Administrative Data | Agricultural Research Collection (AD|ARC)
Synthetic Data

Overview

With permission from data owners, the AD|ARC team have created a low-fidelity synthetic dataset to represent data from the EU Farm Structure Survey, Rural Payments Wales and ONS Census of Population 2011.

Purpose

Researchers can use the low-fidelity synthetic dataset as a training tool, to familiarise themselves with the AD|ARC variables and how they can be used for research. This gives researchers the opportunity to use synthetic data to assess if the real data will meet their needs before applying to access the real data via the Trusted Research Environment (e.g. SAIL databank).

Extra information and limitations

- **Figures are randomly generated within a similar range so do not resemble the original figures.**
- The synthetic data is not suitable for analysis.
- The synthetic data are low fidelity.
- Variable names and labels follow a similar structure to the real variable names but have been amended to include 'syn_' (e.g. syn_equidae).
- Upper and lower limits have been applied to synthetic variables in some cases (e.g. syn_UAA).
- As figures have been randomly generated, totals might not be accurate.
- As the synthetic dataset is low-fidelity, it does not allow for all possible preliminary analyses or statistical testing.
- Means will not accurately represent the real data, and patterns of missingness or relationships between variables have not been preserved.
- Further information is available in the data dictionary.

Contact

Please contact the AD|ARC team if you have any queries or feedback.