministry team need to audit the schemes to ensure quality.

Main issues apear to be around multiple lodgements in a short time frame, and Phil suggested that assessors may use the register for training purposes, but not be willing to inform the scheme, so that "dummy" or duplicate data may be retained. An audit / report showing number of duplicate records with the same UPRN in a short time frame may flag up data quality issues; the use of the live register for training could be eliminated by the provision of a training area.

A record of the number of failed validation checks, and multiple lodgements which do not follow a set of rules (ie too many with the same UPRN in a short period of time) should be kept, with the data associated with the scheme (and maybe assessor) it came from, and regular reports sent to the Data Team.

If there are particular schemes / assessors which regularly fail this ongoing audit, then they can be flagged either for a review of the 3rd party software and pages they use, or for further individual training for assessors.