

KnowFife Dataset Data Quality Procedure

Data Admin Checks

1. Source Data

Although data taken from external sources is likely to have been checked and had quality and control checks applied, it is still beneficial to quickly scan and determine that the correct data is present.

- Ensure the data and indicators are correct in relation to the data you want to upload to the KnowFife Dataset.
- Ensure you are using data for the correct geography and areas
 - Datasone template - 453 Fife Datazones, with codes running from 'S01002572' to 'S01003024'.
 - Comparison template – 2 areas – Fife and Scotland with codes CG001 and CG002 respectively.
 - Other sub-Fife geography templates (check with KFD Co-Ordinator if unsure)
- Check indicator names to ensure correct figures will be copied to KFD template.
- Check all figures remain complete if altering content of the source dataset in any way.

2. KFD Data template (datazone, comparison or other sub-Fife geography)

- Data worksheet**
 - Check indicator names/IDs are correct.
 - Check date and time period is correct for each indicator being loaded.
 - Check data has copied and pasted from source correctly.
 - Check correct number of rows of data matches the number of rows of geographies
 - 453 rows for datazone template
 - 2 rows (Fife and Scotland) for comparison template
 - For rates/percentages, ensure numerator and denominators are in correct columns.
 - Ensure rates/percentages are calculated/formulated correctly.
 - Format and check cells with count, numerator or denominators are **not** using '1000 separator' (, comma).
 - Perform logic checks for row or column totals – is output as expected?
 - Where possible, check key figures/totals against other publications.

- Metadata worksheet**

Ensure all fields are completed.

- Check Element/Indicator name matches each corresponding indicator in data worksheet.
- Check data type is correct (ratio, count...)

- Ensure indicator description is clear and comprehensive, including information on definitions and any exceptions or caveats associated with the data. Be consistent with labelling – e.g. ‘%’ or ‘percent’, ‘0-4’ or ‘0 to 4’.
- Use appropriate keywords to aid the search function in KFD.
- Ensure source of data is correct – provide links if appropriate.
- State what geography the data is available from – datazone upwards or specific geographies only.
- Ensure any disclosure controls applied, either from source or during your population of this template, are clearly explained. (e.g suppression, rounding, averages).
- Apply appropriate and consistent ‘archive due’ dates.
- Ensure ‘data availability’ is the date the data is provided/uploaded to the KnowFife Dataset.
- Ensure ‘time period’ matches datasheet and source data.
- Provide best date for when ‘next version due’ or expected.
- Apply appropriate suppression limit in accordance with disclosure control highlighted above.

3. Data Upload final checks

- Scan over datasheet and metadata to ensure no formatting issues have appeared during creation of dataset and any changes made.
- Ensure you save a copy of the completed data template for this period/data of the data.
- Use the ‘add-ins’ function to create the export file for upload to the KnowFife Dataset – save this with the same name as the excel file.

4. Live data checks

- After the KnowFife Dataset Co-Ordinator has made your data live on the KnowFife Dataset website and informed you, carry out some quick checks of the data outputs.
 - Profiles and Pick and Mix– ensure the data looks correct where it appears in relevant profiles at a couple of geographies and is available through the Pick and Mix section.
 - Check the latest figures with previous period for any apparent errors or unexpected changes to the values.

KnowFife Dataset Co-Ordinator checks

For data added to the system from the data admin group network, there are a series of checks that should be carried out to ensure that data is accurate and has complied with the data quality and disclosure control procedures.

1. Staging Area

- Systematically check the data at datazone and comparison level for any obvious errors or missing values.
- Check the time period is correct for the latest version of the data being loaded.
- Check the metadata fields have all been completed

2. Live data

- Profiles and Pick and Mix– ensure the data looks correct where it appears in relevant profiles .
- Check the latest figures are pulling through in the Pick and Mix section.
- Check the above for a couple of different geographies to ensure aggregations have been applied correctly.
- Check the latest figures with previous period for any apparent errors or unexpected changes to the values.