4.3 Run collection

This sub-process is where the collection is implemented. The different collection instruments are used to collect or gather the information which may include raw microdata or aggregates produced at the source, as well as any associated metadata. It can include the initial contact with providers and any subsequent follow-up or reminder actions. It may include manual data entry at the point of contact, or feedback management, depending on the source and collection mode. It records when and how providers were contacted, and whether they have responded. Depending on the geographical frame and the technology used, geocoding may need to be done at the same time as collection of the data by using inputs from GIS systems, putting a mark on a map, etc. This sub-process also includes the management of the providers involved in the current collection, ensuring that the relationship between the statistical organisation and data providers remains positive, and recording and responding to comments, queries and complaints. Proper communication with reporting units and maximization of the number of non-respondents contributes significantly to a higher quality of the collected data.

For administrative, geographical or other non-statistical data the provider is either contacted to send the information or sends it as scheduled. This process may be time consuming and might require follow-ups to ensure that data are provided according to the agreements. In the cases where the data are published under an Open Data licence and exist in machine-readable form, they may be freely accessed and used.

This sub-process may also include the monitoring of data collection and making any necessary changes to improve data quality. This includes generating reports, visualising and adjusting the acquisition process to ensure the data are fit for use. When the collection meets its targets, it is closed and a report on the collection is produced. Some basic checks of the structure and integrity of the information received may take place within this sub-process (e.g. checking that files are in the right format and contain the expected fields).

Quality Indicators:
- Are there enough staff responsible for dealing with the respondent's questions?
- Support is provided to respondents (e.g. hot lines numbers).
- Quality control is used to manage the quality of data collection and data capture processes.
- Meaningful feedback is provided to interviewers and fieldworkers on a regular basis.
- Monitoring of feedback operations is done during data collection.
- Interviewer performance is measured for CATI, CAPI, PAPI surveys (e.g. interviewers' productivity).
- Domain response rates; representativeness indicators; achieved CVs of key variables in domains of interest.
- Unit nonresponse rate; item nonresponse rate; proxy rate.
- Mode effect when more than one collection mode.
- Outlying error rates; estimate of non-sampling error.
- Deviation between expected and actual start and close of collection.
- Extent to which follow up is based on administrative data.
- Percentage of data transmitted according to the agreements with administrative data owners (e.g. formal, time schedule).

For details, see Quality indicators for GSHRM (version 2.0).