

Issue 17: Add some data integrity validation checking in the Collect phase

In the GSBPM it isn't clear whether some of the validation of administrative datasets took place in the Collect phase (in 4.3), or whether it is all in the Process phase (in 5.3). ABS staff would like to put these editing related activities in the Collect phase, as they need to do some integrity checks and fix any issues, both before and after loading these datasets to shared stores. These checks only deal with the basic integrity of the acquired datasets to the point where they can be loaded to stores. For administrative datasets, these checks might include comparing the size, content (e.g. variables) and format of files received with the expected targets and checking against the standard metadata required for successful loading (e.g. no negative values, numeric only fields, etc.). These administrative data checks are not that different to in-field checks for survey data. Any further editing/validation of input data would still be done in the Process phase.