



August 2013

Technical Brief

How to approach a dataset

Part 1: Database design

Contents

Contents	2
1 Introduction.....	3
a) Why do we need a database?	3
b) Excel as a simple database solution	3
2 Analysis plan	4
a) What should an analysis plan contain?	4
b) Transforming data collection units into reporting units	5
3 Data collection tool	6
4 Designing your data model.....	8
a) Database documentation	8
b) Define your tables	9
c) Define your rows	10
d) Define your columns.....	10
e) Column design steps.....	10
f) Define your data values.....	15
5 Prepare your database for data entry	17
a) Create drop down menus.....	17
b) Setting up named ranges.....	18
c) Creating cascading drop-down menus (advanced)	20
d) Creating a look-up for P-codes	21
6 Testing your database	22
7 Data cleaning and consolidation	22
a) Quality control of data entry	23
b) Validation of rules during data entry	23
c) Consolidating data from multiple sources	24
d) Cleaning of consolidated data	24
e) Categorization of open response questions.....	24
8 Documenting changes	24
9 Additional Resources.....	25

Acknowledgements

This document would not have been possible without the leadership, support and guidance of many people. ACAPS especially wishes to express its gratitude to Emese Csete, Nigel Woof, Aldo Benini and Benoit Munch for their thorough revision and critical insight.

1 Introduction

In an ideal world, a rapid onset disaster would be the instigating event for an equally rapid deployment of a skilled assessment team, with sufficient resources at their disposal to quickly conduct a rapid multi-sector needs assessment capable of guiding decision making. This team might be comprised of an assessment coordinator, an information manager/analyst, sectoral specialists and an IT specialist.

In the potential absence of an information manager/data analyst, this technical note provides guidance in how to set up a simple database suitable for storing small amounts of data as may be generated by a rapid assessment with relatively small sample sizes. It is aimed at supporting non-specialists in information management with a working knowledge of spreadsheet applications to set up a suitable structure rapidly which will support analysis. The document uses an example questionnaire and database used for the Joint Rapid Assessment for Northern Syria (JRANS).

Database design should be undertaken at the same time as designing your data collection tools, methodology and analysis approach – therefore it is recommended that this note is first quickly read in full at the start of the assessment design process, before undertaking any of the steps within.

Analysis planning, data collection tool design and data cleaning are covered in less detail; This note needs to be read in conjunction with the technical note How to Approach a Dataset Part 2: Data Preparation, and How to Approach a Dataset Part 3: Analysis, available on the ACAPS website under <http://www.acaps.org/resources/advanced-material>.

a) Why do we need a database?

A data base is a 'tool that stores data, and lets you create, read, update, and delete the data in some manner'. It does not matter whether you're using paper or a computer software program to collect and store the data – if you have an organised collection of data collected for a specific purpose, then you have a database. Databases offer certain advantages in terms of efficiency:

- Provides a centralised digital **storage** facility for data –easy to share
- **Retrieval** and updating of specific information is made faster and easier, with the possibility of using a number of different search criteria
- Easy **updating** of data
- Facilitates **analysis** by structuring data in such a way that it is simple to conduct calculations

b) Excel as a simple database solution

Choosing the right approach, the right software and deciding how to model your data within your database depends on the processes which you will want to carry out upon your data. Within the context of a rapid assessment where technical database skills are limited, the following requirements are key:

- Fast to set up
- Does not require specialist technical database skills or software licences
- Easy to enter data
- Structures data so as to facilitate analysis

For these reasons, the recommended approach outlined here is to use Microsoft Excel. Whilst this is a spreadsheet application as opposed to database management software, Excel is very good for entering, storing and analysing small amounts of data; it has a much lower learning curve, and also provides built in analysis features which would require significant programming in a database application.

2 Analysis plan

a) What should an analysis plan contain?

It is a common mistake for data collection tool design and sampling design to be seen as a very separate issue from database design, data cleaning/coding and analysis outputs.

These steps are very closely linked; the information output which you get from an assessment will depend on structuring the data in such a way that it can be analysed, and this in turn is constrained by **how** the data has been collected (e.g. one community group discussion per site, or several household interviews), and **what** has been collected (e.g. what question asked). If you design a tool without thought to the other stages, you may come unstuck somewhere along the way by having collected data which does not end up providing the information which you need.

Develop an **analysis plan** at the start of an assessment. This will help you to think through the links between your **information** needs, the **question/data** request which will collect this data (e.g. 'indicate your top three priority sectors from the following list), and your **sampling strategy**. It will contain details about how the output of each question will be **analysed** to provide the desired information. This will ensure that you know in advance how you will transform data into information.

It is important to make the distinction between **data and information**. Data are facts about the world, whereas information is the result of processing raw data to reveal its meaning. Processing can mean carrying out complex calculations, but it can also be as simple as organising data to reveal a pattern, or extracting a key data item (e.g. who was the oldest participant?). In order to reveal meaning, information also requires *context*, i.e. comparison with baseline data. *Good decisions require good information that is derived from raw facts.*¹

Your analysis plan should contain:

- Data collection units. *E.g., community, household*
- Reporting unit. *This tends to be location; either a community, or an administrative unit.*
- Transformation from data collection unit to reporting unit, if required
- Transformations from data inputs into information outputs

The types of data which are collected/reported and the way in which it is collected will affect the design of the structure - and also the ease of analysis.

Quantitative data is data which can be measured and analysed numerically, allowing it to be presented as statistics, tables and graphs (e.g. Number of affected households).

Qualitative data is descriptive, and can be observed but not measured in an exact way (e.g. types of humanitarian needs – shelter, health).

Qualitative data can still be analysed in a quantitative way. For instance, a need for shelter is qualitative data; however, a need for shelter in 50% of sites visited is a quantitative analysis from qualitative data. The benefit of quantitative analysis of qualitative data is that it provides a succinct summary, it is easy to understand/interpret and it allows for simple comparison

¹ Database Systems Design, Implementation and Management (9th Edition), 2010

b) Transforming data collection units into reporting units

In some cases, it may be possible that you do not have a *one-to one relationship* between units of reporting and units of data collection. For instance, you may have the reporting unit of an administrative area, e.g. district, but may have sampled at the household level.

Alternatively, your methodology may involve doing several key informant interviews, community group discussions and direct observation in each community. These are both one to many relationships between the reporting unit (community) and the data collection unit.

Your analysis plan should indicate how you will transform/aggregate/consolidate units of data collection into units of reporting. For instance, in the previous example of a district reporting unit but a household sampling, you may decide that you will take the majority view; in the case of quantitative information, you could average it – so long as it is credible to treat each observation as equivalent. For qualitative data, you could take the most popular response in the case of a single option question. For multiple option questions, this becomes much more complex. You will need to give good thought to whether the outputs of these calculations will be logical, and whether they will allow you to meet your information needs.

You might also face the situation where in each location, several data collection techniques are used a different number of times – for instance, several key informant interviews, community group discussions and direct observations. If these cover some of the same variables (recording the same information but from different sources, e.g. asking both key informants and community group discussions for their priority sector for response), it will be less credible to treat these with equal weight. The outputs of a community group discussion already represent a larger consensus than a key informant, therefore are unlikely to be weighed equally. Also, direct observation is often used as a technique for the assessment team to verify visually what they have been told in key informants or community group discussions. How will this information be cross checked against other information?

In these more complicated scenarios, it becomes difficult to define hard and fast rules for aggregating to the reporting unit. As this requires judgement rather than calculation, this is a task which is best carried out by the assessment team, immediately after the data is collected. Having interacted with respondents, the assessment team are the ones best placed to determine what the 'correct' response is. When sufficient trust is placed in the enumerators and when time constraints are important, final conclusions of the assessment teams after the field visit should be recorded in one single form, the one that will be used in the database.

Finally, consider whether there is validity to maintaining some views separately. For instance, if interviewing individuals, rather than averaging responses at the community level, you may want to average the female and the male view separately and maintain both, so as to be able to disaggregate your analysis by gender. If conducting male and female CGDs and male and female KIs, then you may want to consolidate this information to a female and a male viewpoint for each community, which could be conducted by the assessment team. Figure 1 shows an extract from the Analysis plan from the Pakistan MIRA, demonstrating the linkages between information needs, indicators, required data, data sources and analysis type.

3 Data collection tool

This document does not cover questionnaire design in detail. If a thorough job is done on the analysis plan, the data collection tools should be “fairly” simple to design.

Often, data collection tools are designed before an analysis plan has been written and sampling methodology has been determined. Whilst this is not ideal, the retrospective development of an analysis plan may still help to highlight any pitfalls and limitations to the tool which could limit analysis. Some recommendations to support later analysis are outlined in the Table 1 below.

Table 1. Recommendations for data collection design

Quantitative	
Units	Ensure reporting units are the same – either by stipulating unit of measurement (e.g. number of affected households) or at very least by ensuring that the unit is reported.
Definitions	If the data could be open to interpretation, always provide a definition (e.g. “affected population”)
Null responses	Ensure that it will be possible to differentiate between a response of zero, as compared to a null response – e.g., if a question ask for affected population and is left blank, is it because there are no affected people, or because the respondent does not know? Wherever clarity is required to differentiate this in your responses, consider adding a ‘do not know’ or ‘not applicable’ option, as appropriate to the question.
Qualitative:	
Create categories in advance	Qualitative data can be analysed in a quantitative way if responses are categorised. Categories should be defined in advance, and applied at the point of data collection; this can be done either by presenting a closed question of set options, or by providing an open question, the answer to which is then categorised or recoded by the data collector at the point of information recording.
Don't assume you know it all	The downside of pre-defined categories is that they assume that you already know all the possible options – in order to allow for unanticipated responses, include an ‘other’ category, always ensuring that details of what the ‘other’ response is recorded.
Single response vs. multiple response	<p>Single choice options are suitable in cases where options are mutually exclusive, and will allow simple and intuitive statistics to be carried out, e.g. ‘40% of respondents were displaced, 30% were staying with family and 30% were still living in their homes’.</p> <p>In certain cases, responses may only be fully described through a combination of options (multiple choice), e.g. there is a problem with access to water because water sources are contaminated and security prevents access to water points. The analytical outputs of multiple-response questions can be more complex, but can be a more accurate representation or reality and variability in the context.</p> <p>If it is necessary to categorise responses into exclusive categories, then consider also asking respondents to select their main/priority/dominant category (remember that they are the only ones who can do this!). You can do this by designing the question as a ranking, where the respondent indicates the relative priority of responses.</p>

Figure 1. Extract from Aleppo J-RANS Analysis plan – WASH Section

Information needs	Needs Indicator	Indicator	Data Source	Analysis type	Question	Sample visualization																														
Main issues in water supply	Main problems in water supply as expressed by the population	Frequency of problems reported due to access issues Frequency of problems reported due to availability issues	Local population, relief committees, Head of HH, Water committee, local organization, NGOs	Breakdown per areas	Is there a serious problem regarding water in this neighbourhood? If yes, I am reading a list of possible problems (<i>Select max five most serious problems</i>)	<p>Main Water problems in Aleppo</p> <table border="1"> <tr><td>Not enough water available because water systems, well or pump is broken</td><td>54%</td></tr> <tr><td>Not enough water available because people don't have means to store water</td><td>37%</td></tr> <tr><td>Lack of ways to treat water or heat for boiling it</td><td>14%</td></tr> <tr><td>Water does not taste good or does not look good enough</td><td>8%</td></tr> <tr><td>Not enough water available because water is too far away or difficult to access</td><td>6%</td></tr> <tr><td>Not enough water available because water too expensive</td><td>3%</td></tr> </table>	Not enough water available because water systems, well or pump is broken	54%	Not enough water available because people don't have means to store water	37%	Lack of ways to treat water or heat for boiling it	14%	Water does not taste good or does not look good enough	8%	Not enough water available because water is too far away or difficult to access	6%	Not enough water available because water too expensive	3%																		
Not enough water available because water systems, well or pump is broken	54%																																			
Not enough water available because people don't have means to store water	37%																																			
Lack of ways to treat water or heat for boiling it	14%																																			
Water does not taste good or does not look good enough	8%																																			
Not enough water available because water is too far away or difficult to access	6%																																			
Not enough water available because water too expensive	3%																																			
Main issues in Sanitation	Main problems in Sanitation as expressed by the population	Frequency of problems reported due to access issues Frequency of problems reported due to availability issues	Local population, relief committees, Head of HH, Water committee, local organization, NGOs	Breakdown per areas	Is there a serious problem regarding sanitation and hygiene in this neighbourhood? If yes, I am reading a list of possible problems (<i>Select max five most serious problems</i>)	<p>Main Sanitation & Hygiene problems in Aleppo</p> <table border="1"> <tr><td>Not enough access to water or soap because of high prices</td><td>45%</td></tr> <tr><td>No regular rubbish collection so general waste build up</td><td>26%</td></tr> <tr><td>Not enough access to water, soap or places to wash due to security conditions</td><td>7%</td></tr> <tr><td>Not enough places to wash your body or bath</td><td>3%</td></tr> <tr><td>Not enough access to toilets because they are too far away</td><td>2%</td></tr> <tr><td>Not enough access to separate toilets for men and women</td><td>2%</td></tr> </table>	Not enough access to water or soap because of high prices	45%	No regular rubbish collection so general waste build up	26%	Not enough access to water, soap or places to wash due to security conditions	7%	Not enough places to wash your body or bath	3%	Not enough access to toilets because they are too far away	2%	Not enough access to separate toilets for men and women	2%																		
Not enough access to water or soap because of high prices	45%																																			
No regular rubbish collection so general waste build up	26%																																			
Not enough access to water, soap or places to wash due to security conditions	7%																																			
Not enough places to wash your body or bath	3%																																			
Not enough access to toilets because they are too far away	2%																																			
Not enough access to separate toilets for men and women	2%																																			
Affected groups	Ranking of groups the most at risk as reported by the population	Groups the most vulnerable in the WASH sector	Local population, relief committees, Head of HH, Water committee, local organization, NGOs	Breakdown per areas and priority rank	Regarding the lack of safe water, which group is most at risk? (<i>rank top three: 1=first rank, 2=second rank, 3=third rank</i>)	<p>WASH Risk to Vulnerable Groups</p> <table border="1"> <tr><th>Group</th><th>Very High</th><th>High</th><th>Intermediate</th><th>Low</th></tr> <tr><td>Resident population who have not been displaced</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> <tr><td>Displaced people in host families</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> <tr><td>Displaced people in vacated buildings</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> <tr><td>Resident population hosting displaced persons</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> <tr><td>Displaced people in collective shelter (hotels, camps, etc.)</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> </table>	Group	Very High	High	Intermediate	Low	Resident population who have not been displaced	0%	0%	0%	100%	Displaced people in host families	0%	0%	0%	100%	Displaced people in vacated buildings	0%	0%	0%	100%	Resident population hosting displaced persons	0%	0%	0%	100%	Displaced people in collective shelter (hotels, camps, etc.)	0%	0%	0%	100%
Group	Very High	High	Intermediate	Low																																
Resident population who have not been displaced	0%	0%	0%	100%																																
Displaced people in host families	0%	0%	0%	100%																																
Displaced people in vacated buildings	0%	0%	0%	100%																																
Resident population hosting displaced persons	0%	0%	0%	100%																																
Displaced people in collective shelter (hotels, camps, etc.)	0%	0%	0%	100%																																
Response capacity	Type and regularity of assistance provided in the WASH sector	Frequency of intervention reported in the WASH sector Type of assistance provided	Local population, relief committees, Head of HH, Water committee, local organization, NGOs	Breakdown per areas Breakdown by humanitarian actor	Which organizations have been providing regular water, sanitation or hygiene support in this neighbourhood over the past 30 days? <ul style="list-style-type: none"> Type of organization Organisation responsible Regular or one off support 	<p>Percentage of neighbourhoods receiving WASH support</p> <table border="1"> <tr><th>Area</th><th>Support Received</th><th>No Support Reported</th></tr> <tr><td>All Areas</td><td>92%</td><td>8%</td></tr> <tr><td>South West</td><td>93%</td><td>7%</td></tr> <tr><td>South East</td><td>92%</td><td>8%</td></tr> <tr><td>North East</td><td>92%</td><td>8%</td></tr> </table> <p>Type of agencies providing WASH assistance</p> <table border="1"> <tr><th>Area</th><th>NGOs</th><th>Local relief provider</th></tr> <tr><td>North East</td><td>1</td><td>0</td></tr> <tr><td>South East</td><td>2</td><td>0</td></tr> <tr><td>South West</td><td>1</td><td>0</td></tr> </table>	Area	Support Received	No Support Reported	All Areas	92%	8%	South West	93%	7%	South East	92%	8%	North East	92%	8%	Area	NGOs	Local relief provider	North East	1	0	South East	2	0	South West	1	0			
Area	Support Received	No Support Reported																																		
All Areas	92%	8%																																		
South West	93%	7%																																		
South East	92%	8%																																		
North East	92%	8%																																		
Area	NGOs	Local relief provider																																		
North East	1	0																																		
South East	2	0																																		
South West	1	0																																		
Severity of conditions	Severity of problems	Severity status on a life-saving scale	Local population, relief committees, Head of HH, Water committee, local organization, NGOs	Breakdown per areas	Overall, which of the following statements describes best the general status of water supply? (<i>Circle one right answer</i>)	<p>Water supply severity status in Aleppo per area</p> <table border="1"> <tr><th>Area</th><th>Very High</th><th>High</th><th>Intermediate</th><th>Low</th></tr> <tr><td>North East</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> <tr><td>South East</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> <tr><td>South West</td><td>0%</td><td>0%</td><td>0%</td><td>100%</td></tr> </table>	Area	Very High	High	Intermediate	Low	North East	0%	0%	0%	100%	South East	0%	0%	0%	100%	South West	0%	0%	0%	100%										
Area	Very High	High	Intermediate	Low																																
North East	0%	0%	0%	100%																																
South East	0%	0%	0%	100%																																
South West	0%	0%	0%	100%																																

4 Designing your data model

a) Database documentation

A lack of documentation of records and variables is a commonly identified issue when a new dataset is received by an analyst who did not design it². In order to ensure that the database is never separated from this information, create supporting documentation in worksheets in the same workbook as your database. Table 2 summarises the various spreadsheets which will eventually be needed, some of which will be covered in later sections of the document.

Table 2. Supporting documents

Database	Where all data will be stored. If your data model contains several tables, there will be one worksheet per table.
Data dictionary/ codebook work	Worksheet containing all variable names, types, data formats, categorical values. Demonstrates how each database field relates back to the data collection tools.
Domains	Contains all of the lists of categories used within the database.
P-code lookup	Contains all of the administrative boundary information.
Change log	For keeping a record of all modifications made to data within the database

A **data dictionary** contains the ‘metadata’ for your data model, essentially all of the information necessary for someone else to understand your database. Data dictionaries document what each of your variables are, their names, what data type they are, and codes for categorical values. A **codebook** is a similar document, but provides a technical description of a **data file**, describing how the data are arranged in the files, what the various numbers and letters mean, and any special instructions on how to use the data properly.

For the purpose of a rapid assessment, you can combine both codebook and data dictionary functionality in one document which describes not just your database, but relates the database design back to your data collection instruments. An example codebook can be found within the example database attached to this document. In your code book, you should include the following:

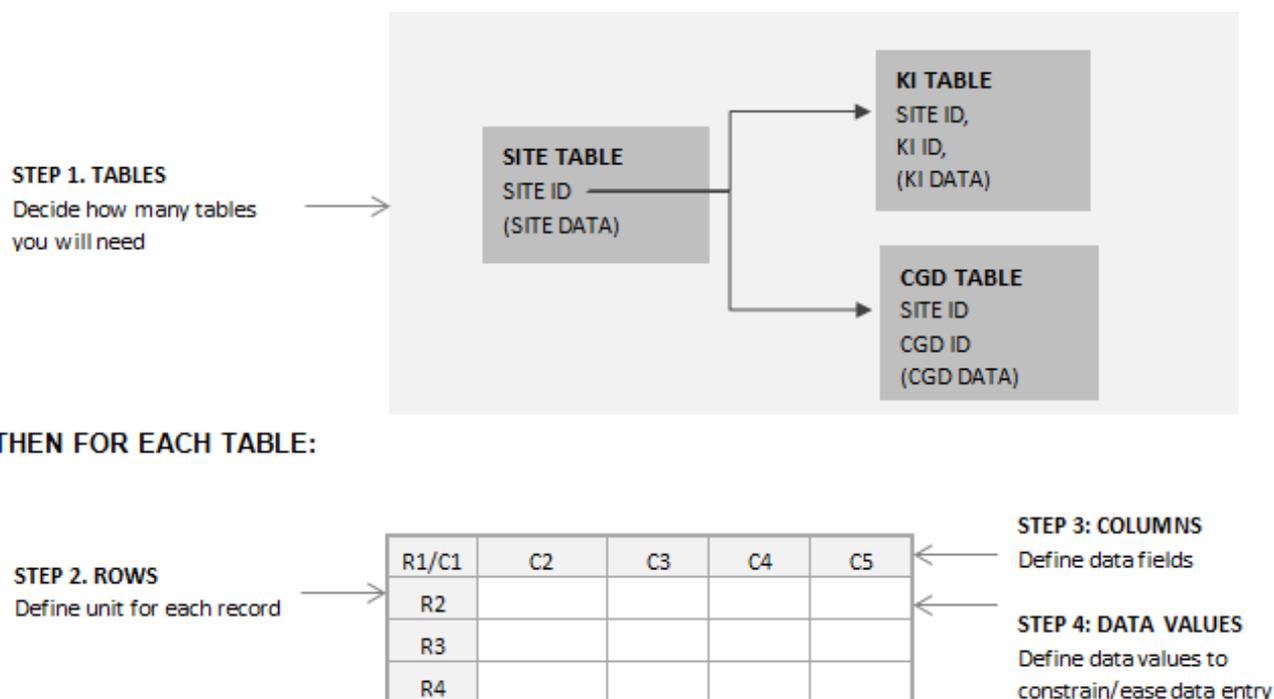
- Name of each column
- Variable type (relating the data back to the relevant data collection item, or if the variable is calculated, indicating how it was calculated)
- Data format (number, category, text, date)
- Categorical values

² Technical brief: How to approach a dataset. Part 1: Data preparation. Aldo Benini, 15/03/2013

b) Define your tables

Figure 2 shows the four steps in designing the structure of your database. The first step is to decide how many tables you need. This will relate strongly back to your analysis plan. As the aim is simplicity, the ideal would be one table, where one rows represent one unit of reporting, and each column holds the different variables. This will make your life much easier when it comes to analysis.

Figure 2. The four steps in designing your data model



More than one table means you have a relational database; whilst Excel can be used to model either non-relational or relational database, it lacks the functionality of full database applications (Access, SQL and Oracle) which helps manage these relationships – therefore keeping the number of tables to a minimum will avoid the manual management of complex relationships. Key principles are as follow:

- Data should be organised to support analysis, not to reflect the data collection tools.
- Minimise the extent to which you have to analyse across tables.
- Wherever you have the same data structure, store information together (e.g. if you have a questionnaire which can be applied to both male and female KIs, store these in the same table).

Databases are information containers, which hold digital data in a way that allows a user to interact with it.

The easiest way to store small quantities of data is as a **non-relational database** - which is a two dimensional array of data - rather than as a **relational database**, which models data in terms of the relationships between different sections of the data (resulting in several tables of information interlinked through the use of unique keys, allowing one to many relationships between different data sections).

Non-relational databases are ideally suited to data with the same number of data fields for each record. They have some drawbacks compared to relational database models (such as more redundancy), but for small amounts of static data such as from a rapid assessment, the advantages in terms of ease of analysis outweigh these constraints:

- **User-friendly:** This is a very simple model which most people can easily visualise, and which can easily be modelled using a spreadsheet application.
- **Ease of analysis:** This approach structures data immediately into a format which will easily allow analysis.

c) Define your rows

One row should represent one record/one sample, with each record being the basic unit of *reporting* for your assessment. This should be evident from the information needs outlined within your analysis plan, and will probably correspond with one set of responses at the data collection stage. For instance, if your information need is to know the proportion of households assessed who need shelter, then your rows would represent households. Structuring one row per response unit will allow for an easy analysis, by allowing meaningful calculations at the level of each column.

d) Define your columns

Each column, or database field, should contain one variable/discrete unit of information. Each column should be defined in your code book (see Figure 3), giving it a variable name and relating the field back to the relevant section of the data collection tool. In some cases, it will be possible to structure one field to contain the response to one question, though often more than one field will be necessary. Table 3 provides examples of how certain common question types can be mapped across columns. Section e contains column design steps.

Figure 3. Questionnaire section with corresponding code book excerpt

J1. What percentage of children (6-14 yrs of age) is regularly attending school in this neighbourhood? <input type="checkbox"/> 0-25% <input type="checkbox"/> 26-50% <input type="checkbox"/> 51-75% <input type="checkbox"/> 76-100%		283	284	285	286	287	288	289	290
J2. What are the reasons why children are not attending schools? (Select all that apply) <input type="checkbox"/> Schools not functioning (damaged, destroyed or occupied) <input type="checkbox"/> Lack of school materials (stationery, books, etc.) <input type="checkbox"/> Safety - fear of schools being bombed/targeted <input type="checkbox"/> Lack of water and hygienic sanitation facilities in schools <input type="checkbox"/> Lack or absence of teachers <input type="checkbox"/> Other (specify): _____		J. Education							
		J1. % of children attending school		J2. Main reasons for not attending schools					
		What percentage of children (6-14 yrs of age) is regularly attending school in this neighbourhood?	Schools not functioning (damaged, destroyed or occupied)	Safety - fear of schools being bombed /targeted	Lack or absence of teachers	Lack of school materials (stationery, books, etc.)	Lack of water and hygienic sanitation facilities in schools	Other	Other - specify
		j1_att	j2_sc_nf	j2_sc_bom	j2_sc_teach	j2_sc_mat	j2_sc_sanit	j2_sc_oth	j2_sc_0

e) Column design steps

Unique ID in first column: In the first column, keep a unique ID which allows you to cross reference between rows in the database and original paper/digital questionnaires or forms, to allow you to trace data back to its source. Using an alpha alpha-numeric code can be useful for recording both the identity of the assessment team as well as the number of the returned questionnaire (e.g., 'EC07'). Whilst alpha-numeric codes cannot be sorted as effectively as numeric codes, an additional column containing containing numeric values can be added to the database as the record number, to facilitate sorting.

Unique column headers: Create unique column headers (necessary for managing data entry if done through a form, and also for cross-tabbing data during analysis). The column headers should be kept

short (e.g. the question number, not the whole question) and should allow you to easily reference between the column contents and the questionnaire. When many columns refer to the same question number (this is the case with multiple choice, multiple option questions), extend the code to be question number plus a code to represent the response (e.g. see Figure 3; each response to question J2 is coded with 'j2_sc_' then a code representing the answer, e.g. 'nf' for 'Schools not functioning').

Additional column headers: In your code book, add additional header rows to help relate the responses back to the questionnaire text. See figure 3.

Stratification data near start: In the columns following the unique ID, include columns for the main factors which will be used to stratify the analysis (major elements of comparison, e.g. displaced/host, urban/rural, male/female, geographical area - as outlined within your analysis plan).

Geographical information: Geographical information must be recorded accurately in order to ensure that information is attributed to the correct location, allowing mapping and comparison against other data. Assessment information is often collected and recorded by administrative unit. The best way to record the administrative unit is by using the P-code; get an up to date p-code list which you can use to design the geographical information columns. If your information is not being gathered by administrative unit but by specific location, geographical coordinates should be recorded. In the later case, separate fields should be implemented for latitude and longitude – do not put both into one field. See box on page 15 for more on recording geographical information.

Map remaining questions: Map out the remaining questions across the other columns. Stick to the same order of items as in the data collection tool to allow easier navigation through the data.

Table 3. Mapping questions to columns

Example question	Example response structure	Example database structure																
Number field where unit is specified: one column																		
B1. Estimated population in neighbourhood:	<table border="1"> <tr> <td>Estimated number of population in neighbourhood:</td> <td>2,500</td> </tr> </table>	Estimated number of population in neighbourhood:	2,500	<table border="1"> <thead> <tr> <th>Responses</th> <th>B1_popn</th> </tr> </thead> <tbody> <tr> <td>R1</td> <td>2,500</td> </tr> <tr> <td>R2</td> <td></td> </tr> <tr> <td>R3</td> <td></td> </tr> </tbody> </table> <p>Data values: <i>Number: age</i></p>	Responses	B1_popn	R1	2,500	R2		R3							
Estimated number of population in neighbourhood:	2,500																	
Responses	B1_popn																	
R1	2,500																	
R2																		
R3																		
Number field where unit is variable: two columns																		
B1. Estimated population in neighbourhood:	<table border="1"> <tr> <td>Estimated population in neighbourhood:</td> <td>2,500</td> <td>Units:</td> <td>Individuals</td> </tr> </table>	Estimated population in neighbourhood:	2,500	Units:	Individuals	<table border="1"> <thead> <tr> <th>Responses</th> <th>D7_sitepop</th> <th>D7_sitepop_u</th> </tr> </thead> <tbody> <tr> <td>R1</td> <td>2,500</td> <td>Individuals</td> </tr> <tr> <td>R2</td> <td></td> <td></td> </tr> <tr> <td>R3</td> <td></td> <td></td> </tr> </tbody> </table> <p>Data values: <i>Number</i> Data values: <i>Freetext: units</i></p>	Responses	D7_sitepop	D7_sitepop_u	R1	2,500	Individuals	R2			R3		
Estimated population in neighbourhood:	2,500	Units:	Individuals															
Responses	D7_sitepop	D7_sitepop_u																
R1	2,500	Individuals																
R2																		
R3																		
Single choice option list: one column																		
E2. Is there a serious problem because people are not able to get adequate health care for themselves?	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know	<table border="1"> <thead> <tr> <th>Responses</th> <th>E2_HC_YN</th> </tr> </thead> <tbody> <tr> <td>R1</td> <td>Yes</td> </tr> <tr> <td>R2</td> <td></td> </tr> <tr> <td>R3</td> <td></td> </tr> </tbody> </table> <p>Dropdown menu: <i>Yes, No, Do not know</i></p>	Responses	E2_HC_YN	R1	Yes	R2		R3									
Responses	E2_HC_YN																	
R1	Yes																	
R2																		
R3																		

E2. If yes, I am reading a list of possible problems. **Select the main reason.**

	Not enough health facilities available
X	Lack of medicines
	Lack of medical staff
	No access to health services due to physical/logistical constraints
	No access to health services due to security constraints
	No access to health services due to limited economic resources

Responses	E2_HC_main
R1	Lack of medicines
R2	
R3	

Dropdown menu:
Not enough health facilities available, lack of medicines, etc.

Single choice option list with 'other' category: two columns

E2. If yes, I am reading a list of possible problems. **Select the main reason, or use 'other'.**

	Not enough health facilities available
	Lack of medicines
	Lack of medical staff
	No access to health services due to physical/logistical constraints
	No access to health services due to security constraints
	No access to health services due to limited economic resources
X	Other (give details) : No birthing service

Responses	E2_HC_main	E2_HC_oth
R1	Other	No birthing service
R2		
R3		

Dropdown menu **Data values:**
Not enough health facilities available, lack of medicines, etc. *Freetext: Details of 'other'*

Multiple choice list: as many columns as options

E2. If yes, I am reading a list of possible problems. **Select all which apply**

X	Not enough health facilities available
	Lack of medicines
	Lack of medical staff
X	No access to health services due to physical/logistical constraints
	No access to health services due to security constraints
	No access to health services due to limited economic resources

Responses	E2_1	E2_2	E2_3	E2_4	E2_5	E2_6
R1	1				1	
R2						
R3						

Data values: 1 or 0 (Boolean)

Multiple choice list with ‘other’ category: as many columns as options +1

E2. If yes, I am reading a list of possible problems. **Select all which apply**

<input checked="" type="checkbox"/>	Not enough health facilities available
<input type="checkbox"/>	Lack of medicines
<input type="checkbox"/>	Lack of medical staff
<input checked="" type="checkbox"/>	No access to health services due to physical/logistical constraints
<input type="checkbox"/>	No access to health services due to security constraints
<input type="checkbox"/>	No access to health services due to limited economic resources
<input checked="" type="checkbox"/>	Other (give details) : No birthing service

Resp.	E2_1	E2_2	E2_3	E2_4	E2_5	E2_6	E2_7	E2_o
R1	1				1		1	No birthing service
R2								
R3								

Data values:
1 or 0 (Boolean)

Data values:
Freetext: Details of ‘other’

Ranking of pre-set options: as many columns as options

F5. Which group is most at risk of having not enough food to survive in this neighbourhood? **(rank top three: 1=first rank, 2=second rank, 3=third rank)**

<u>1</u>	Displaced people living in host families
----	Displaced people in collective shelter (schools, camps, etc.)
<u>3</u>	Displaced people in vacated buildings
<u>2</u>	Resident population hosting displaced persons
----	Resident population who have not been displaced

Responses	f5_fv_dh	f5_fv_dc	f5_fv_dv	f5_fv_rh	f5_hv_rn
R1	1		3	2	
R2					
R3					

Dropdown menu: 1,2, or 3. You could also make the dropdown menu show ‘first’, ‘second’ and ‘third’, so that it is easy to differentiate this ordinal data from nominal data.

Ranking of non pre-set options: as many columns as ranks

E3. Which specific health interventions are most urgently required in this neighbourhood? **(Enter short description)**

First rank :	Treatment for skin infections
Second rank :	Birthing services
Third rank :	Vaccinations

Responses	e3_hi_r1	e3_hi_r2	e3_hi_r3
R1	Treatment for skin infections	Birthing services	Vaccinations
R2			
R3			

Data values: Freetext

Geographical data / P-codes

P-codes are unique identification codes, represented by combinations of letters and/or numbers to identify a specific administrative area or location. These are commonly used when mapping data, in order to allow information to be linked to geographical boundaries and represented on a map. As well as ensuring a unique reference system, using P-codes rather than a names also avoids issues of inconsistent admin name spelling, which are not uncommon occurrences when names have been translated from their original language.

Geographical areas are often a key disaggregation factor in rapid assessments. The accurate recording of the location where information was collected is essential; however, the names of administrative units can sometimes be duplicated within a country, within different regions (for instance, in the US there are 13 cities, 11 towns, and 14 townships named Springfield). In order to make sense of administrative areas, these must be recorded in a way in which they are unique. This can be by reporting the **full administrative hierarchy** within which it is situated, e.g. Springfield, Massachusetts. This is a very commonly applied approach often seen within assessments, where location will be recorded as Admin level 1, Admin level 2, Admin level 3 (e.g.: Province, District, Commune).

Another way to uniquely identify geographical areas is by recording corresponding **P-Codes** (codes allocated to each administrative area as a means of identifying them uniquely). Recording the p-code for any location based information will ensure that the information is not attributed to the wrong location at a later stage due to name duplication, and will also allow the information to be easily imported into a GIS system, linked to boundary data and displayed on a map. It will also allow for an easy comparison between assessment outputs and data generated by other institutions (for instance, census data showing pre-disaster population figures).

Whilst the actual names of the administrative areas will be used for reporting (no-one tends to memorise the list of p-codes!), it is desirable to 'translate' this into p-coded data when entering the information into the database. To do this, ensure that you have an up to date copy of the p-coded administrative areas for the country. These can normally be found on each country's Humanitarian Response website, under the Common Operational Datasets (CODs), a registry of which can be found on the main humanitarian response website, here: <http://cod.humanitarianresponse.info/>. OCHA are a good point of contact for up to date P-code lists in country.

In cases where information is not being collected and recorded by administrative unit, it may be desirable to record the location of data collection for later use. Assessment teams will need access to and knowledge of how to use a GPS, and coordinates should be collected and collected in one pre-agreed coordinate system (e.g. WGS84 Latitude/Longitude coordinates) and reported in a common format, preferably decimal degrees. This will eliminate the need for coordinate conversion, which can be both timely and error prone if coordinates are not transcribed correctly.

f) Define your data values

In your codebook, define the data type for each column (e.g. number, text, date). If the data value is one of a list of options, define the data type as a domain, and list the potential options (domain values) to each question (see figure 4). These will be used in the next section to constrain data entries. Where necessary, allocate codes to differentiate between zero and 'no reply' (Table 4). If you choose to code

all of your responses, you should list the codes and the response which they correspond to (see box below on coding responses).

Figure 4. Example code book with response options

	1	73	74
1	Section	D. INFORMATION	
2	Sub-Section	D1.Existence of humanitarian assistance	D1.Level of information
3	Question wording	Is humanitarian assistance provided in this neighbourhood over the past 30 days?	If yes, are people generally
4	Code	d1_hum_ass	d1_info_lev
5	Data type	Domain	Domain
6	Domain	Yes	Well informed about humanitarian assistance
7	values	No	Poorly informed about humanitarian assistance
8		Do not know	Not at all informed about humanitarian assistance
9			

Table 5. Defining data values

Data type	In your code book, define the format of the data to be entered, e.g. number, text, date, domain (categories).
Domain values	Where a set of responses can be defined, create a list of all of the option. This will be used as the basis for drop-down lists, which will speed up data entry and ensure spelling consistency.
Null value codes	Where you are recording numbers, identify a code for null responses, in order to differentiate them from zero. For instance, in a field where you are recording age (a number), if a data collection form is returned with an incorrect entry of 'male' and no age, then this should be recorded in the database with a specific code. Choose a non-numeric code (e.g. 'Null') to ensure that descriptive statistics can be generated without errors.

To code or not to code?

In some situations (see example in Figure 4 for question D1), your response options may be very wordy. Good database etiquette would be to replace these options with a code (e.g. 1, 2 or 3 in the previous example); storing only one number per respondent reduces redundancy in the database. In a relational database, the 'translation' for the code is normally also stored at the database level. While you could implement this for your database, the issue of storage space is unlikely to be a consideration when dealing with small data sets from rapid assessments. Furthermore, introducing a process of translation from text to code at data entry, then from code back to text during analysis, is both time consuming and leaves you more open to errors. For this reason, the examples in this document are all encoded.

5 Prepare your database for data entry

Your code book now contains the outline of the structure of your database. Create a copy, keeping the column header rows and delete the data values – this will be your final database. If you will be conducting data entry directly into the spreadsheet, these header rows will help to ensure that the person doing the data entry can orientate themselves easily between the questionnaire and the database. If conducting data entry through a form, you may need to reduce this to one column header, in which case keep the column ID header, and remove the other header rows.

a) Create drop down menus

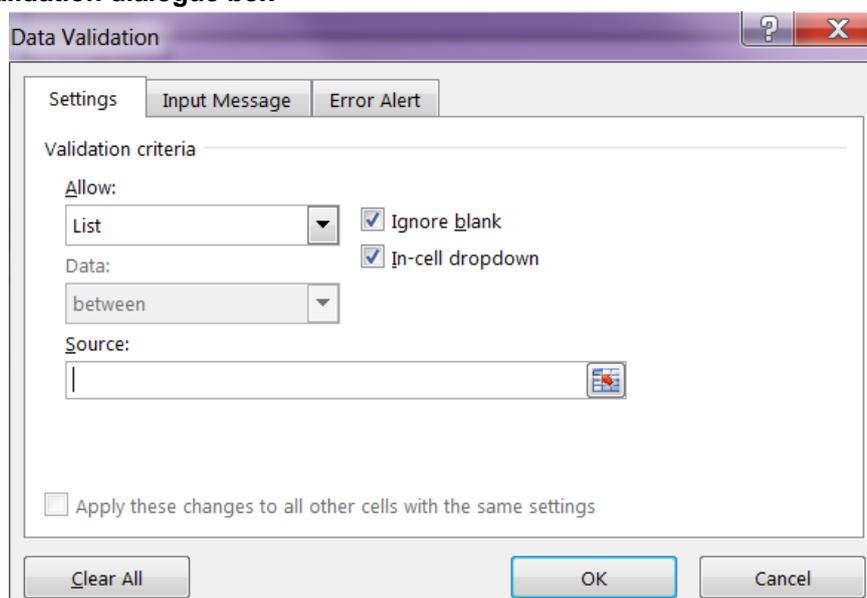
The final step is to create drop down menus to support data entry, and to validate the incoming data. If you will be entering data directly into the spreadsheet, you will do this in the cells of the database spreadsheet. If you will be setting up a form for data entry, you can do it in the form. As form design is beyond the scope of this document, this section will focus on setting up the validation within the workbook.

Drop down menus help to constrain data entry, the advantages being:

- Data entry will be faster
- Erroneous values will be prevented
- Spellings will be constrained – this is essential for ensuring that you will be able to automatically sum, filter and cross tabulate information. For instance, if a field containing 'gender' contains both the values 'female' and 'woman', it will not be possible to automatically sum up all 'female' responses unless data is cleaned. Even additional white spaces can prevent Excel from matching different values.

To create drop-down menus in Excel, use the 'data validation' function. This is set up at the level of the cell, allowing you to specify the list of acceptable options. Highlight the cell which should have the dropdown list, then go to the 'Data' menu and select data validation. In the data validation dialogue box (figure 5), select the method of validation to be a list.

Figure 5. Data validation dialogue box



For the source of the list, whilst you could reference the list of option in the codebook directly by typing in the reference of the cells containing the options (e.g. '=A1:A5', or selecting them manually), it is preferential to work with **named ranges**. These are names assigned to a range of cells, allowing the range to be referenced in formula rather than hard coding the reference of the location of the cells. To reference named ranges, type an equals sign followed directly by the name of the range (e.g. '=elec_func'). The following section details how to set up and manage named ranges.

Only set up your data validation in the **first** cell immediately below the column header. Once you have set this for all columns, you can use the **autofill** function to copy all of these validation rules to the remaining spreadsheet – this will save you copy/paste time, and will also ensure that you don't have any erroneous rows with the wrong formula.

Why work with named ranges?

The advantage of working with named ranges rather than hard coding absolute cell references relates to database upkeep. If there are any changes to the questionnaire structure (last minute changes to the current data collection format, or re-assessment at a later stage with slight modifications to the questionnaire), the database and code book will also need to be updated.

If you have lists of response options which are used several times within the database (in the example database, see 'source reliability', with options '1= reliable', '2= fairly reliable', '3=unreliable'), any change to this list (e.g. to add '0=no response'), it would be necessary to go through the entire database and change the absolute cell reference in every column which references this list.

Once you have set up cell validation, there is no way to easily see which cells are referenced within the validation without clicking on each cell and inspecting the formula for the source. If you have several columns within your database that reference the same set of cells where a list of options is given, any additions or subtractions to list of options would need to be changed in all of those cell validations.

However, if you have set up a named range which references the three options, and have set up your data validation in all of the relevant columns to reference that named range, you will only need to update the named range to include the additional option, rather than having to hunt through your worksheet to find all references to it.

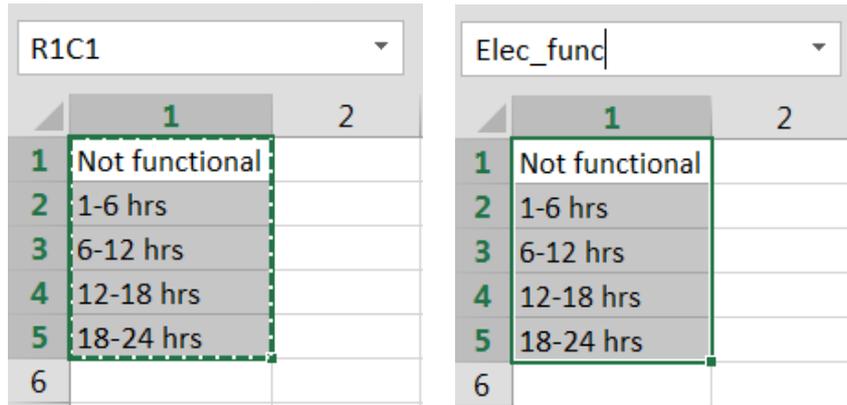
The named range will exist across the whole workbook, which is particularly useful if you have chosen to create a data model across several different worksheets. Another advantage is that it makes your formulas easier to read, therefore easier for the database to be used by others.

b) Setting up named ranges

Each range which you will reference should have a unique name. If you re-use the same list of options repeatedly within the database, you will only need to set up one named range.

To set up a named range, highlight the cells containing the options. Whilst highlighted, edit the **name box** next to the formula bar, entering a unique and meaningful name for the range, and press enter (figure 6).

Figure 6. Named range creation and storage



Set up a separate spreadsheet specifically for the domains (Figure 7, from the sheet 'domains' in example workbook), avoiding duplication if several questions have the same lists of response options.

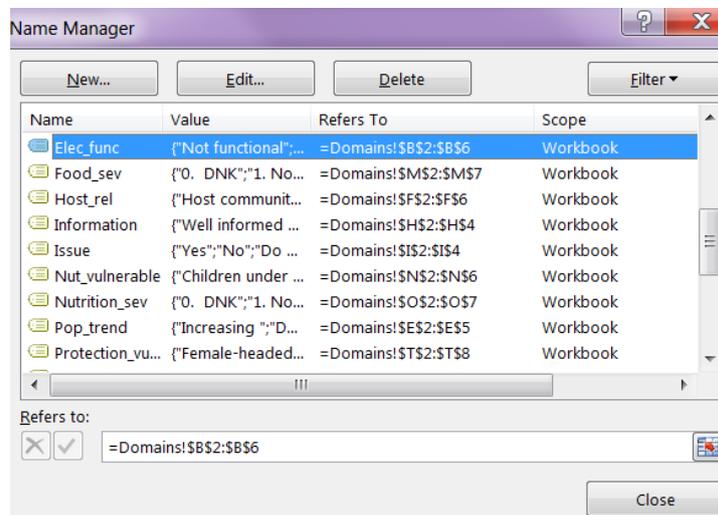
Figure 7. Named ranges stored with names

	1	2	3	4	5
1	Range name	Elec_func	Source_rel	Reg_status	Pop_trend
2	Values	Not functional	1=reliable	Yes (completed)	Increasing
3		1-6 hrs	2=fairly reliable	Yes (under way)	Decreasing
4		6-12 hrs	3= unreliable	Not yet, but scheduled	About the same
5		12-18 hrs		No	DNK
6		18-24 hrs			
7					

Show each range below the name of the range to make it easier to read and therefore manage. The domain sheet will need to always be kept with the database, to ensure these named ranges can be used within the database.

All of the names ranges within a workbook can be viewed and managed through the **Name Manager**, which can be found under the **formula** menu (Figure 8).

Figure 8. Name manager



c) Creating cascading drop-down menus (advanced)

In some circumstances, you may want the list of options in a dropdown menu to vary according to a previous choice. This is the case for administrative areas – once an Admin 1 area (in the example workbook, this is the Governorate) is chosen, offering only the admin 2 levels (District) which are found within that Governorate helps both to reduce the number of options to a more manageable list, and also ensures the validity of responses.

This section outlines one method of achieving cascading drop down menus, though there are several ways to add this functionality. It is recommended that you keep all of the administrative data in one sheet in the workbook, in which you can create all of the relevant lists.

Set up lists: In your spreadsheet of administrative areas, set up your administrative data so that you have:

- One list of unique Admin 1 (Governorate) areas
- One list of unique Admin 2 (District) areas, alongside a column of the Admin 1 areas these relate to.

Set up variables: Create named ranges which will be referenced in formulas:

- Admin 1 names (in the example workbook, the named range is called Governorate).
- The column header of the list of Admin 1 names belonging to the Admin 2 list (in the example, this is called StartGov).
- The whole column of the list of Admin 1 names belonging to the Admin 2 list (in the example, this is called ColGov).

Define formula: Create another named range containing the formula for looking up the set of Admin 2 names (in the example workbook, the named range is called District). You will need to do this from the Name manager console. The following formula can now be used, where Database!RC[-2], is the reference to the cell, two columns to the left which contain the selected Admin 1 name.

```
District=OFFSET(StartGov,MATCH(Database!RC[-2],ColGov,0)-1,1,COUNTIF(ColGov,Database!RC[-2]),1)
```

This works by finding the starting position of the Admin 1 name in the list alongside Admin 2 names, and returning the corresponding all Admin 2 names which have the same Admin 1 name.

Set validation: In the database, set up validation:

- For the Admin 1 name column (Column G in the example database), data validation is simply the list on unique Admin 1 names (Governorate)
- For the Admin 2 name column (Column I in the example database), data validation is set as the named range containing the formula from step c (in the example, District).

This process can be repeated for further cascading menus; in the example database, Admin 3 (Sub-district) dropdown menus are also dynamic, referencing the named range called SubDistrict, which contains the following code:

```
SubDistrict=OFFSET(StartDist,MATCH(Database!RC[-2],ColDist,0)-1,2,COUNTIF(ColDist,Database!RC[-2]),1)
```

This uses the selected District two columns to the left (Database!RC[-2]), and uses this to look through a third table which contains a unique list of Subdistricts, listed alongside their District and Governorate names, as shown in Figure 9.

Figure 9. Spreadsheet of administrative areas with additional Index columns for P-code lookup.

	6	7	8	9	10	11	12	13
1	Governate	Governate Code	District Name	District Code	Sub-district Name	Sub-district Code	DistrictIndex	Sub-districtIndex
2	Idleb	SY07	Ariha	SY0705	Ariha	SY070500	IdlebAriha	IdlebArihaAriha
3	Idleb	SY07	Ariha	SY0705	Ehsem	SY070501	IdlebAriha	IdlebArihaEhsem
4	Idleb	SY07	Ariha	SY0705	Mhambal	SY070502	IdlebAriha	IdlebArihaMhambal
5	Idleb	SY07	Jisr-Ash-Shugur	SY0704	Jisr-Ash-Shugur	SY070400	IdlebJisr-Ash-Shugur	IdlebJisr-Ash-ShugurJisr-Ash-Shugur
6	Idleb	SY07	Jisr-Ash-Shugur	SY0704	Badama	SY070401	IdlebJisr-Ash-Shugur	IdlebJisr-Ash-ShugurBadama
7	Idleb	SY07	Jisr-Ash-Shugur	SY0704	Darkosh	SY070402	IdlebJisr-Ash-Shugur	IdlebJisr-Ash-ShugurDarkosh
8	Idleb	SY07	Jisr-Ash-Shugur	SY0704	Janudiyeh	SY070403	IdlebJisr-Ash-Shugur	IdlebJisr-Ash-ShugurJanudiyeh

d) Creating a look-up for P-codes

Storing the full administrative hierarchy safeguards against information being misattributed to the wrong location. A further step of adding the p-code will greatly simplify the process of comparing this data to other datasets, mapping the information, and also providing it for others in an easy to use format.

By adding a few formulas at the time of database design, these p-codes can be automatically filled in. As with cascading menus, there are a number of ways In Excel for looking up information. The following method is simple to implement.

Create index columns: In your Administrative worksheet, ensure that you have the whole table of Administrative levels and corresponding p-codes.

- Add a column containing a concatenation of the Admin 1 and Admin 2 names using CONCATENATE formula:

=CONCATENATE(RC[-6],RC[-4])

- Add a further column concatenating Admin 1, Admin 2 and Admin 3 names (see Figure 9).

Create named ranges: Create the following:

- A named range of the whole look-up table (example: Lookuptable)
- Set up named range referencing the newly created Indexes (DistrictIndex and SubDistrictIndex) and also an Index for Governorates (GovernorateIndex). For this last Index, you can use the existing column of Governorate names (column 6 in the example).

Define formula for admin 1: In the cell where you would like the p-code for Admin 1 to be added, write the following formula, where RC[-1] references the cell with the Admin 1 name:

=INDEX(LookupTable,MATCH(RC[-1],GovernorateIndex,0),2)

This matches the Governorate name with the GovernorateIndex column, and returns the corresponding Governorate P-code in column 2 of the LookupTable.

Define formula for admin 2: In the cell where you would like the p-code for Admin 2 to be added, write the following formula, where RC[-3] is the cell containing the Admin 1 name, and RC[-1] is the cell containing the Admin 2 name:

```
=INDEX(LookupTable,MATCH(RC[-3]&RC[-1],DistrictIndex,0),4)
```

This matches the combination of the Governorate and District names in the adjoining cells with the DistrictIndex column, and returns the corresponding District P-code in column 4 of the LookupTable.

Define formula for admin 3: In the cell where you would like the p-code for Admin 2 to be added, write the following formula, where RC[-5] is the cell containing the Admin 1 name, RC[-3] is the cell containing the Admin 2 name, and RC[-1] contains the Sub-district name:

```
=INDEX(LookupTable,MATCH(RC[-5]&RC[-3]&RC[-1],SubDistrictIndex,0),6)
```

This matches the combination of the Governorate, District and Sub District names in the adjoining cells with the SubDistrictIndex column, and returns the corresponding Sub District P-code in column 6 of the LookupTable.

6 Testing your database

Once you have implemented all of the functionality outlined within the previous sections, there will be a number of different formulas, lookups and calculations embedded. Before starting data entry, you should have someone thoroughly test the database. This is best carried out by someone other than the developer, and should be done in conjunction with reviewing the data collection tools, to ensure that all of the data items in the data collection tools will be able to be recorded within the database.

Test the first empty row of the database. Once this is verified, the contents of these cells can be copied to subsequent rows.

Check the following:

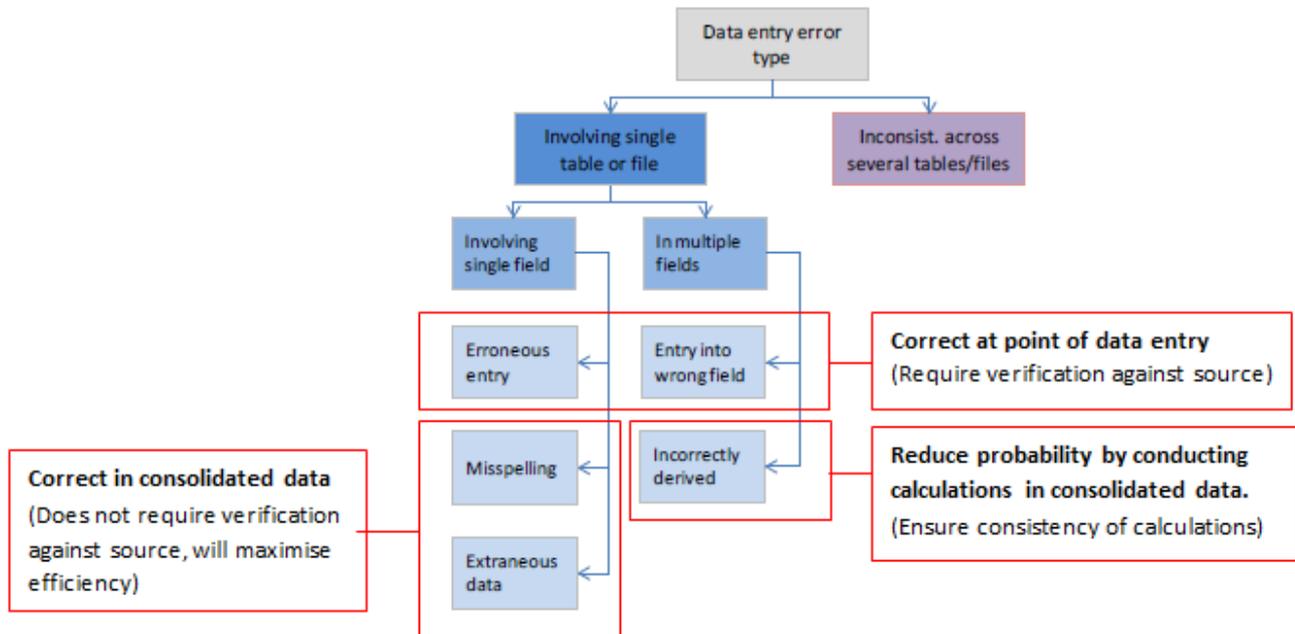
- Is there a corresponding field in the database for every data item within the data collection tools?
- Are all dropdown menus working correctly?
- Are all dropdown menus showing the right set of options?
- Are cascading menus working correctly?
- Are p-codes being filled in correctly?
- Have formulas in the first row been copied to subsequent rows?

7 Data cleaning and consolidation

Data cleaning is a critical stage before beginning analysis. Care should be taken to ensure data is as accurate and consistent (e.g. spellings, to allow aggregation) as possible. Whilst some errors may only be discovered during analysis, it is by far more effective to correct these in advance in order to avoid having to re-conduct analysis. This section focusses on deciding on your data cleaning strategy, as opposed to the data cleaning process.

When deciding upon an approach to data cleaning, it is useful to consider the different types of errors which can be made (see Figure 10 for a typology of data errors), and to plan at what point in your process you will try to identify them. Assuming a system where data entry is distributed across field locations and consolidation occurs in a different location, Figure 10 suggests where these errors should be corrected.

Figure 10. A typology of data errors and suggested correction points



a) Quality control of data entry

Erroneous errors (e.g. where data entered is different from the source yet valid, e.g. 26 instead of 25) and entries into the wrong field can only be rectified by comparison to source data, therefore should be done at the time of data entry when original sources are close at hand.

Missing values should also be examined during data entry. Depending on the question type, it may be necessary to differentiate a ‘no reply’ from a ‘do not know’ or a ‘no’ or a zero. During the data model design, ‘null’ codes should have already been implemented for such fields - ensure that data entry staff know when and how to use them.

These errors are easiest to identify when a **quality control** procedure exists, ensuring that a second pair of eyes compare source data to data entered. This will be particularly important when there is a process of translation at data entry, to ensure consistency/accuracy of translation.

b) Validation of rules during data entry

If there are additional ‘rules’ which should have been followed during data collection, ensure that data entry staff are familiar with them, so that these can be identified early on and verified/rectified, rather than adding them to the database (e.g. rules such as ‘pick only three’ or ‘must add to 100%’). If data entry will be conducted in distributed locations, document the rules to follow, where focus should be given, and how to solve errors/issues.

Additional functionality can be added in Excel to highlight rule violations, such as conditional formatting. The decision to include this in the database must be pragmatic, weighing up the merits of having errors detected and rectified by data entry staff, versus the time required to set this up.

c) Consolidating data from multiple sources

Your methodology design may require that data is entered in a number of different locations, by different people. This will not present an issue, so long as the same database structure is used; it will be possible to copy and paste additional rows of data into one master version, so long as no alterations are made to the structure (e.g. no additional columns added. Ensure that IDs are not duplicated across different database versions – this can be done by allocating a range of ID numbers to each data entry staff.

d) Cleaning of consolidated data

Misspellings and inconsistent spellings will have been largely avoided if you have implemented drop-down menus. If you have some fields where drop down menus were not used, you can quickly check for spelling inconsistencies by switching on auto-filtering functionality. When using the filter, each of the unique entries in the column will be listed, making it easy to spot items which should be the same but have been spelt in different ways. The find and replace function can then be used to quickly replace these.

Extraneous errors (where additional irrelevant information had been added) are best removed in the consolidated database, allowing a consistent approach to be applied.

Incorrectly derived errors (where some calculations have been applied incorrectly) can be reduced by conducting all calculations after consolidation to ensure consistency (e.g. converting households to individuals or vice versa).

e) Categorization of open response questions

If your database contains some 'open response' questions, or if you have added 'other' options to some of your categorical questions, you will be likely to need to categorise these into common responses.

Best practice is to add an additional field to contain the categorised responses (which is a derived field), leaving the original text behind – this allows you to always trace back to the original response and not to over clean your data.

8 Documenting changes

Documentation of error, alterations, additions and error checking is essential to:

- Maintain data quality
- Avoid duplication of error checking by different data cleaners
- Recover data cleaning errors
- Determine the fitness of the data for use
- Inform users who may have used the data knowing what changes have been made since they last accessed the data

In order to manage this process and track changes, create a **change log** within your workbook, where you will store all information related to modified fields. This will serve as an audit trail showing any modifications, and will allow a roll back to the original value if required. Within the change log, store the following fields:

- Table (if multiple tables are implemented)
- Column, row
- Date changed
- Changed by
- Old value
- New value
- Comments

Figure 11. Sample change log from J-RANS Assessment

Date	Time	Changed By	Change	New value
29-Jan	10:45	Georges	Cleaning actors across all sectors	
29-Jan	11:00	Henry	Added recategorisation of priority needs	Refer code book
04-Feb	21h55	Henry	Addition of the last questionnaire Deir ez Zor	
05-Feb	11h33	Georges	Duplication between questionnaire 49 and 59 solved. Contact with the enumerators. It was decided to delete the questionnaire 49 (no possible contact with the enumerators) and to replace it by questionnaire 59 (after debriefing), considered more reliable	Questionnaire 49 deleted and replace by 59
05-Feb	15h37	Christine	Double and triple check on Q59 pop figures. Initial figures reflected pop number at district level and not sub district. Contact with inside informant to change pop figures	Change from 256.000 to 345.000
05-Feb	16h15	Georges	Dana agencies updated after MdM input	
06-Feb	22h31	Christine	Number of arrested change in Q49. DNK instead of 20000	from 20.000 to DNK
07-Feb	17h05	Henry	Complete cleaning of the demographic section. Creation of a second tab called Database pop figures OK where sum of IDPs in public building+host+vacated building = total number of IDPs. See tab Changes in pop figures for more details	

Always make this information available when sharing the dataset internally or externally (i.e. by enclosing the change log in a separate worksheet).

9 Additional Resources

Database design in Excel:

<http://www.und.edu/dept/cndtrain/Excel/database.pdf>

http://spreadsheets.about.com/od/datamanagementinexcel/ss/excel_database.htm

<https://intranet.birmingham.ac.uk/as/claddivision/skills/documents/public/excel3.pdf>

Change logs:

<http://www.codeproject.com/Articles/105768/Audit-Trail-Tracing-Data-Changes-in-Database>

ANNEX: Joint Rapid Assessment for Aleppo City Form 2013 (J-RANS 2013)

Questionnaire ID:		Contested: (y/n)		Names of neighbourhoods covered (identical to MAP):	1.	4.
Date (dd/mm/yy):		# of Neighbourhoods covered in this form:			2.	5.
Team name/code:					3.	6.

A. Damages by Conflict				
Type: (INGO, Committee, local group, health staff, other):				Reliability**
Main Source				
A1. Due to conflict number of persons:*				
	Total	Male	Female	Of whom Children < 5 yrs
Dead				
Injured				
Missing				
Arrested				
**(Rating: 1=reliable, 2=fairly reliable, 3=unreliable)				
A2. Due to conflict damages of physical infrastructure (enter in %) Total for each column should be 100%				
Type: (INGO, Committee, local group, health staff, other):				Reliability rate
Main Source:				
Description	Private Buildings (houses, apartment buildings, etc.)	Public Infrastructure (schools, health centres, etc.)		
No damages				
Slight damages: light repairs required (windows, doors)				
Moderate damages: Under 30% roof damage, fire damage, can be repaired				
Heavy damage: Over 30% roof damage, severe fire damage, can be repaired				
Destruction: Unusable, houses levelled, can't be repaired				
A3. Electricity (per day, over the past 30 days)				
<input type="checkbox"/> Not functional <input type="checkbox"/> 1-6 hrs <input type="checkbox"/> 6-12 hrs <input type="checkbox"/> 12-18 hrs <input type="checkbox"/> 18-24hrs				
B. Demography*				
Type: (INGO, Committee, local group, health staff, other):				Reliability**
Main Source				
B1. Estimated # of population in neighbourhood:			Total	% Female
Total # of pre-conflict population (2011)				
Of whom # who have fled the neighbourhood				
Current total # of population (resident population + new arrivals, at this moment)				
- Of whom total # of displaced population (total # of below groups)				
- # Displaced people living in collective accommodation				
- # Displaced people hosted by local families				
- # Displaced people in vacated buildings				
* '0'=not present ; 'DNK'=Don't know ; otherwise provide point estimate **(Rating: 1=reliable, 2=fairly reliable, 3=unreliable)				
B2. Have the displaced / crisis-affected people been registered in this neighbourhood?				
<input type="checkbox"/> Yes (completed)		<input type="checkbox"/> No		
<input type="checkbox"/> Yes (under way)		<input type="checkbox"/> Not yet, but scheduled		
If yes, which organization conducted the registration in this neighbourhood?				
B3. Is the population increasing, decreasing, or staying about the same in this neighbourhood? Ask this question to more than one person –LCC/Local authorities, IDPs, neutral party (i.e. NGO)				

<input type="checkbox"/> Increasing <input type="checkbox"/> Decreasing <input type="checkbox"/> About the same <input type="checkbox"/> DNK				
B4. How is the relationship between the displaced and the host community in this neighbourhood? Select only one				
<input type="checkbox"/> Host community willing to assist for as long as necessary <input type="checkbox"/> Host community willing to assist, but for limited time		<input type="checkbox"/> Tensions already exist <input type="checkbox"/> Other (specify _____) <input type="checkbox"/> Not applicable		
C. Humanitarian access				
C1. Humanitarian Access: Are there problems to gain access to aid in this neighbourhood? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know				
If yes, how severe are the following problems: (Tick only box one per problem)				
	Severe problem	Problem	Limited problem	No problem
Restriction of movement for people				
Interference into humanitarian activities				
Violence against personnel, facilities and assets				
Restriction and obstruction of access to aid				
Active hostilities				
Presence of mines and explosives				
D. Information				
D1. Is humanitarian assistance provided in this neighbourhood over the past 30 days? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know				
If yes, are people generally: (Select only one)				
<input type="checkbox"/> Well informed about humanitarian assistance <input type="checkbox"/> Poorly informed about humanitarian assistance <input type="checkbox"/> Not at all informed about humanitarian assistance				
E. Health				
E1. Health Status: Is there a serious problem regarding health in this neighbourhood? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know				
If yes, I am reading a list of possible problems: (Select max five most serious problems)				
<input type="checkbox"/> Numerous cases of psychological trauma (anxiety, depression, phobia, etc.) <input type="checkbox"/> Numerous injured less than 6 months ago <input type="checkbox"/> Numerous injured more than 6 months ago <input type="checkbox"/> Numerous disabled with limitation to move (amputation, spinal cord injury, brain injury, or peripheral nerve injury) <input type="checkbox"/> Numerous cases with other disabilities (hear, see, speak)		<input type="checkbox"/> Incidents of communicable diseases (measles, tetanus, scabies, cholera, etc.) <input type="checkbox"/> Numerous cases of chronic diseases (arthritis, dialysis, etc.) <input type="checkbox"/> Numerous cases of diarrhoea <input type="checkbox"/> Numerous cases of fever <input type="checkbox"/> Numerous cases of respiratory diseases <input type="checkbox"/> Numerous cases of pregnancy related diseases <input type="checkbox"/> Other: _____		
E2. Health Care: Is there a serious problem because people are not able to get adequate health care for themselves in this neighbourhood? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know				
If yes, I am reading a list of possible problems: (Select max five most serious problems)				

ANNEX: Joint Rapid Assessment for Aleppo City Form 2013 (J-RANS 2013)

<input type="checkbox"/> Not enough health facilities available <input type="checkbox"/> Lack of ambulance services <input type="checkbox"/> Lack of medicines <input type="checkbox"/> Lack of mobility devices (wheelchairs, prosthetics, others) <input type="checkbox"/> Not enough rehabilitation services <input type="checkbox"/> Lack of medical staff	<input type="checkbox"/> Not enough access to health services due to physical/logistical constraints <input type="checkbox"/> Not enough access to health services due to security constraints <input type="checkbox"/> Not enough access to health services due to limited economic resources (lack of money) <input type="checkbox"/> Other: _____
---	---

E3. Which specific health interventions are most urgently required in this neighbourhood? (Enter short description)
 Do not know

First rank:

Second rank:

Third rank:

E4. Overall, which of the following statements describes best the general status of public health in this neighbourhood? (circle right answer)

0. DNK
1. No concern – situation under control
2. Situation of concern that requires monitoring
3. Many people will suffer if no health assistance is provided soon
4. Many people will die if no health assistance is provided soon
5. Many people are known to be dying right now because of insufficient health services

Main reason for selecting category: (add short text)

E5. Distance and capacity of next functional hospital:

Distance (in travel time) _____ minutes

E6. Which group faces the biggest health risks in this neighbourhood? (Rank top three: 1=first rank, 2=second rank, 3=third rank)

- ___ Displaced people living in host families
- ___ Displaced people in collective shelter (schools, camps, etc.)
- ___ Displaced people in vacated buildings
- ___ Resident population hosting displaced persons
- ___ Resident population who have not been displaced

E7. Which organisations have been providing regular health care services in this neighbourhood over the past 30 days?

Type (INGO, Local Org, Self-help group, other)	Organisation responsible	Type of regular support (excluding one-offs)

F. Food

F1. Is there a serious problem regarding food in this neighbourhood? Yes No Do not know

If yes, I am reading a list of possible problems: (Select max five most serious problems)

<input type="checkbox"/> Not enough food available (including in markets, etc.) <input type="checkbox"/> Not enough diversity in food <input type="checkbox"/> Not enough access to markets due to physical/logistical constraints (transport) <input type="checkbox"/> Not enough access to food sources (i.e. markets) due to security constraints <input type="checkbox"/> Not enough access to markets due to limited economic resources (income)	<input type="checkbox"/> Price increase of basic food items <input type="checkbox"/> Agricultural production is disrupted <input type="checkbox"/> There are not enough cooking facilities or utensils <input type="checkbox"/> Not enough cooking fuel <input type="checkbox"/> Loss of economic assets due by conflict (livestock, machinery, seeds, etc.) <input type="checkbox"/> Other: _____
---	---

F2. Which specific food security interventions are most urgently required in this neighbourhood? Do not know

First rank:

Second rank:

Third rank:

F3. Are there functional bakeries regularly providing bread to the people in this neighbourhood? Yes No Do not know (bag = 6-7 loafs)

If yes, what is their normal capacity (tons of wheat flour processed per day) _____ (tons)

What is their current output (tons wheat flour processed per day) _____ (tons)

Price of subsidized bread (per bag): _____ SYP

Price on the street (per bag, not subsidized): _____ SYP

F4. Overall, which of the following statements describes best the general status of food security in this neighbourhood? (Circle right answer)

0. DNK
1. No concern – situation under control
2. Situation of concern that requires monitoring
3. Many people will suffer if no food assistance is provided soon
4. Many people will die if no food assistance is provided soon
5. Many people are known to be dying right now due to lack of food

Main reason for selecting category: (add short text)

F5. Which group is most at risk of having not enough food to survive in this neighbourhood? (rank top three: 1=first rank, 2=second rank, 3=third rank)

- ___ Displaced people living in host families
- ___ Displaced people in collective shelter (schools, camps, etc.)
- ___ Displaced people in vacated buildings
- ___ Resident population hosting displaced persons
- ___ Resident population who have not been displaced

F6. Which organizations have been providing regular food support in this neighbourhood over the past 30 days?

Type (INGO, Local Org, Self-help group, other)	Organisation responsible	Type of regular support (excluding one-offs)

G. NUTRITION

G1. Nutritional Status: Is there a serious problem regarding nutrition in this neighbourhood? Yes No Do not know

If yes, who in this neighbourhood do you think are the most vulnerable to the issue of poor nutrition: (Select only one most vulnerable group)

ANNEX: Joint Rapid Assessment for Aleppo City Form 2013 (J-RANS 2013)

<input type="checkbox"/> Children under 6 months <input type="checkbox"/> Children under 5 years <input type="checkbox"/> Children over 5 years <input type="checkbox"/> Pregnant and lactating women <input type="checkbox"/> Other: _____		
G2. Are mothers facing a problem with feeding their babies? If yes, what are some of the reasons mothers are facing trouble feeding:		
<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know		
If yes, I am reading a list of possible problems: (Select max five most serious problems)		
<input type="checkbox"/> Women are unable to breastfeed due to stress/fear <input type="checkbox"/> Women are unable to breastfeed due to insufficient food availability <input type="checkbox"/> Women are unable to breastfeed due to lack of privacy <input type="checkbox"/> Women are unable to access breastfeeding support	<input type="checkbox"/> Lack of infant formula in the markets <input type="checkbox"/> Lack of fuel/water/sterilizing equipment for preparation of infant formula <input type="checkbox"/> Unsolicited / untargeted distributions of infant formula (<i>milk or powder</i>) ongoing <input type="checkbox"/> Other: _____	
G3. Which specific nutrition interventions are most urgently required in this neighbourhood?		
<input type="checkbox"/> Do not know		
First rank:		
Second rank:		
Third rank:		
G4. Overall, which of the following statements describes best the general nutritional status in this neighbourhood? (Circle right answer)		
0. DNK 1. No concern – situation under control 2. Situation of concern that requires monitoring 3. Many people will <u>suffer</u> if no nutrition assistance is provided soon 4. Many people <u>will die</u> if no nutrition assistance is provided soon 5. Many people are known to be <u>dying right now</u> because of insufficient nutrition services		
Main reason for selecting category: (add short text) _____		
G5. Which group faces the biggest risks of malnutrition in this neighbourhood? (rank top three: 1=first rank, 2=second rank, 3=third rank)		
<input type="checkbox"/> Displaced people living in host families <input type="checkbox"/> Displaced people in collective shelter (schools, camps, etc.) <input type="checkbox"/> Displaced people in vacated buildings <input type="checkbox"/> Resident population hosting displaced persons <input type="checkbox"/> Resident population who have not been displaced		
G6. Which organisations have been providing regular nutrition services in this neighbourhood over the past 30 days?		
Type (INGO, Local Org, Self-help group, other)	Organisation responsible	Type of regular support (excluding one-offs)
H. Places to live in and non-food items (NFI)		
H1. Is there a serious problem in this neighbourhood regarding shelter?		
<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know		
If yes, I am reading a list of possible problems: (Select max five most serious problems)		

<input type="checkbox"/> Not enough shelter space available <input type="checkbox"/> Not enough protection against cold (<i>snow, wind, rain</i>) <input type="checkbox"/> Not enough access to privately rented shelter space <input type="checkbox"/> Not enough access to collective shelter space (<i>lack of facilities/overcrowded</i>)	<input type="checkbox"/> Not enough access to building materials due to physical/logistical constraints <input type="checkbox"/> Not enough access to building materials due to security constraints <input type="checkbox"/> Not enough access to building materials due to limited economic resources (<i>income</i>) <input type="checkbox"/> Other (Specify): _____	
H2. Which specific shelter interventions are most urgently required in this neighbourhood?		
<input type="checkbox"/> Do not know		
First rank:		
Second rank:		
Third rank:		
H3. Is there a serious problem in your neighbourhood regarding Non Food Items?		
<input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know		
If yes, I am reading a list of possible problems: (Select max five most serious problems)		
<input type="checkbox"/> Lack of cooking utensils (<i>pots, dishes, utensils</i>) <input type="checkbox"/> Lack of household lights <input type="checkbox"/> Lack of adult clothing/shoes <input type="checkbox"/> Lack of child clothing/shoes <input type="checkbox"/> Lack of baby supplies (<i>diapers, etc.</i>)	<input type="checkbox"/> Lack of personal hygiene products (<i>nail clippers, toothbrush</i>) <input type="checkbox"/> Lack of female hygiene products (<i>sanitary pads, underwear</i>) <input type="checkbox"/> Lack of mattresses and blankets <input type="checkbox"/> Other (Specify): _____	
H4. Which specific NFI interventions are most urgently required in this neighbourhood?		
<input type="checkbox"/> Do not know		
First rank:		
Second rank:		
Third rank:		
H5. Overall, which of the following statements describes best the general status of Shelter and NFIs?		
0. DNK 1. No concern – situation under control 2. Situation of <u>concern</u> that requires monitoring 3. Many people will <u>suffer</u> if no shelter assistance is provided soon 4. Many people <u>will die</u> if no shelter is provided soon 5. Many people <u>are known to be dying right now</u> due to lack of shelter		
Main reason for selecting category: (add short text) _____		
H6. Which group is most at risk due to lack of shelter and NFIs? (rank top three: 1=first rank, 2=second rank, 3=third rank)		
<input type="checkbox"/> Displaced people living in host families <input type="checkbox"/> Displaced people in collective shelter (schools, camps, etc.) <input type="checkbox"/> Displaced people in vacated buildings <input type="checkbox"/> Resident population hosting displaced persons <input type="checkbox"/> Resident population who have not been displaced		
H7. Which organizations have been providing regular shelter and NFI support in this neighbourhood over the past 30 days?		
Type (INGO, Local Org, Self-help group, other)	Organisation responsible	Type of regular support (excluding one-offs)
I. Water, Sanitation and Hygiene		
I1. What is the main water source in this neighbourhood?		

<input type="checkbox"/> Piped water system	<input type="checkbox"/> Rain water harvesting	
<input type="checkbox"/> Stream, river or hillside spring	<input type="checkbox"/> Private well	
<input type="checkbox"/> Water truck	<input type="checkbox"/> Other (Specify): _____	
I2. Is there a serious problem regarding water in this neighbourhood? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know		
If yes, I am reading a list of possible problems: <i>(Select max five most serious problems)</i>		
<input type="checkbox"/> Lack of jerry cans	<input type="checkbox"/> Not enough water available because water is too far away or difficult to access	
<input type="checkbox"/> The water available is not safe for drinking	<input type="checkbox"/> Not enough water available because people don't have means to store water	
<input type="checkbox"/> Water does not taste good or does not look good enough	<input type="checkbox"/> Not enough water available because water system, well or pump is broken	
<input type="checkbox"/> Lack of ways to treat water or fuel for boiling it	<input type="checkbox"/> Other (Specify): _____	
<input type="checkbox"/> Not enough water available because water too expensive		
I3. Overall, which of the following statements describes best the general status of water supply? (Circle one right answer)		
0. DNK		
1. No concern – situation under control		
2. Situation of <u>concern</u> that requires monitoring		
3. Many people <u>will suffer</u> due to lack of water		
4. Many people <u>will die</u> if insufficient water remains available		
5. Many people <u>are known to be dying right now</u> due to lack of water		
Main reason for selecting category: (add short text) _____		
I4. Regarding the lack of safe water, which group is most at risk? (rank top three: 1=first rank, 2=second rank, 3=third rank)		
___ Displaced people living in host families		
___ Displaced people in collective shelter (schools, camps, etc.)		
___ Displaced people in vacated buildings		
___ Resident population hosting displaced persons		
___ Resident population who have not been displaced		
I5. Is there a serious problem regarding sanitation and hygiene in this neighbourhood? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know		
If yes, I am reading a list of possible problems: <i>(Select max five most serious problems)</i>		
<input type="checkbox"/> Not enough places to wash your body or bathe	<input type="checkbox"/> Not enough access to toilets due to security constraints	
<input type="checkbox"/> Not enough access to water, soap or places to wash due to security constraints	<input type="checkbox"/> Not enough access to toilets because they are too far away	
<input type="checkbox"/> Not enough access to water or soap because the cost is too expensive	<input type="checkbox"/> Not enough access to toilets because they are not segregated	
<input type="checkbox"/> Not enough toilets available for men	<input type="checkbox"/> No regular rubbish collection so general waste builds up	
<input type="checkbox"/> Not enough toilets available for women	<input type="checkbox"/> Others: _____	
I6. Which specific water, sanitation, and hygiene interventions are most urgently required? <input type="checkbox"/> Do not know		
First rank:		
Second rank:		
Third rank:		
I7. Which organizations have been providing regular water, sanitation or hygiene support in this neighbourhood over the past 30 days?		
Type (INGO, Local Org, Self-help group, other)	Organisation responsible	Regular support (excluding one-offs)

J. EDUCATION	
Number of functional schools in this neighbourhood before the conflict	Number of functional schools today in this neighbourhood (used for education)
J1. What percentage of children (6-14 yrs of age) is regularly attending school in this neighbourhood?	
<input type="checkbox"/> 0-25% <input type="checkbox"/> 26-50% <input type="checkbox"/> 51-75% <input type="checkbox"/> 76-100%	
J2. What are the reasons why children are not attending schools? (Select all that apply)	
<input type="checkbox"/> Schools not functioning (damaged, destroyed or occupied)	<input type="checkbox"/> Lack of school materials (stationery, books, etc.)
<input type="checkbox"/> Safety - fear of schools being bombed/targeted	<input type="checkbox"/> Lack of water and hygienic sanitation facilities in schools
<input type="checkbox"/> Lack or absence of teachers	<input type="checkbox"/> Other (specify): _____
J3. Are education activities taking place in other locations? (e.g. home, mosque, etc.) <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know	
If yes, add main type of location: _____	
J4. What percentage of children (6-14 yrs of age) is regularly receiving education in these other locations?	
<input type="checkbox"/> 0-25% <input type="checkbox"/> 26-50% <input type="checkbox"/> 51-75% <input type="checkbox"/> 76-100%	
K. Protection	
Type: (INGO, Committee, local group, health staff, other):	Reliability rate
Main Source:	
K1. Is there a serious problem in your neighbourhood regarding protection issues for vulnerable groups? <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Do not know	
If yes, what are the three main problems: <i>(See separate list for guidance)</i>	
First rank:	
Second rank:	
Third rank:	
K2. Which specific interventions to protect vulnerable persons are most urgently required in this neighbourhood? <input type="checkbox"/> Do not know	
First rank:	
Second rank:	
Third rank:	
K3. What are the most vulnerable groups in this neighbourhood? (Select only one)	
<input type="checkbox"/> Female-headed households	<input type="checkbox"/> Families belonging to ethnic / religious minorities
<input type="checkbox"/> Elderly headed households	<input type="checkbox"/> Children without appropriate family care / orphans
<input type="checkbox"/> Households with disabled persons	<input type="checkbox"/> Other. Specify _____
<input type="checkbox"/> Destitute families	
K4. What are the structures in the area that are responsible for the protection of vulnerable persons in this neighbourhood? (Select max five)	
<input type="checkbox"/> Local Council	<input type="checkbox"/> Local police
<input type="checkbox"/> Community based structures / groups / committees	<input type="checkbox"/> Family
<input type="checkbox"/> Local charities	<input type="checkbox"/> No structures responsible for protection in the area
<input type="checkbox"/> Religious leaders	<input type="checkbox"/> Other. Specify: _____
<input type="checkbox"/> Schools	

K5. Which groups contain the most vulnerable people in this neighbourhood? (Rank top three: 1=first rank, 2=second rank, 3=third rank)

Displaced people living in host families
 Displaced people in collective shelter (schools, camps, etc.)
 Displaced people in vacated buildings
 Resident population hosting displaced persons
 Resident population who have not been displaced

K6. Which organisations have been providing regular protection services in this neighbourhood over the past 30 days?

Type (INGO, Local Org, Self-help group, other)	Organisation responsible	Type of regular support (excluding one-offs)

L. Sector Prioritization

After these specific questions, we want to recapitulate. In terms of which sector poses the most serious problems, can you say which is the most serious, second most, third most, fourth most, and fifth most serious? I read you a list of 7 sectors:

L1. Priority Level. Rank a maximum of 5: 1=first priority, 2=second priority, 3=third priority., 4=fourth priority; 5= fifth priority

	Health
	Food Security
	Nutrition
	Water, Sanitation, Hygiene
	Places to live and Non-Food Items
	Education
	Protection

L2. Are there any other urgent problems in this neighbourhood, which I have not yet asked you about? (Please write down bullet points only)

L3. Any further observations from the assessment team on the difficulty to collect information or the situation in the neighbourhood (Please elaborate as required)