4.4 Finalise collection

This sub-process includes loading the collected data and metadata into a suitable electronic environment for further processing. It may include manual or automatic data capture, for example, using clerical staff or optical character recognition tools to extract information from paper questionnaires, or converting the formats of files or encoding the variables received from other organisations. It may also include analysis of the metadata and paradata associated with collection to ensure the collection activities have met requirements. In cases where there is a physical collection instrument, such as a paper questionnaire which is not needed for further processing, this sub-process manages the archiving of that material. When the collection instrument uses software such as an API or an app, this sub-process also includes the versioning and archiving of these.

Quality Indicators:
- Discrepancy between planned versus actual collection costs
- Percentage of collection activities that met requirements assessed through analysis of metadata
- Outgoing error rates, estimates of non-sampling error
- The rate of over-coverage: the proportion of units accessible via the frame that do not belong to the target population (are out-of-scope). The rate of over-coverage is applicable to all statistical processes (including use of administrative sources).
- Error producers:
  - Unit nonresponse rate, item nonresponse rate, proxy rate
  - Percentage of materials adequately archived (easily retrievable, properly labelled, retention period indicated) (survey and ADE)
- Technical checks on ADE (e.g. 1) controls on the readability of the file (e.g. unknown format, a corrupted file, a file with an unfamiliar character set, or a file that cannot be decoded); 2) file declaration compliance (examples of problems in this area are a file with a missing metadata description and a file with a layout that does not comply to the layout agreed upon); 3) Convertibility, it focuses on the conversion of the file to the NEO-standard format (examples of problems in this area are file errors while decoding and corrupted data in the file after conversion)

For details, see Quality Indicators for GSBPM (version 2.0)