

Issue #17: Proposal to change text descriptions of Phase 5 Process and 5.1 Integrate data

Proposal to change text descriptions of phase 5 Process

1. We suggest to delete the statement in brackets of the last sentence from paragraph 45:

The sub-processes in this phase can apply to data from both statistical and non-statistical sources (with the possible exception of sub-process 5.6 (Calculate weights), which is usually specific to survey data).

Weighting may still be required for outputs that use non-survey data or mixed data sources so we suggest the last sentence to state:

The sub-processes in this phase can apply to data from both statistical and non-statistical sources.

1. We suggest to modify the first sentence from paragraph 47:

This phase is broken down into eight sub-processes, which may be sequential, from left to right, but can also occur in parallel, and can be iterative.

This does not hold true for statistical outputs which use admin data. E&I/code/derive new variables & units may need to happen first before integration; therefore, we suggest to modify the sentence to read:

This phase is broken down into eight sub-processes, which may be sequential but can also occur in parallel, and can be iterative.

Proposal to change text description for 5.1 Integrate data

We suggest to rewrite paragraph 48 to better reflect admin data use within the GSBPM as well as include more recent examples of integrated data.

1. This sub-process integrates data from one or more sources. It is where the results of sub-processes in the "Collect" phase are combined. The input data can be from a mixture of external or internal data sources, and a variety of collection modes, including extracts of administrative data. Administrative data can substitute for all or some of the directly collected survey variables. This sub-process also includes harmonising or creating new figures that agree between sources of data. The result of this sub-process is a set of linked data. Data integration can include:
 - combining data from multiple sources, as part of the creation of integrated statistics such as national accounts
 - data pooling, with the aim of increasing the effective number of observations of a phenomena
 - matching / record linkage routines, with the aim of linking micro or macro data from different sources
 - data fusion - integration followed by reduction or replacement
 - prioritizing, when two or more sources contain data for the same variable, with potentially different values.