

YOUNG LIVES SUMMARY OF KEY DATA MANAGEMENT ISSUES

The table below summarises some key data management issues to consider.

	Issues to consider	<i>Young Lives</i> approach
Software	Use software that can deal with data stored at different levels (e.g. household and community level) and that can link these files easily. This avoids unnecessary duplication of data and preserves data integrity.	MS Access was used for data management purposes (one country used Delphi). From the Access database, data can be exported into statistical packages for analysis.
Computer capacity	Analyse likely computer capacity requirements carefully. Databases often take up more space than anticipated, and it is important to allow enough capacity for storage of data.	Databases in <i>Young Lives</i> countries ranged from 15 to 70 Mb. Most PCs now have a minimum hard disk size of 20 Gb. This should be adequate. However, in one <i>Young Lives</i> country, digital photos, used for tracking, took up 20 Gb of disk space.
Back-up routine	Back up to prevent loss of some or all of your data files.	Daily intermediate back-ups (of files that had changed since the last full back-up) and weekly full back-ups were done of all files, including questionnaires, data files, programme files and results files. The files were saved onto CDs and zip disks; <i>WinZip</i> was used to compress files. <i>Young Lives</i> recommended that at least 2 copies were saved, dated and stored in separate safe locations.
Audit trail	Document everything you do and when you do it (e.g. data entry, validation, corrections) and include notes about any queries. Recording changes on paper to the master dataset provides an audit trail. You can also have an electronic logging system that records all changes and is only accessible to one person. Include notes on data extracted for analysis, programmes that were run on these data and the results produced. This makes it easier to go back and	The audit trail was a <i>Word</i> document (Delphi does more auto-recording). <i>Young Lives</i> recommended that each entry contain the date, name of the person making the entry, a brief description of what was done, and names of any files created or altered.

	repeat the process or for others to find files or information.	
Database merging	With several data entry clerks working on the data, and several copies of the database, you need a method for merging the data at the end of the data entry phase.	Access has a system called <i>Replication</i> , which allows you to have one 'Master' copy of the database, with other copies referred to as 'Replicas'. In theory, data can then be entered into each of the replica databases and these records can be merged with the master database through a process of synchronisation, one by one. In practice, this created more problems than it solved, so <i>Young Lives</i> devised a series of Append Queries in the database, which appended data from each table in turn into an otherwise blank database.
Double data entry	Use double data entry (DDE). This is a good method of data checking, as it is unlikely that two people will make the same mistake.	The idea with DDE is to have two people entering the same data into separate copies of the database. The data from the two databases are then compared and any differences checked back to the questionnaires. The Statistical Services Centre together with CARE Malawi have developed a DDE program for Access databases. This is still in the early stages of testing but was used successfully in some of the <i>Young Lives</i> countries.
Consistency checks	Checks that can be made on the data include checking that no mother has had more surviving children than live births or that values stay within a realistic range.	The Access database included checks in the data entry forms. It also has a 'skip and fill' option, so that answers to sub-questions can be coded as NA (not applicable) depending on the answer to the filter question. The database also includes a report that checks through the data and produces a list of the most obvious inconsistencies. These lists are checked against the questionnaire forms to identify whether the error is at the data collection or the data entry stage. Further checks were done on the data after it had been transferred to Statistical Package for the Social Sciences (SPSS) for analysis. Corrections were then made to the data in the Access databases and the data exported again to SPSS.
Management of	Take care of the	<i>Young Lives</i> recommended that a set

paper questionnaires	questionnaires, making sure they are protected from the elements and kept confidential.	of shelves or box-files be allocated for the questionnaires, stored according to sentinel site and child identification (ID) for efficient access. The questionnaires from future rounds of data collection can be added according to child ID.
Data handling	Sign questionnaires at each stage of the collection-checking-entry process. This is a useful way to keep track of which data should be on the database and which is still to be entered, and of consolidating the number of completed questionnaires with the number of records in the database.	Near the beginning of each questionnaire there is a section for data handlers, with a space for a supervisor to sign and date the questionnaire after checking that it is complete, and for the data entry clerk to sign (or stamp) the questionnaire after the data have been entered. The fieldworkers should also sign and date the questionnaires immediately after the interview. Names of the fieldworkers, supervisors and data entry clerks should be entered into the database so it is clear who to check back with if there are problems. The stored questionnaires should have three signatures and dates.
Codes	Use identification codes to represent a person, place or item to avoid entering lengthy names into the database.	Codes were created for child ID, form numbers, data handler identification and community identification. The child ID is the most important as it identifies the child in each of the tables, thereby linking the data at the child and household level.
Storage and archiving	Develop a system for archiving. Keep raw data, not just summary tables.	Raw data, and records of what the data represent, including codes used and how they were collected, were kept. Young Lives also kept analysis files (SPSS data, syntax and output files).