

Aldo Benini

A note for ACAPS

A template for managing data in needs assessments

Summary

In needs assessments as in countless other fields, data management forms the bridge between the collection of data and the analysis. This note speaks to assessment teams using a particular format: assessments combining several levels (sites, sectors and key issues), yet centered on the severity of needs.

ACAPS has led several recent assessments in this line. The experience in Bolivia, Yemen and Bangladesh is particularly valuable. However, the way the data was structured in spreadsheets was not always optimal.

A new generic Excel workbook template improves on the previous handling. It strives for easy adaptation to new disaster contexts, speedy data entry, and readiness for analysis. This tool may be appropriate for the management of data collected at the community level during Phase 2 of assessments.

This note details the rationale for the template as well as its likely benefits and drawbacks. It lays out the step-by-step instructions for its use in the current shape and for later adaptation. A companion note on the subsequent analysis phase, using real data from Yemen, complements it.

Contents

Summary	1
Introduction.....	4
A generic template for data management	5
Expected benefits	6
Expected challenges.....	7
[Sidebar:] Predecessors of the template	7
The nature of entities	9
[Sidebar:] The severity of needs - How many levels?	10
Architecture.....	11
[Sidebar:] Required user skills.....	13
Usage and adaptation	14
"Once the template is ready, how do I use it?"	14
[Sidebar:] Tags.....	17
"How do I combine data from several computers or from several locations?"	18
"How do I recode problems and recommendations?"	21
[Sidebar:] Free-text entries - an illustration	22
"How do I adapt the template to the requirements of this new assessment?"	23
Changing variables.....	23
Changing category sets	24
Changing sector-specific category sheets	24
Outlook	27
Appendix.....	29
Entity relationships	29
References.....	30

Figures and tables

Figure 1: Yemen assessment data arrangement (segment)	8
Figure 2: Color scheme for severity ratings, four levels	10
Figure 3: Alternative severity rating scheme, five levels	10
Figure 4: Template architecture	12
Figure 5: Drop-down menu for the "Problem" field	15
Figure 6: Example of a recoding scheme	20
Figure 7: Example of data validation settings for "Problem" drop-down menus	26
Figure 8: Entity relationship diagram	29
Table 1: Free-text entries of problems, Bolivia (segment)	22
Table 2: Examples of dynamic named ranges	25

Introduction

In needs assessments as in countless other fields, data management forms the bridge between the collection of data and its analysis. The ways data is managed are fashioned by a mixture of universal best practices and particular institutional factors¹. For example, the correct formatting of a data table in a spreadsheet follows a small set of rules that serve all users well. By contrast, many other operations vary widely across the data management community. Technology, professional and institutional traditions, and the objectives of analyses are the main forces that stimulate diversity in tools and formats.

An example from each may suffice: The spread of inexpensive and decentralized computing has made rapid, yet sizable surveys possible; the Internet and mobile telephony are further accelerating such capabilities. The widespread mastery of spreadsheets among professionals and volunteers facilitates networked data collection and exchange while at the same time distinct formats and styles are enshrined in particular institutional fields. And while one may naively think that the best ways of managing data should serve any analytical purpose equally well, practically we see data generated, managed and transformed in the mould of guiding analytical concerns.

These general considerations hold also for the management of data that is collected in multi-sectoral needs assessments after disasters. The methodology for these has been advanced under the "Multi-Cluster/Sector Initial Rapid Assessment (MIRA) Approach" (NATF 2011) coordinated by UNOCHA and supported by the work of ACAPS.

In fact, ACAPS has led several inter-sectoral needs assessments in recent times. Their data management exhibited three common characteristics:

1. The time pressure was enormous, all along from design to reporting, and thus also for the data management.
2. The data speaks to a hierarchy of entities, for which in less improvised environments a relational database would be suitable, but which had to be handled in a spreadsheet-based flat table.
3. The core analytic interest was built around ratings and rankings expressing severity and priority, and giving rise to ordinal-type data, with its own formatting (and subsequently analytical) challenges.

Similar formats may be continued in future assessments. They may face new and potentially steep challenges if they are used during the so-called "Phase 2" of assessment after a sudden-onset disaster. Phase 2 is expected to produce, within two weeks, the "MIRA Report" or, as it was called earlier, the "Initial Rapid Assessment" (IASC 2007).

¹ We follow in this note and elsewhere the consistent abuse of the term "data" as a collective singular and therefore use expressions like "the data is ..", "all this data ..", despite the obvious Latin plural form.

Under this stringent deadline, data management will require adaptations not only to local contexts, but also to improved data handling. This is where this note comes in.

We recognize that there are many suitable data management formats for rapid needs assessments. What follows is largely fashioned by one particular tradition - formats that grew under IASC premises and have been used in several assessments. Notably, this line of assessment is centered on the concept of "key issues" and on how severe the needs associated with those issues are (the "key issues" have frequently been called "problems", the term this note uses henceforth). The severity judgments, by key informants or field teams, elicited sector-by-sector while the assessment proceeds in a given community or site, are the core data. The practical execution is most clearly demonstrated by the assessments that ACAPS led in Bolivia and Yemen in 2011.

A generic template for data management

This note introduces a generic template for the management of such data. Its purpose is to

- make data entry and subsequent analysis efficient and thus mitigate time pressure
- maintain multi-entity data handling in spreadsheets, on account of their great flexibility and familiarity among users.

"Data management" here means:

- data entry,
- combining the contributions from several entry locations,
- preparation for analysis.

The next sections discuss benefits and limitations of the template, walk the reader through its features, and explain how it can be modified. We encourage reading these pages side by side with the opened demonstration workbook. A companion note and workbook address the analysis of data arranged through this template; they make use of the Yemen assessment to demonstrate the process. For some readers, this demonstration, using real data, may make several of template's features better understandable than the generic exposition here.

Again, we repeat: the template has been created for the handling of data collected in a particular assessment tradition. Other traditions will call for different data architectures - although we hope that they would easily and usefully adopt elements of this template.

Expected benefits

Design flexibility

The template allows assessment teams to add, subtract or modify variables in flexible manner. While being flexible, it preserves a core structure for the placement of problems and their associated severity ratings.

Pre-formatted records

When data arrives to be entered about another site, a macro button click adds a block of pre-formatted records at the bottom of the data table. The pre-formatting supports navigation and reduces entry error. In addition, two core variables - "Problem" and "Recommendation" - are entered using drop-down menus, with options controlled in separate tables, one for each sector. These entries can be automatically recoded, if desired, with values specified in the same sector tables.

Preparations for efficient analysis

The extent of built-in analysis preparation is minimal, to maintain flexibility. Yet the template provides useful devices that are often underrated in the humanitarian information management community:

- Sufficient record identifiers to preserve the original sort order and locate the affected community in the administrative hierarchy
- Tagging of records for entities of different levels - site, site-sector combinations, problems -, so that cross-tabulations can be appropriately filtered
- Rapid recoding of two core variables during analysis - "Problem" and "Recommendation" - to regroup categories for meaningful summaries.

Suitable for different assessment scenarios

Taken together, these features:

- easy modification of variables
- incremental addition of pre-formatted records when more data arrives
- drop-down menus for the entry of problems and recommendations, and
- optional automatic recoding of problems and recommendations

are helpful for several assessment scenarios. They facilitate, if logistically desirable, *decentralized data entry* and the efficient integration of partial (e.g. by province) data tables into one combined central (e.g. national) master table. They enhance and speed up analysis because two- and three-way *tabulations can be built and filtered in flexible ways*. The recoding feature allows assessment teams to *experiment with different formulations* of problems and recommendations and quickly see the effect of changes on findings. It makes it possible for regional teams to use their own problem categories and later to *integrate sets of diverse categories in a unified scheme*.

Expected challenges

At the same time, the template inherits ambitions (and add some of its own) that challenge data management and analysis:

- The template packs several layers of information all into one spreadsheet table. This demands extra vigilance on the part of the data manager and careful briefing of team members not familiar with the intent of the template.
- In keeping with past assessments, the template pre-formats, for each combination of sites and sectors, seven blank records to receive information on problems and recommendations. An eighth record is added in order to hold a summary evaluation. This space is more than enough to accommodate the information except in exceedingly rare cases of long and detailed problem lists. However, it entails that the data table gets bloated with numerous empty records. Such "vertical incompleteness" is frowned upon by data managers. A built-in tag makes it easy to filter them out and, once the database is final, to delete them. But there is always a residual risk of misunderstanding and mischief by insufficiently briefed team members.
- This risk looms also about three variables that are automatically updated through built-in formulas. The record number, recoded problem and recoded recommendation fields depend on them. Once the database is final, the record numbers should be made static (i.e., the formulas replaced with their values). Until then, users must not sort the table. If and when the assessment team shares the table with outsiders, also the recoded problems and recommendations should be frozen. These requirements may strike some users as unusual and burdensome.

It is also recognized that ACAPS and its partners have not finally determined some major design elements, such as the use of the so-called HESPER Scale (WHO and King's College London 2011). The template includes some place-holders for their possible addition.

[Sidebar:] Predecessors of the template

The template reflects a tradition that grew out of deliberations in the Needs Assessment Task Force, of which ACAPS is a member. The instrument that they produced received approval from the United Nations Inter-Agency Standing Committee (IASC 2009; NATF 2010). Its core elements - the hierarchy of sites, sectors and problems rated for the severity of their attendant needs - were practiced in assessments in Bolivia, Yemen and, to a smaller degree, in Bangladesh. Some are included also in Cluster-specific instruments, e.g. in the Joint Education Needs Assessment Toolkit (Global Education Cluster 2010).

The Yemen case is particularly illuminating. Key elements for our purpose are easily illustrated with a screenshot from its master workbook.

Figure 1: Yemen assessment data arrangement (segment)

1		Sector	Problems Identified in priority order (max 5)	Severity of Needs				Total	Recommendations for Intervention
				Rank					
Governorate	Amran	Livelihood	Income/employment			x		3	Income/Employment
District	Amran		Cash for basic services			x		3	Cash programming
Site Name	Alhboba-BeirHirab		Lack of skills		x			2	Vocational training
Site category	Village/part of town							NA	
Urban Rural	Urban							NA	
Total # population	11680								
Total # IDPs	3240								
		Synthesis				x		3	
Target Group	Vulnerable IDPs	WASH	Water supply/management			x		3	Provision of water, tanks
Gender	Male		Water sources		x			2	Rehabilitation of sources
Average age	30		WASH NFIs status	x				1	Provision of Hygiene Items
			Water quality			x		3	Water treatment (filters)
								NA	
Priority 1							NA		
Sector	Cross-cutting	Synthesis				x		3	
Intervention	Provision of support to nev	Shelter	Shelter Security/Condition			x		3	Shelter repairs
AT comment			NFI status					NA	Provision of NFIs
Priority 2			Heating/Cooking		x			2	Provision of heating fuel
Sector	Health							NA	
Intervention	Access to life saving emerg						NA		
AT comment									
Priority 3		Synthesis				x		3	
Sector	Food Security	Food Security	Food availability			x		3	Food supply
Intervention	Food rations and gas for th		Food accessibility			x		3	Registration, advocacy
AT comment	Winterization (blankets, et							NA	
								NA	
								NA	
		Synthesis				x		3	

The segment is from a table called "Data". But the format is a far cry from any standard data table. The "Data" layout rather mimics part of the assessment process in the field; as such it makes it easy to intuit the transfer from filled-in questionnaires to a data entry mask of this shape. The following elements are constitutive, regardless of how many sector records and how many assessed sites are eventually stacked upon each other:

1. The upper left corner holds site-level identifying as well as basic descriptive information. Since only one field is provided each for target group and gender, by implication the presence of two groups or genders at the same site will necessitate separate assessment records.
2. The core element consists of the four colored columns in the center and the fields to their left and right. These are pre-formatted spaces, with an equal number of rows per sector to receive descriptions of problems, their severity ratings as well as recommended interventions. The color codes have been defined, with variations, in numerous assessment methodology documents. For each sector, more correctly: for each site-sector combination, the communities (as done in Yemen) or the assessment teams (conceivably) give a combined severity rating called "Synthesis".
3. In the lower left, space is set aside to designate the three most important sectors. These are ranked choices. Sector-level recommendations and optional comments are noted here.

This third element has been handled differently in other contexts. In a recent assessment in Bangladesh, it was replaced with a humanitarian severity rating for the entire site. In Bolivia, sectors were not ranked.

Our template dispenses with this questionnaire-mimicking entry mask. It uses a standard data table format with bold-faced variable names in the top row and one row exclusively for one entire record. The severity of the problem is recorded in one variable. Nevertheless, the constitutive elements of the Yemen format are all found in this template.

The nature of entities

In a formal data management and analysis perspective, the template (like the tradition form which it emerged) is noteworthy, if not always self-explanatory, for a number of features. These the reader should keep in mind for the correct understanding of analytic choices that underpin the data management side:

- The data table combines entities from three different levels - site, sector and problem. The basic record is about the combination of a site, a sector and a specific problem. The table holds records also at the sector level (called "Synthesis") and, in columnar ordering, the site level. If priority sectors, named by key informants, are included, they are properties of the site, not of the sectors.
- Sites and sectors are drawn from complete lists. The assessed sites are from a purposive sample; the sectors were written into the questionnaire. By contrast, problems surface from narratives and observations. They are not instances from a defined list until the assessment team categorizes them. Put differently, there is "no agreed-upon standard subsector list."
- Of the three key dimensions of Phase-2 assessments - sector, region, group -, sectors have multiple records (problems and syntheses). The region is a defining element of the site. The group is its dependent property.
- Sector priorities express true, if incomplete *rankings* (there is only one first priority, one second priority, etc.). Problem and sector synthesis severity *ratings* are ordinal variables (there can be many at the same severity level). The underlying severity continuum may not be uniform; the ratings cannot be treated as true rankings.

The complex relationships between levels and entities are diagrammed in an appendix. Here, our task is to develop a data table template that lends itself to efficient analysis. This will remain a compromise among current practice, openness for future requirements and sound data management.

[Sidebar:] The severity of needs - How many levels?

In this tradition, the severity ratings have been elicited over four levels, represented by colors from green to red. The explanations of the colors have slightly varied across agency documents, but this version comes close to most.

Figure 2: Color scheme for severity ratings, four levels

Low	Relatively normal situation (or good data) or local population able to cope with crisis; no further action required
Medium low	Situation of concern, lack of data/unreliable data: further assessment and/or surveillance required
Medium high	Situation of concern, serious risk and lack of data/unreliable data: further assessment and/or surveillance required
High	Severe Situation: Immediate intervention required <u>to save lives</u>

There are two criticisms leveled at this scheme:

1. It mixes two concerns - data quality and the severity of needs.
2. The color scale may be intuitively understood by key informants, in as much as it strictly expresses the severity of needs. However, response levels such as "surveillance" and "immediate intervention" are humanitarian agency concepts foreign to most people.

ACAPS is contemplating a five-level scheme on the lines below.

Figure 3: Alternative severity rating scheme, five levels

Problem	Intervention required	Timing	Numeric code	Color code
No	No	-	1	
Yes	No	-	2	
Yes	Surveillance	Asap	3	
Yes	Response	Later	4	
Yes	Response	Asap	5	

From the data management viewpoint, the number of levels does not matter. It becomes a concern when severity ratings mix elements that would better be recorded into more than one variable.

One might argue that the color scheme expressing severity of needs should have identical meaning for both assessment teams and key informants. It should thus be unidimensional, and not speak to organizational decision premises such as whether an intervention seems required and how soon. For these latter elements - to be determined by the assessment teams, not the key informants - separate variables should be provided, next to the variable "Recommendation".

Architecture

We assume:

- Teams continue to hold site, sector and problem information in one table. They use a spreadsheet application, not a relational database.
- Even with a sampling plan, the number of sites eventually returning usable data is not fixed in advance.
- Data are entered incrementally, site by site (and not by instances of other entities, e.g. target group by target group).
- Entry may take place in several computers, in several locations and even in assessment teams that use different category sets (e.g. different drop-down lists for recommendations). In the extreme, entry operators - defying common instructions - may return tables with differences in their sets of variables.
- The first version of problem and recommended intervention statements - whether entered as free text (as in Bolivia) or in fixed categories (as in Yemen) - will likely need reformulation.

The proposed template responds to these assumptions in the following ways:

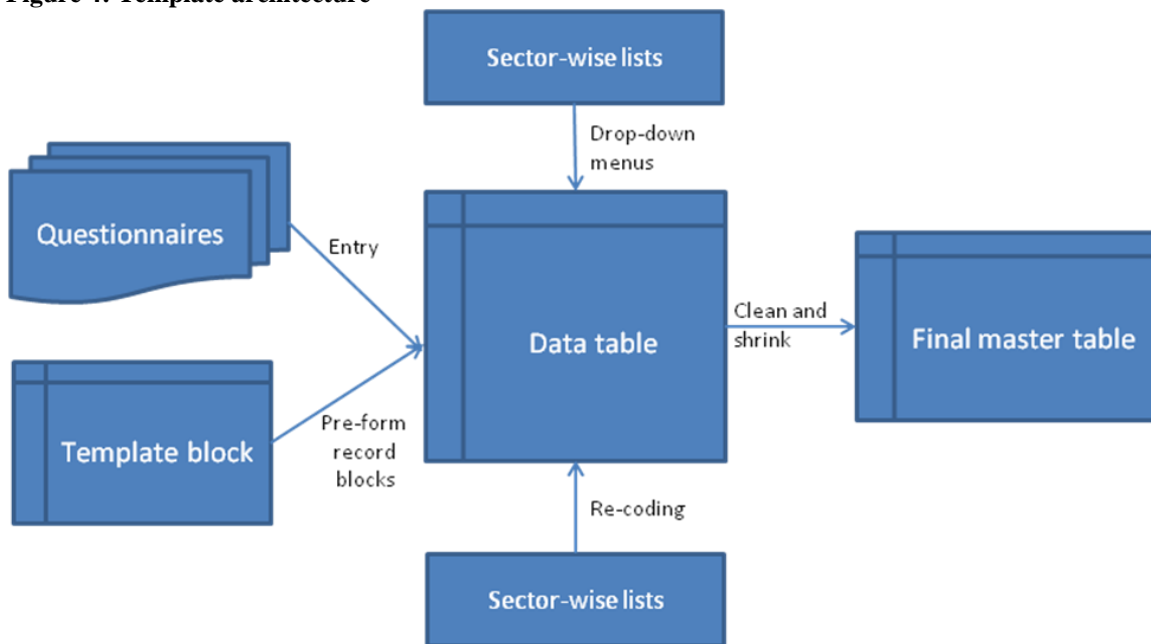
- The data table has clearly marked variable sets (groups of columns marked by color and with colored superheadings). Distinct sets are for metadata (e.g., date of the assessment) and identifiers, for site-level information including priority sectors, for sector level information, and for problems and recommendations. At the far right of the table, two columns are reserved for the "automatic" recoding of problems and recommendations.
- Not only columns, also records (rows) are pre-formatted. The pre-formatting takes place in a separate template sheet, so that revisions can be experimented with, as and when needed (chiefly during the initial contact with the new theater). Every activation of the template sheet provides sufficient records for one more site. When the questionnaire arrives at the entry computer, clicking a macro button adds a set of blank records to the bottom of the data table. Such blocks of pre-formatted spaces, with automatic record numbering, can be called as often as needed.
- At each level, data is entered only once. While the data collected at a site may easily fill 20 rows or more (one for each problem and sector synthesis), the site specific information is entered only once, in boxes in the first row of the site block. This data can then be copied to the remainder of the records of this site, to make it accessible in Pivot tables.
- So-called tagging variables mark one record per level - the site tag for the first record for the given site, the "Synthesis" tag for every sector-level record. Tags,

used as filters, allow Pivot tables to summarize variables defined for a particular level.

- Problems and recommendations can be entered as free text or as options from drop-down menus. Drop-down menus are the default. They are controlled from separate lists, one for each sector. The sector-wise lists work with dynamic named ranges, meaning that every addition at the bottom is immediately added as drop-down option. In free-text entry, the problems and recommendations need to be categorized subsequently; again, the template architecture helps.
- The sector-wise lists accommodate also the definitions for recoding problems and recommendations. Dynamic names and lookup formulas wired into the data entry template make recoding efficient and available for repeated trial-and-error at the time of analysis. In multilingual environments - "water", "eau", "agua" - or with frequent spelling variants among entry operators -, this device produces a unified version.

The diagram schematically represents the process for a one-computer entry process. If data entry takes place in several locations, the master table results from appending several contributing data tables. The point here is that this architecture connects a data table, a template for adding pre-formatted records, and tables controlling drop-down menus and recoding rules. These elements are held in separate sheets of the same workbook.

Figure 4: Template architecture



[Sidebar:] Required user skills

Using the template calls for basic MS Excel skills for data entry and minor template adaptations. Pivot tables and recoding are the essential workhorses harnessed to the analysis. Whether the former are to be considered a basic or an intermediate skill is a matter of taste; the fact is that they are fundamental to data analysis in Excel.

The latter is a generic categorical operation that does not depend on this computer application; the housewife who, planning to buy apples, oranges and bananas, writes down "Apples and some tropical fruit" has just done a recoding. However, the template invites more complex recoding schemes to be written into the sector sheets and then automatically applies them in the data table. Filling the recoding schemes requires none but the most basic Excel skills. To appreciate this, it will help to go through the companion analysis note and workbook.

Intermediate Excel skills are required to perform some of the more demanding, if rarely needed template adaptations, such as the addition of sheets to hold new sector-wise lists. They are also helpful for the curious who wish to understand the silent mechanics of the template, and for the enterprising who look to combine elements of the extant with a novel template of their own design.

Specifically, these elements may be noted by intermediate-skill users:

- The template uses R1C1, instead of A1, cell reference style. This makes navigation and modifications of named ranges easier because the columns are numbered.
- Some modifications need work with named ranges. These are best accessed through the Name Manager.
- The named ranges in the sector-wise list sheets are so-called dynamic ranges² - ranges that update their definitions when their environment changes (meaning here: when elements are added to, or deleted from, the list).
- The recoding machinery works through the combination of these dynamic ranges and the lookup functions written into the recoding space of the data table³.

Finally, we should mention the macro that adds record blocks to the data table at data entry. It can be accessed in the VBA editor (by typing Alt + F11). While the code may elude the basic and some intermediary-level users, it is still a simple macro. Foreseeably, it will never need modification. But chances are always that some users will be able to improve on it. Regardless, the basic idea is that it does its work silently in the background.

² The MS Excel help function does not offer entries on dynamic ranges, but the Web is full of its help and formula collection pages. OzGrid (Undated) is a possible starting point. The Excel Help does feature the function OFFSET, which is the one used in dynamic ranges in the template.

³ The lookup function VLOOKUP reads in the original problem or recommendation value, then looks up its recoded value in the concerned sector-wise lists, or more correctly, the dynamic ranges that hold those values in the sector sheets. VLOOKUP is an Excel feature almost as important as Pivot tables and is often the hinge between data tables for different entity levels and the reason why relational data can be managed in spreadsheets.

Usage and adaptation

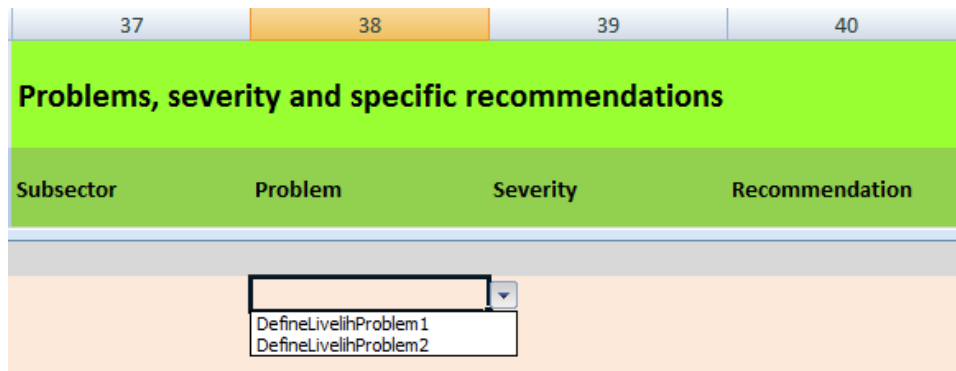
We first discuss the work process assuming that the template for blocks of records is ready. The next chapter will turn to questions of how to adapt the template.

"Once the template is ready, how do I use it?"

The template is ready in a separate sheet, which the companion demo workbook names "DataEntryBlockTemplate". The data entry takes place in the sheet called "DataEntrySheet".

1. Make sure Excel is macro-enabled. It is also recommended that you use R1C1 reference style (to be found under Excel options - Formulas).
2. Check the data entry sheet. It must be named "DataEntrySheet". In the demo workbook, row 1 and 2 are already populated with chapter headings and field names. The screen is split, with the upper left quadrant showing row 1 and 2 and column 1 - 3. If, for any reason, this initial arrangement is missing or compromised, delete "DataEntrySheet". Create a new sheet, name it again "DataEntrySheet", copy row 2 and 3 of the "DataEntryBlockTemplate" into row 1 and 2 of the new "DataEntrySheet" and split the screen.
3. In the template sheet, click the macro button once. This takes you to the entry sheet, which has now been pre-formatted to facilitate the data entry for the first site. Using the slides in the lower right quadrant, explore the sheet horizontally and vertically. The macro has created a site-level record (grey) as well as eight records for each of eight sectors.
4. Enter site level data in the red-framed cells, beginning with the *site number* or identifier in column 2 (the *site name* is entered in column 10). Since site identifier conventions may vary between assessments, the macro does not create it.
5. Enter the sector level data for this site if any variables have been defined for this level. In the demo workbook, placeholders are in column 35 and 36, with the sector-level boxes for this data in dark-blue frames. The essential sector-level variable, however, is the severity rating ("Synthesis", in column 39). Conveniently, if illogically, it has been placed together with the problem-level severity ratings in the same column (In the analysis, severity ratings are filtered, either to the sector or the problem levels).
6. Prepare for the problem-level data entry. Check if drop-down menus work, by selecting cell R4C38, which should display these options:

Figure 5: Drop-down menu for the "Problem" field



7. These reflect the default options in sheet "Livelih_Lists". Visit this and other sector-wise list sheets. In column 1 and 6, replace the defaults with meaningful options, adding as many as needed downward. Unless problem and recommendation lists have been decreed in advance (not recommended!), this will be an iterative process, with frequent switching between reading questionnaires and adding further options. Options already used in the data entry sheet should not be deleted in the lists; rather add better options at the bottom. This way, entries that you reject for more suitable alternatives can be collectively replaced in the later recoding process; they will not require individual attention in the data sheet.
8. You need not yet bother about column 2, 3, 7 and 8 in the sector-wise list sheets. However, if you know for sure that you will want to replace a problem or recommendation category with something else, you may at any time enter desired replacement terms in columns 3 or 8. You will find the replacements made automatically in column 41 or 42 in the data entry sheet.
9. If you wish to use free text rather than drop-down menus to enter problems and recommendations, remove the options from column 1 and 6 in the sector-wise list sheets. Do not delete the column and column headers, though. There is no need to remove the data validation in the template or in the data entry sheets. The drop down arrow will continue to appear in column 38 and 40, but free-text data entry will not be impeded or complicated by warnings. Recoding free-text entries calls for an additional step, which we will describe further below.
10. For each problem noted in column 38, enter the severity rating in column 39 and, if one was made, the recommendation in column 40. In the questionnaire, the severity rating is expressed as a check in one of the colored columns. Traditionally, there are four such columns. If so, enter "1" for green, "2" for yellow, "3" for orange, "4" for red. There is no need to enter "N/A" or the like in rows left unused. See the sidebar

further above regarding the number of severity levels and whether it has any importance for the data management.

11. Enter the severity rating for the site-sector combination on hand, in the "Synthesis" record, using the same coding as above. Leave it blank if no problems were recorded for this site and sector.
12. Continue to enter the data in the segment "Problems, severity and specific recommendations" for all sectors of this site. The variable "Subsector" (column 37) is a placeholder; its use depends on a subsector list on which your team agreed.
13. Ignore the segment "Recoding space" (column 41 and 42) and the "Include" tag (column 43) for the time being.
14. When you are done entering the data for this site, review your work. You will notice that in column 3, the "Keep_record" indicator has been updated. It says "Yes" in all rows in which some entry was made in the "Problem" field, in addition to the site-level record and all the "Synthesis" records. "Yes" has been turned on deliberately also for sector synthesis records in sectors without any problem recorded. On this, more later.
15. If you are satisfied with the entries, copy the site number (column 2) down to all records for this site. Similarly, copy all the other site-level data, from column 6 to column 32, down to all the records for this site. Since these columns are all contiguous, the copy-paste operation can be done for all of them together⁴. The purpose is to make the site-dependent information available for any sector or problem-level tabulations later on. For example, break-downs by target group depend on this fill-up.
16. Save your work.
17. Enter data for more sites.

⁴ There are several ways of doing this - the ordinary copying followed by pasting into the selection; selecting the site-level cells to be copied downwards and pulling the fill-handle on the lower right of the rightmost cell. For esthetic reasons, one may want to remove, after pasting, the repeated red cell frames. Double-clicking the handle sends copies down automatically, as far as an adjacent column to the right or left is filled. However, this is error-prone in case you inadvertently already pre-formatted records for any following sites. Manual control while pulling the handle and watching the record numbers as you go down is safer.

[Sidebar:] Tags

Tagging variables are devices that "tag" or - some would prefer - "flag" records that meet particular conditions. In statistical applications that combine in one table variables with properties from different entity levels, tags are used to create subsets of records for procedures that use data only at a particular level. For example, a table may list all members of all households surveyed. If this table also holds data on the annual household incomes (a household-level variable), we need to tag one record per household in order to compute valid income statistics.

The situation for this template is in part similar. The data table combines information on three entities, as we have seen. There is a need, therefore, to have tags that restrict tabulations during analysis to the appropriate entities.

We make a distinction between formal and substantive tags.

Formal tags respond to the structure of the data table. The template comes equipped with three such tagging variables:

- "Keep_record" (column 3) tags (with "Yes") records that are either site level, synthesis level, or filled-in problem-level records. Blank problem-level records receive a "No".
- "Site_level_record" tags (column 4) one record within the 65-record block for a site - the first record, the one in which the site-level information is entered.
- "Synthesis_record" (column 5) does what its name says: it tags synthesis records. A convenience copy is placed in column 34, for optical reasons when the user scrolls to the sector and problem regions.

The formal tags are used to filter the Pivot tables to the records appropriate for the purpose. Once the table can be sorted (i.e. after the record numbers have been made static), the tags enable particular sorts (e.g. to see all the site information, one record for every site, together) and let uses delete blank records.

Substantive tags set records apart for which the analyst has a substantively motivated preference, but which cannot be identified by the usual filtering techniques. For example, in the companion analysis note and workbook, such a substantive tag is created for problem and synthesis-level records whose sector the affected community elected as a priority sector. This information is scattered over three variables (there are three priority sectors); if any of these equal the sector of the record, the tag kicks in. This cannot be handled directly in the Pivot table filter. A formula combining IF and OR-functions computes the tag; it is then this tag that is used in the Pivot table filter.

Another situation potentially calling for a substantive tag arises from the occasional practice to record more than one assessment for a given site population, notably one from interviews with male key informants, and a second with female informants. This may not be done consistently in all sites; thus simple gender-segregated statistics will not work. The analyst may need to express a preference for each site as to which assessment to include, such as on the basis of presumed better reliability. The tag may need to be set manually, or in a small summary table in a separate sheet from which tags are computed via a lookup formula.

At present, the template holds only a placeholder for a substantive tag, in column 43, called "Include". At this point, it is populated uniformly with the value 1 (instead of "Yes"). When used, records to be excluded would be tagged 0 (instead of "No"). The numeric format allows the

tagging variable to be counted in a Pivot table, a possibility that comes in handy when the frequency of cases by some other variables is sought. It is obvious that the meaning and formula for the substantive tag has to be documented.

Tagging is common in statistical work. There is, as mentioned, some confusion in terminology, with the same basic device being called tag, flag or even indicator. Regardless, tags facilitate speedy and valid analysis. Therefore, the template provides a minimum of tagging variables.

"How do I combine data from several computers or from several locations?"

We assume:

- You are the person authorized to combine the contributions in a master table.
- You are authorized to recode problems and recommendations or to apply a recoding scheme that the assessment team approved.
- The site identifiers (column 2) used in different contributing tables do not overlap.
- The number of tables to combine is small; doing it manually takes little time and is desirable because the tables can be visually inspected beforehand.

The general idea is to use a copy of the data table in one of the workbooks as the starting point for the master table. Data from additional workbooks are sequentially added. The operation is known by different terms, such as append or union (but not: join or consolidate!). We prefer to think of copy (from subsequent tables) - paste (into the master table) operations. Visual inspection is highly recommended before any actual copying.

1. Make working copies of each contributing workbook, ideally with a *year.month.day_time_contributor* prefix to identify different versions of the same table in case updates or corrections arrive.
2. Save the first workbook with a name that suggests a combined database. Keep the screen split in the data table ("DataEntrySheet" unless it was renamed).
3. For all subsequent workbooks, take these steps:
 - 3.1. Open the next workbook from which data will be appended to the combined workbook. Copy the data table, including the column header row.
 - 3.2. Paste it into the master table sheet, selecting the cell in column 1 that is *about 2-3 rows down* from the bottom of the master table. This helps visual orientation while scrolling through the sheet.

- 3.3. Scroll horizontally across all columns and check if the variable names and order are the same in both tables. They should be⁵.
 - 3.4. Pull the horizontal split line down a few rows so that you can see the data of the master part in the upper quadrants. Scroll down vertically and check whether the site identifiers in column 2 are truly unique at the site level across both tables. They should be if operator or regional prefixes were used, such as "West-01", "West-02", etc. and "East-01", "East-02", etc. Else, create distinguishing prefixes⁶ or change identifying numbers in one of table being appended, such that none overlaps with those in the master part. The record numbers (column 1) will overlap until the append operation is completed; they will update automatically in the next step.
 - 3.5. Once these conditions - identical variable order, unique site identifiers - are satisfied, delete the blank rows between the tables and delete the header row of the appended table. Its record numbers should now change to being consecutive for all records in the combined table.
 - 3.6. Before we are done with this contributing table, we check the sector-wise list sheets in the originating workbook. If we find options in problem and recommendation categories different from the ones used in the master workbook, we append (copy-paste) them to the corresponding master workbook definitions. We do this also if the differences concern only spelling variants. Both substantive and spelling differences may eventually need to be dealt with in the recoding schemes. The appended elements may be highlighted and, if any specific rationale is known, annotated, pending resolution during the analysis phase. There must be no blank rows left between the master categories and the freshly appended ones.
4. Append all remaining contributing data tables likewise.
 5. Until now, the record numbers in column 1 have been formula-based. We make them static so that they are preserved when the master table is sorted in different order. Copy row 1 and paste it (values only) in place.

At this point, it is a matter of discretion to delete the blank records (i.e., rows pre-formatted to receive problem descriptions, but not used) or to keep them. If more data or substantial corrections of existing data are expected, we should keep them. If the master

⁵ If not, postpone working with the workbook that deviates from the original common template until you have inspected the rest of the contributing workbooks. Decentralized data entry, experience shows, sometimes prompts operators to adapt templates to what they consider legitimate special considerations. Ultimately, you will need to establish - through insertion, deletion and reordering - a common order.

⁶ Efficiently done by building a CONCATENATE function in a temporary column to the far right, then copy - paste (values only) its results back to the site number column.

workbook is final, we should delete them. As a general precaution, we should delete them in a workbook saved under a different file name (with timestamp). After sorting on "Keep_record", the "No" rows are deleted, and then the "Keep_record" column may be deleted as well. Note that we keep "Synthesis" records even if the community did not report any problem for this sector. These "missing value" synthesis records can have a function as part of denominators in the later analysis.

6. Delete or keep the blank records as discussed.
7. Select the entire data range including the field header row (but not the superheading row above it) and name it "Database" (e.g., by typing Database in the address box to the left of the formula bar, or by going to Formulas - Defined Names - Name Manager).
8. With "Database" selected, go to Formulas - Defined Names - Create from Selection. Only "Top Row" must be checked. Hitting OK, we have created ranges of all column vectors in the data table conveniently named after the field headers. These ranges are now available for the analysis, such as for quick counts, in the sector-wise lists, of how often categories were used. Here is an example of a livelihoods problem recoding scheme from the Yemen analysis. Note the COUNTIF formula using the column range "Problem". It computes the number of records in which the original problem formulation was used in the data table. These frequencies are key when we decide how to combine categories in the recoding process⁷.

Figure 6: Example of a recoding scheme

	1	2	3
1	Problems	Frequency	Problems_Recoded
2	Cash for basic services	29	Financial
3	Income/employment	45	Financial
4	Lack of skills	19	Human resources
5	Loss of capital/property	14	Financial
6	Limited economic resources	21	Financial

The problem categories are from the Yemen assessment. The recoded values are arbitrarily set for this note.

⁷ In every sector sheet of the template workbook, a formula for the frequency of the original problem formulations has already been written into cell R2C2, and one for the frequency of original recommendations in R2C7. These can be conveniently copied downward once the actually used formulations have been filled in. The error message "#NAME?" will disappear and the frequency will appear when the columnar named ranges, including "Problem", have been created, as described in step 8.

9. Save the master workbook.
10. When you share the data with outsiders not familiar with the template, its built-in features may not be of interest or may even prompt harmful manipulations. In particular, the shared data table should no longer contain any formulas. You may also wish to strip the shared workbook of all sheets other than the data sheet (with the blank records already deleted) and the sheet "Variables". Before doing so, make sure that the formulas in the data sheet have all been replaced with their results. To do this, save the workbook under a different name and copy the entire data table and paste it (values only) in place. This assumes that you are already finished with the recoding work, or are continuing it a workbook version that preserves the formulas. The workbook that you share should be clean; therefore, go into the name manager and delete all names that no longer have a reference.

"How do I recode problems and recommendations?"

The term "recoding" combines a number of different operations in the gray area between data management and analysis. Common to them is the replacement of some or all of the values of a variable by other values ("values" standing here for elements of text as well as of numeric ranges). In simple situations, this is done with the replacement command, in the original variable or in a renamed copy. In situations that need syntactic (deciding spelling variants) or analytic experimentation (reducing category sets by grouping), recoding schemes should be defined and documented outside the data table. This is likely the situation in assessments that need to streamline problem categories for meaningful reporting.

The purposes of recoding are multiple:

- eliminating spelling differences
- adding a prefix to the category names so as to force a particular sort order
- categorizing free-text response
- renaming and/or grouping categories
- translating categories in a different language
- (in conjunction with another variable) renaming and splitting categories

The mechanism in Excel relies on the function VLOOKUP in the recoded variables in the data table and on look-up ranges in the sector lists. Both are part of the template; the user does not have to create formulas or name ranges unless new sectors are defined. This possibility is discussed in the section on adapting the template.

1. If problems and/or recommendations are in free-text form (the Bolivia case), then an additional operation is required. Make a Pivot table with the frequencies of all problem (or recommendation) variants in a separate sheet, filtered to one sector at a time. Copy - paste (values only) the results (text strings and frequencies) into the concerned sector-wise list space (never mind that this will overwrite the COUNTIF

formulas in cells R2C2 and R2C7 - you already have the frequencies!). Do this for all sectors.

[Sidebar:] Free-text entries - an illustration

Free-text entries, if they are carefully done, preserve much of the original key informant or assessment team information and intent. They are particularly valuable in more qualitative processing of the problems and recommendations.

Free-text entries were used in Bolivia. Key informants listed 85 problems in the WatSan sector. No fewer than 68 different formulations were used. Most were not only distinct, but also unique. Here is a segment, showing the first nine members of the alphabetically ordered list:

Table 1: Free-text entries of problems, Bolivia (segment)

Problems	Frequency
Agua no potable	1
Agua turbia	1
Alta tasa de enfermedades en niños	1
Baños tapados	1
Capacitación IEC en agua y saneamiento.	1
Conocimientos débiles de infección	1
Consumo agua cruda, pozos semisurgentes	1
Contaminación del agua en los pozos	1
Diarreas	1

Such numerous different formulations await a steep compression into an analytically feasible category set. This too is best done in the sector-wise list sheet. The recoding in the data table then is automatic.

-
2. If, by contrast, problems and recommendations were entered using drop-down menus, then sets of categories already exist in the sector-wise list sheets. Determine what syntactic (pre-fixes, spellings, leading and trailing spaces), semantic (improved formulations) and analytic (grouping of categories, substitutes within particular sectors) modifications are appropriate. Copy the existing p (resp. Recommendations) lists wholesale to the _Recoded columns (i.e., from column 1 to 3, resp. from 6 to 8). Make the changes, if any, in here, in column 3, resp. 8.
 3. Go back to the recoding space in the data sheet (column 41 and 42 in the template). Check whether any cells still display errors. If so, inspect the original problem or recommendation. The error occurs either because a drop-down option was overwritten with a value that is not an element of the sector-wise list, or because this element has no corresponding entry in the recoding scheme. If the original value is ok,

add it to the sector list, including in the recoding scheme. Else, replace the original with a legal option. In both events, the error should disappear.

4. In subsequent analyses, use the recoded problem and recommendation variables in the Pivot tables. If the results suggest that the recoded category sets are not appropriate, modify the recoding schemes in the sector-wise list sheets. Refresh the Pivot tables and re-evaluate.

"How do I adapt the template to the requirements of this new assessment?"

We assume:

- Every needs assessment requires a format that is different at least in some minor degree from previous formats.
- Nevertheless, some core definitions and data structures remain common.
- Most information can be arranged hierarchically, in terms of affected communities met at particular sites, sectors of concern to them, and problems and needs meaningfully assigned to sectors. If so, the basic template structure remains.
- The need for adaptation arises in the selection and naming of variables, category sets for variables, and (rarely) the selection and naming of sectors under which problem and needs are recorded.
- In special situations, other components of the template may need modifications, such as in multi-lingual settings that call for translation as part of recoding.

The template elements possibly affected are:

- The data entry block template and the range name that tells the macro what blocks of pre-formatted records to add to the data entry sheet (but not the macro itself)
- The data entry sheet with its chapter and field name rows
- The sector-wise list sheets controlling drop-down menus and recoding schemes.

We discuss changes in each of these:

Changing variables

Variables are changed in the sheet "DataEntryBlockTemplate". They can be renamed and reordered. We recommend leaving the identifying and tagging variables in column 1 - 5 as they are. Variables can be deleted, and new ones can be inserted (preferably after column 5). New variables can also be created contiguously on the right side of the existing table.

1. Make the requisite changes by renaming, reordering, inserting, deleting or right-side adding variables.

2. Reflect these changes accordingly in the sheet "Variables", which documents the names, order, meaning and special characteristics of the variables.
3. The macro copies blocks of records for the existing variables, referenced in a range named "BasicBlock". This range definition is unaffected by, or adjusts itself automatically to, most changes in the composition of variables. Only if variables are added to the left (not recommended) or to the right (ok) of the existing block, will it need adjusting. Open the NameManager and change the reference (currently =DataEntryBlockTemplate!R4C1:R68C43) as appropriate.
4. Delete the current contents of the "DataEntrySheet". Copy row 2 - 3 in the template sheet and paste them into row 1 - 2 of the data sheet.
5. Click the macro button. Check the integrity of the block pasted into the data sheet, and its alignment with the field headers in row 2. In the Recoding space (column 41 and 42), all cells in problem-level rows should display the error "#N/A".

Changing category sets

The sector-wise list sheets in the template workbook do not hold substantively defined categories, just placeholders. Replacing these with, and adding further, meaningful categories, therefore does not structurally change the template.

Remember, however, to list categories without gaps in the vertical, always starting in row 2. The underlying dynamic ranges that pass the categories to the drop-down menus in the data sheet expand vertically until they meet the first blank cell; they do not capture cells separated by gaps.

Changing sector-specific category sheets

Currently, the template includes category sheets for eight sectors:

1. Livelihoods
2. Water, sanitation and hygiene (WASH)
3. Shelter and non-food items
4. Food security
5. Health
6. Education
7. IDP (or refugee) return
8. Protection

We distinguish the case where the adapted template lists eight or fewer sectors from adding an ninth sector or more. The steps needed in the second case are only slightly more demanding.

A. Replacing an existing sector with a new sector

Rededicating an existing sector sheet to a new sector name is essentially just the task of renaming the concerned parameters in the renamed category sheet and in the sector block of the "DataEntryBlockTemplate". This means:

Every sector-wise list sheet has four dynamic named ranges. For example, "Education":

Table 2: Examples of dynamic named ranges

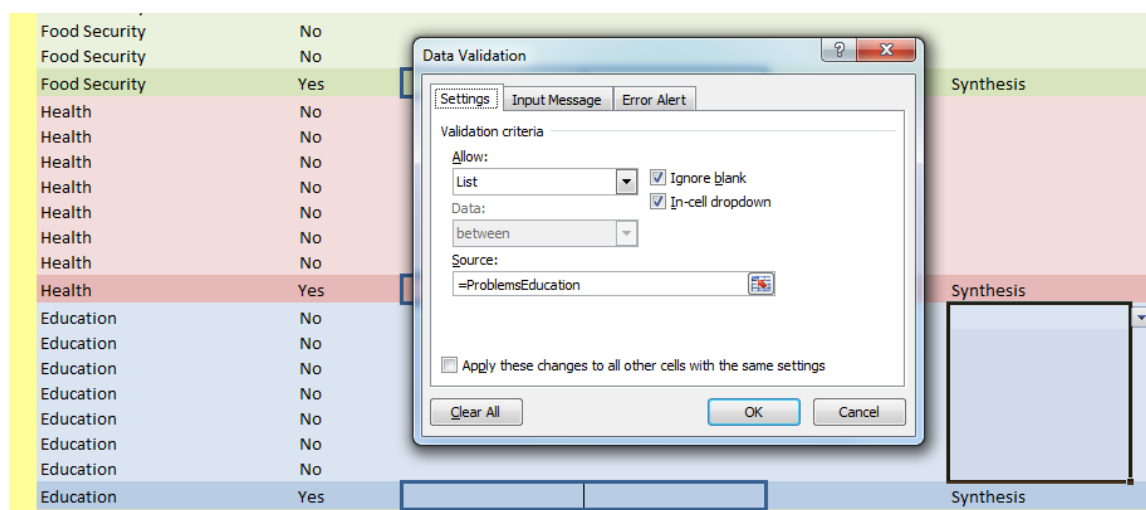
Name	Definition
ProblemsEducation	=OFFSET(Education_Lists!R2C1,0,0,COUNTA(Education_Lists!C1) - 1,1)
RecodeEducation	=OFFSET(Education_Lists!R2C1,0,0,COUNTA(Education_Lists!C1) - 1,3)
RecommEducation	=OFFSET(Education_Lists!R2C6,0,0,COUNTA(Education_Lists!C6) - 1,1)
RecRecodEducation	=OFFSET(Education_Lists!R2C6,0,0,COUNTA(Education_Lists!C6) - 1,3)

The first and third names in this list are used in the data validation rules for the problem and recommendation entries. The second and fourth appear in the lookup formulas of the Recoding space.

Thus, if you wanted to replace "Education" with some other sector, you would need to make several changes:

1. In the "Sector" column (column 33 of the sheet "DataEntryBlockTemplate"), "Education" is replaced with the name of the new sector.
2. The name of the category sheet changes from "Education_Lists" to "NEWSECTOR_Lists", where NEWSECTOR is the name of the new sector.
3. In the name manager, the four names and their definitions are changed accordingly.
4. In column 38 ("Problem") and 40 ("Recommendation") of the sheet "DataEntryBlockTemplate", the data validation rules are adapted. The figure below suggests where and how this is to happen for column 38. In the "Source" field of the wizard, the name "ProblemsEducation" is replaced with "ProblemsNEWSECTOR". Similarly for column 40, replacing "RecommEducation" with "RecommNEWSECTOR".

Figure 7: Example of data validation settings for "Problem" drop-down menus



5. In column 40 and 41, the recoding formulas

*=VLOOKUP(RC38,RecodeEducation,3,FALSE), and
=VLOOKUP(RC40,RecRecodEducation,3,FALSE).*

are adjusted to the new names.

B. Adding new sectors

The case of working with more than eight sectors will be rare. It largely follows the same logic:

1. At first, however, it will be necessary to add rows to the template block for problems and syntheses, as appropriate. Name the sectors in column 33. Add formulas for the record numbers in column 1 and add values as appropriate in the tagging variables in column 3, 4, 5 and 34. If the rows are added at the bottom, the definition of the "BasicBlock" range will have to be adapted.
2. Add new sector-specific list sheets, as appropriate, and format them in the same way as other list sheets. Add placeholders, as in the existing ones.
3. Create the dynamic names in the Name Manager, on the pattern exemplified above.
4. Create the drop-down menus in the new rows in column 38 and 40. Mark the synthesis records with "Synthesis" in column 38.

5. Insert the lookup formulas in the Recoding space. Check if they as well as the drop-down menus work as expected.

C. Wholesale recreation of dynamic ranges

This option is for users familiar with VBA for Excel macro code.

If multiple sectors are to be modified against the current template, or if you wish to change the architecture of the recoding schemes, it may be appropriate to re-create an entirely new, consistently formatted set of dynamic ranges. To create these efficiently, the template workbook carries a convenience sheet named "AddNamesMacroCode". The macro commands for creating new names, such as

```
ActiveWorkbook.Names.Add Name:="ProblemsLiveli",  
RefersToR1C1:="=OFFSET(Liveli_Lists!R2C1,0,0,  
COUNTA(Liveli_Lists!C1) - 1,1)"
```

are built by concatenating fixed as well as adaptable elements that can be copy-pasted in blocks. Static copies of all the name commands can simply be inserted between

```
Sub CreatedynamicNames()  
'Insert command block here:  
  
End Sub
```

in the VBA editor. Run the macro once.

The situations A. - C. will be rare. The point here is to demonstrate the cooperation between the data sheet template and the sector lists, using dynamical named ranges in the latter.

Outlook

"Data management" means very different things to different people and professions. The Wikipedia article (2011a) lists ten major topics with over forty sub-topics. A minority only signal concerns that are directly relevant to needs assessments. Here, the purview of data management is more modest. In this field, data management is basically a conglomerate of simple norms and tools. We share computer application pieces (such as this Excel-based workbook template), user instructions (such as part of this note), and implicit or explicit standards or maxims (such as our emphasis on recoding).

If there is any dominant guiding consideration in its design, it may be said that this template tries to help busy assessment teams to handle data in ways that

1. facilitate adaptation to new disaster contexts,
2. speed up data entry, and
3. produce a master table ready for analysis.

We have already pointed out that the template is a compromise. The data structure militates for a relational database. The known skills in assessment personnel and the time pressure on all of their work, including on adapting and learning the template, impose a spreadsheet solution. This formal tension as well as the substantive challenges of measuring problems (or "key issues", as some documents call them) and needs will not go away. They let us expect periodic dissatisfaction and instability in the development, use and evaluation of any such template, this one included.

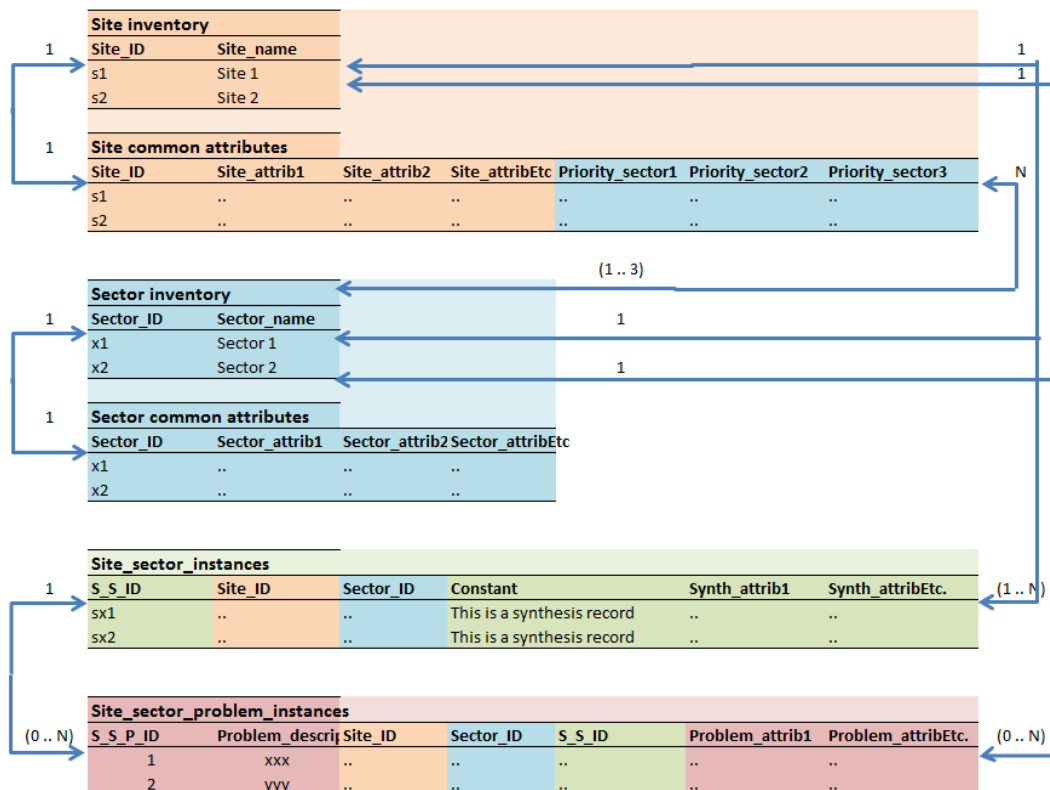
The adaptability of this template may be its best advantage. Whether the color scheme from green to red for severity survives the next assessments or not, whether HESPER Scale items are included or not, whether "problems" have to be recorded in finer granularity over more fields or not - a data manager worth his salt will be able to incorporate the requirements within two hours of concentrated work. These two hours are a literal bet, but they do not take care of the heavier lifting: having colleagues pre-test the template; setting-up a workflow for entry and inspection; recognizing defects and correcting them while assessment teams are already turning in questionnaires.

Appendix

Entity relationships

The structure of the data accommodated in one table - the data table - of the template workbook is complex. In orthodox data management, it would be handled through several linkable data tables. This diagram visualizes the generic relationships among entities⁸.

Figure 8: Entity relationship diagram



The current template does not hold space for any sector-level attributes (the blue table below "Sector inventory"). There is no need for them. This table is added here only for logical completeness. The placeholder variables under "Sector", in columns 35 and 36 of the data table template, are, by strict logic, for site-sector combinations.

⁸ The diagram uses an abbreviated version of the Chen entity diagramming. In the usual didactic example to make this clear: "A person is born in one location; in a location, zero, one or many persons are born", the arrow between person and location is marked "N" on the person side, and "1" on the location side (most of us would do it the other way round!). The Wikipedia (2011b) article is helpful.

Also, there is no "problem" or "key issue" inventory. Problem information is captured either in free-text entries or through the team's process of categorization. This peculiar situation is apparent in the fact that problems are elicited sector by sector, but there is no commonly agreed-upon classification of subsectors. Problems are neither finite, nor of uniform granularity. Both "Water is scarce" and "The water supplied by tanker is not clean" may be recorded for the same site. The ontology of problems challenges the analysis. There is no natural denominator for the severity-of-needs ratings associated with the recorded problems.

Target groups are handled as one of the site attributes. Sites that host more than one target group necessitate separate assessments. This requirement is grounded not so much in the data structure than in the interviewing process.

References

- Global Education Cluster (2010). The Joint Education Needs Assessment Toolkit. Geneva, Education Cluster Unit, Save the Children.
- IASC (2009). INITIAL RAPID ASSESSMENT (IRA): FIELD ASSESSMENT FORM. Geneva, United Nations Inter-Agency Standing Committee.
- NATF (2010). Multi-cluster/Sector Initial and Rapid Assessments. Technical Guidance. Annex to the Operational Guidance for Coordinated Assessments in Humanitarian Crises [Working Draft 03 March 2010]. Geneva, UNOCHA and IASC Needs Assessment Task Force.
- NATF (2011). The Multi-Cluster/Sector Initial Rapid Assessment (MIRA) Approach. Process, Methodologies and Tools. Provisional version as of 2 September 2011. Geneva, IASC Needs Assessment Task Force.
- OzGrid. (Undated). "Dynamic Named Ranges." ozgrid.com. Retrieved 4 July 2010, from <http://www.ozgrid.com/Excel/DynamicRanges.htm>.
- WHO and King's College London (2011). The Humanitarian Emergency Settings Perceived Needs Scale (HESPER): Manual with Scale. Geneva, World Health Organization.
- Wikipedia. (2011a). "Data management." Retrieved 21 October 2011, from http://en.wikipedia.org/wiki/Data_management.
- Wikipedia. (2011b). "Entity-relationship model." Retrieved 6 October 2011, from http://en.wikipedia.org/wiki/Entity-relationship_diagram.

AB/21 October 2011/Revised 5 and 24 January 2012

The Excel workbook template is currently called:
120105_Assessment_DataManagement_Template.xlsm