August 2013

Technical Brief

Compared to what?

Analytical thinking and needs assessment

1. Contents

Acknowledgements

This document would not have been possible without the assistance, support and guidance of many people. ACAPS especially wishes to express its gratitude to Susan Erb, James Sparkes and Richard Garfield for their thorough revision and critical insight.

2. Introduction

Analysts of humanitarian needs find themselves faced with the daunting task of analysing data, but lack clear approaches to that task. Needs assessment literature historically falls silent on providing useful guidance on *how to do analysis,* and instead, jumping over analysis techniques straight to discussions on how to best document and communicate findings. This technical brief seeks to redress that imbalance by breaking down the analysis process into simple steps to show that analysis consists of a fairly limited set of basic moves.

More than just a set of skills, analysis is a frame of mind, an attitude toward experience. It is a form of detective work that typically pursues something puzzling, something you seek to understand rather than something you are already sure you have the answers to. Analysis finds questions where there seemed not to be any and makes connections that might not have been evident at first. It breaks things down to search for meaningful patterns, or to uncover what we had not seen at first glance, and to understand more closely how and why the separate parts work as they do.

Understanding simple techniques used in analysis can remove some of the uncertainty and provide a clear way into the work. Each of the proposed steps outlined in this technical brief serves the primary purpose of analysis: to figure out what something means: Why it is as it is and why it does what it does.

This technical brief is based on three years of experience in analysing needs in emergency settings. It is the first of three ACAPS documents on the analysis of humanitarian needs. Readers are advised to complement this reading with the technical brief *How Sure Are You?* which explores how to judge the quality and strength of evidence in data analysis and the technical brief *What Is the Most…?* which explores needs prioritization.

3. What is data analysis?

Data analysis is the process of bringing order, structure and meaning to the mass of collected data. It does not proceed in a linear fashion; rather it searches for general statements about relationships between categories of data [1], moving from a description of *what is the case* to an explanation of why *what is the case* is the case[2].

a. Analysis and interpretation

Analysis is generally defined as *what we do to make sense of information*, an activity which always demands human input. It involves two very different sets of complementary activities: understanding what the data says (analysis), and determining what the data means (interpretation)

While data collection is the systematic compilation of information, data analysis involves the uncovering and identification of patterns, trends and relationships in data. It can be as qualitative as recounting or recoding a narrative story and as quantitative as working out averages from sets of numeric values.

Data interpretation involves going beyond the data and giving meaning to patterns and trends through contextualisation, use of experience, and selection of most important findings to enable decision making.

Activities related to the evaluation of the precision and accuracy of the data at hand are not detailed in this document, but are a crucial step of the analysis process. Data quality, completeness, and usability will considerably influence the scope of analysis and the extent to which conclusions can be drawn from the data. This topic is covered in detail in the ACAPS Technical brief *How Sure Are You?*

[1] Marshall and Rossman, 1990:111.

[2] Hitchcock and Hughes 1995:295.

b. Deductive and inductive reasoning

An effective approach to analysis can be drawn from the practice of scientific and logical reasoning, specifically *induction* and *deduction.*

Deductive reasoning involves confirming or finding evidence to support specific ideas. It is a targeted and narrow approach concerned with validating a theory through the testing of a hypothesis. A deductive approach will involve a predetermined sense of what stories might be interesting, relevant, and potentially available within the data. Deductive reasoning involves pursuing curiosity by interrogating available data (e.g. extracts from focus group discussion or a dataset from field assessment) to substantiate or refute the analyst's hypothesis.

Inductive reasoning is more open-ended and exploratory and works the opposite way. When the analyst is not sure precisely what the interesting stories might be, analytic techniques are used to try and unearth potentially interesting discoveries, forming different and evolving combinations of data questions. Fundamentally, inductive reasoning is about using analysis to determine the relationships that exist within raw information materials and to recognize the most important and relevant associations.

Both inductive and deductive reasoning call for a number of interrelated processes that summarize, arrange, and transform data into information. Technically, the analysis, irrespective of whether the data is qualitative, quantitative, or a combination of both, seeks to describe the data, identify relationships between data points, compare variables, and forecast outcomes. This entails defining significant parts and how they are related, both to each other and to the subject as a whole.

Analysis will often involves activities such as defining, categorizing, inspecting, editing, evaluating, interpreting, illustrating, explaining, clarifying and modelling data to highlight useful information, suggest conclusions, and support operational and strategic decision making.

Data analysis tends to involve following up several ideas in parallel rather than trying to find a single optimum solution. By deploying a disciplined and sensible balance between deductive and inductive enquiry, analysts can efficiently and effectively identify relationships and navigate towards the source of the most important stories and findings.

c. Analysis and needs assessment

In a needs assessment context, data is pulled from multiple sources (both secondary and primary) and in multiple formats (text, photo, numbers, observations, interviews, group discussions or media etc.). Primary data is gathered from interviews and observations[3] at the field level. Pre- and in-crisis secondary data are compiled through secondary data review. The purpose of the analysis and interpretation steps is to transform the data collected (observations and narratives) into credible information about the humanitarian needs faced by the disaster affected population.

Analysis of humanitarian needs assessment data generally include:

- Comparing the severity of the conditions between various affected groups and locations
- Explaining association and underlying factors
- Predicting/forecasting the evolution of the impact of the disaster
- Prioritizing most important issues and target groups
- Supporting the definition and selection of appropriate and proportionate response modalities.

The challenge to analysis lies in making logical sense of all (or most of) the data collected, reducing the volume of information to identify

[3] Observations refer to a set of measurements or an individual item (one person, one site, one sector, or a combination of those). Such data points or observations can be related to either a single member of a population or to a summary statistic calculated for a given sub-population.

significant patterns, and constructing a framework for communicating the essence of what the data reveals as well as the confidence you have in the conclusions drawn from the analytic process.

The analysis pyramid

4. Key Facts

Data literacy is required for efficient data analysis. It includes basic mathematical literacy and an understanding of how to work with qualitative and quantitative data, how they were produced, how to connect various data sets, and how to interpret them.

Analysis is better done in a group setting, including people familiar with the context, who have expertise in multiple sectors, and who are familiar with emergency programming.

Analysis requires careful planning upfront as the needed data becomes available. Analysis needs to be linked to clear and agreed research questions. Each needs assessment should start by asking, as a minimum, the following:

- What are the questions that need to be asked?
- What are the answers that help us move from data to decision making?
- How can we shift insight into action?
- How do we tie information to the program cycle?
- Who needs what information and at what time?
- How often should this information be updated, delivered, and shared?

There is no straightforward model, standard algorithm, or generic framework to support humanitarian needs analysis. The data available will determine what can and cannot be asserted.

All analysis has weaknesses, and it is more honest and responsible to acknowledge them.

Data analysis and collection is an iterative process moving back and forth. New data are compared and contrasted to old so as to note patterns, etc. This iterative process continues until the researcher is able to make *assertions* which describe the reality. Theses emerge as data is collected and should be tested, refined, and retested against new information until explanations are repetitive.

Analysing and understanding data is by definition a time intensive process, requiring immersion and flexibility. Fitting this with a short lived assessment process is challenging and requires attention and planning. *Ensure there is enough time* and resources to turn data into information. Start analysis as soon as data becomes available.

Know when to stop. The process of analysis can potentially go on endlessly, with seemingly infinite combinations of variables to explore and compare.

Some researchers believe numbers to be more accurate than words, but it is the quality and rigor of the analysis process and not the type of data, that matters. Quantitative analysis is not more accurate than other types of analysis, and qualitative analysis is not easier than quantitative analysis. Data analysis is not about numbers — it uses them. Data analysis is about the world, asking, always asking, *How does it work? So what?*

Data rarely speak for themselves and must be interpreted and contextualized to acquire meaning. *Data analysis is done by humans, not computers.* At best, computers can assist the process, but never replace it.

A visual display of data is essential for both data exploration and communication of findings. Visualization highlights trends and patterns in data that might not otherwise be apparent. Designing, understanding and interpreting graphs and other visual forms of data is a critical skill for data analysts.

Garbage in, garbage out: the quality of your analysis (output) ultimately depends on the quality of your data (input). Learning about the numbers and metrics used and checking for completeness and quality of your raw material is required to perform efficient analysis and will help communicate about the confidence and the uncertainty of the results.

5. Analytical steps

Data analysis generally revolves around four key steps, which are detailed below:

1. Define purpose of the analysis (research questions)
2. Organize and summarize data to allow for discovery of meaningful facts
3. Identify and describe patterns and relationships within the data through exploration and application of scientific reasoning and argument to that data
4. Determine what those patterns and relationships within the data mean. Make conclusions about that data, including understanding what caused it to occur, and identify the next steps (*so what? therefore…*).

Step 1 - Define purpose

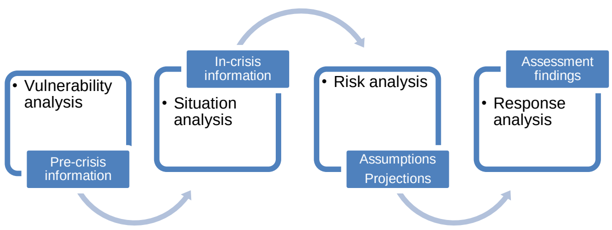
The first step in data analysis is to identify questions to be answered through analysis of the data. Clearly identifying the decisions or key documents to be informed (Flash appeals, Consolidated Appeal Processes, Cluster strategy, etc.) will help determine what type of information needs to be collected, what the most appropriate and relevant sources for that information are, and what needs to be learned through the data analysis.

Understanding your audience and the decision making process is also key to define the level of analysis required. *What decisions do they make? What questions do they need answered? How much information do they need to choose between response options? When do they need this information?*

Preparing an analysis plan allows analysts to answer the following questions:

- What are the research questions?
- What is the overall purpose of the analysis?
- How does each data collected (or to be collected) connect to the research question?
- What is the timeframe for delivering the results? How will the analysis be conducted? What resources and expertise are needed for conducting the analysis?

Rapid needs assessments typically aim to answer questions about population conditions and status through the following analytical flow:



Different types of information call for different types of analysis. Ultimately, each analysis blocks will build on the analysis performed in the previous step and aggregate knowledge incrementally until problems, issues, and risks are clearly identified and ranked. This allows for effective planning of interventions in the response analysis stage.

Most common questions attached to each step are:

Vulnerability Analysis

- Prior to the crisis, who and where were the most vulnerable segments of the population?
- Which pre-existing vulnerabilities might have been exacerbated by the current situation?
- What seasonal patterns are likely to influence the way the crisis will unfold in the future (i.e. rainy season, winter, hunger gap season, religious holidays, etc.)?

Situation Analysis

- What happened? Describe the type of emergency or disaster.
- Where did it happen? Identify the geographic areas affected by the disaster and their environmental conditions.
- Who/what was affected, and to what extent? Identify the people and resources affected by the emergency or disaster. What is the scope and scale of the impact?
- How severely have people, resources, and assets been affected? How have different groups have been affected?
- What resources and capacities already exist? Which government capacities are functioning, and which organisations are present and operational? What resources are available in terms of functioning offices, vehicles, pre-positioned goods and materials, etc.?
- What are the humanitarian access conditions and operational constraints? What is the security context and level of access to affected areas?
- What are the humanitarian needs and gaps? What are the priorities for humanitarian assistance?

Risk Analysis

- How might the (impact of the) disaster develop? Highlight special concerns about existing or emerging risks/threats and how they might unfold in the future and positively (opportunities) or negatively (risks) impact the current situation.
- What are the most important issues and risks? Which affected groups require the most immediate assistance?
- Where are the most affected geographic areas?

Response Analysis

- Is humanitarian intervention required? Define the objectives of the intervention, bearing in mind local preparedness measures and financial and human resource capacity.
- What are the priorities for action and resources in terms of sectors, affected groups, and geographical areas in both the short and longer terms?
- How should humanitarian stakeholders intervene? What are the recommended interventions, modalities, and practicalities required (e.g. support to the Ministry of Education, partnerships, etc.)?
- What are the information gaps and needs? How can information gaps be addressed in further phases of assessment. Provide recommendations on assessment design and coordination structure.

Identifying which of these questions should be answered and with which level of detail will help to shape and plan the overall scope of the analysis.

Whenever possible, analysts should reduce the range of inquiries and resist the temptation to include too much information. A fine balance should be sought between the scope (geographical areas, numbers of sectors, diversity of affected groups, etc.) of the assessment and the depth and granularity of information required.

Step 2 - Organize and summarize observations

The next step in data analysis is to organize and categorize what you have seen, heard or been counting, and impose a common structure upon the data that will facilitate further enquiry and comparison.

a. Data organization

Organizing data involves[4]:

- Gathering all forms, questionnaires, and observations in one place
- Checking for completeness, relevance, quality and usability of the data[5]
- Removing, filtering or separating those observations that are incomplete, not of immediate use, or which do not make sense. Recording and documenting your decisions
- Assigning a unique identifier to each observation.

The use of spread sheets for this step is highly recommended and allows data to be looked at from a number of different perspectives. All observations should receive specific attributes that will later ease the summarization or aggregation process[6]. Two generic sets of attributes can be identified: dimensions and measures.

Dimensions are additional information or descriptive details about observations or measures, allowing for detailed analysis and contextualisation.

The most useful dimensions for collected observations during assessments are[7]:

- The date the data was collected or the information it refers to
- The location or geographical area the data is applicable to
- The sectors and/or sub-sectors the observations represent
- The population segment or affected groups the observation derives from.

Dimensions categorize and describe measured objects in ways that help the analyst understand the meaning of the measures. They are the observation's metadata, generally categorical, context adapted and strongly linked to the analysis strategy defined earlier in the process (step 1, define purpose of the analysis). For example, *time dimension* allows to compare measures at different points over time, and ease comparison before/after. A *group dimension* allows to compare needs between different affected group, etc.

Many dimensions contain a hierarchy of attributes that support drilling[8] up and down. These include:

- Geographical locations, generally composed of a hierarchy of administrative levels: 1> 2>3>X
- Date dimensions can contain a hierarchy of year> quarter> month> week
- Sector dimensions can be composed of several sub-sectors, i.e. WASH which may include water supply, sanitation, hygiene, waste management and vector control
- Affected group dimensions (i.e. IDPs) can be composed of several types of affected sub-groups, for instance: IDPs in self-settled camps; IDPs in public buildings; and IDPs in host families, etc. These can be further broken down (e.g. by gender, age, or livelihood).

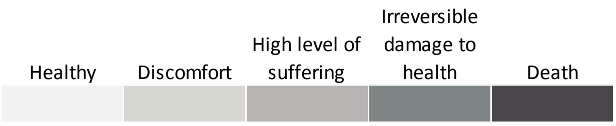
At the early stage of a disaster, information will be collected at a generic level (i.e. *overall affected area, affected people*, IDPs), and the degree to which lower hierarchical dimensions can be measured will be limited. Only later, can assessments successfully focus on more granular data and allow for more exhaustive measurements.

Measures are quantities being calculated or estimated. Three essential measures are recommended for use during assessments:

1. The severity level of the discrepancy created by the need and recoded, for instance, on a severity scale, i.e. from 1 to 5 as shown in the

[4] Procedures for data editing, cleaning, and coding goes beyond the scope and focus of this document and are not detailed here.
[5] See ACAPs technical Brief How sure you are?
[6] It is important to note that the assigning of attributes to individual pieces of information is not a license to extract data, especially qualitative data, from its holistic context. Qualitative data, like quantitative data, derives meaning first and foremost from its context.
[7] For qualitative researchers, dimensions can also encompass socio-cultural perspectives and perceptions.
[8] For more information on drilling, see Step 3, section C: Iteration and interaction.

following chart:



2. The number of people affected by a specific issue, if available. It can either be a qualitative quantifier (None, few, many, etc.) or an absolute number, if available.

3. The reliability level of the information. As illustrated in the following scale:

Ultimately, all available observations should be assigned dimensions and/or measures to facilitate the aggregation of similar responses into categories of analysis, as defined in the analysis plan.

Ensuring that each dimension is attached to one or several measures will further facilitate comparison, i.e. between magnitude or degree of severity as well as allowing for a credibility check across observations to verify the consistency of the findings.

b. Summarize/aggregate observations

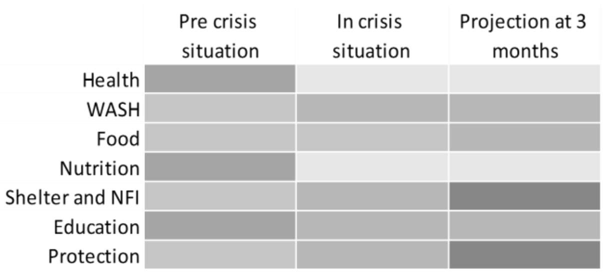
Summarizing data entails grouping like data with like data and aggregating the related measures. Aggregation is the process of consolidating multiple values into a single value.

For example, the number of admissions to a nutritional centre can be collected on a daily basis and aggregated into a value for the week; the weekly data can be aggregated into a value for the month, and so on. Likewise perceptions of access to health services, though qualitative in format, can be aggregated into measures of severity where perceptions may range from adequate, to poor, to life threatening. Aggregation allows patterns in the data to emerge which then become the basis for analysis and decision making.

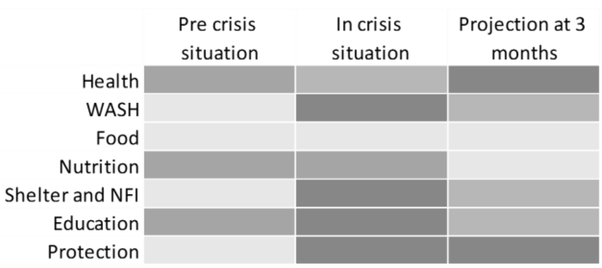
Different types of aggregation exist, such as sum, average, count, minimum or maximum values, first, last, etc. Aggregation can be done at different levels and across different dimensions. For example, the sum of food insecure people (measure) across several provinces or affected groups (dimensions) can be estimated. Likewise, the average percent of children not going to school (measure) and its evolution over time (dimension) can be calculated.

In the chart below, *observations* related to the severity of conditions have been aggregated at sector level and for two distinct affected groups, IDPs and residents affected.

Province Beta - *Resident affected*



Province Beta - *IDPs*



The grey colour gradient encodes a summary measure of the severity score provided by multiple observations, aggregated at the sector level, and organized around different point of time (pre-crisis, in-crisis and a forecast estimate). In this chart, the darker colour indicates a higher level of concern and gives a visual sense of priority that immediately attracts the analyst's attention.

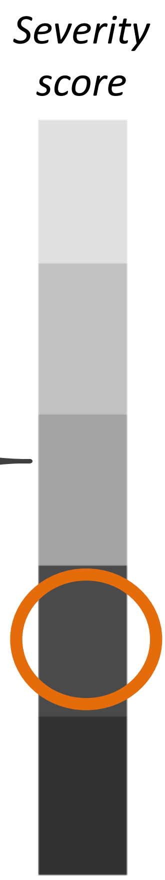
In those charts, the WASH in-crisis severity level was processed using data originating from several NGO assessment reports, minutes from various cluster meetings, and initial findings from field visits (key informant interviews and direct observations). By using the attributes previously discussed, data can be filtered to process only in-crisis observations related to the WASH sector, for Province Beta and for IDPs. Then the severity score is aggregated for this category of analysis, as outlined in the table below:

Observations → Summarization → Final value

Key informant interviews
Fewer than three litres/day/person available on average. No alternative safe water point available. Important increase of water diarrhoea cases reported at the health centre.

Direct observation
>3 hours queue time at the only water point. Limited water storage capacity at household level. Most often dirty jerry cans are used by households. Proximity of sewage system to the water point.

Secondary data review
Meteorological office reports rainfall will continue for at least two weeks and contamination risks will increase. UNICEF bacteriological test confirm high level of pollution in the water. UNICEF will distribute chlorination tablets and NFI next week to the affected households.



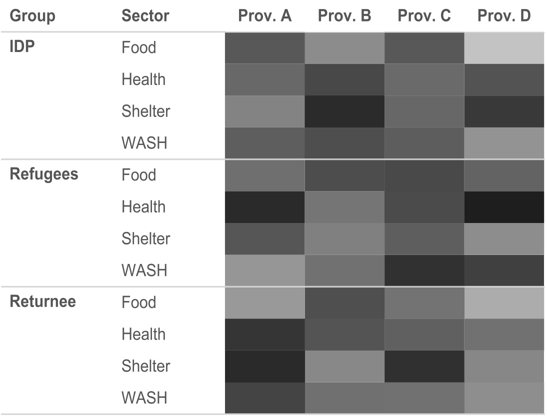
This operation can be repeated for other data points or observations and as needed. Having each observation *labelled* and *tagged* around dimensions and measures allows for quick classification, filtering and reorganization of data, depending on analytical needs.

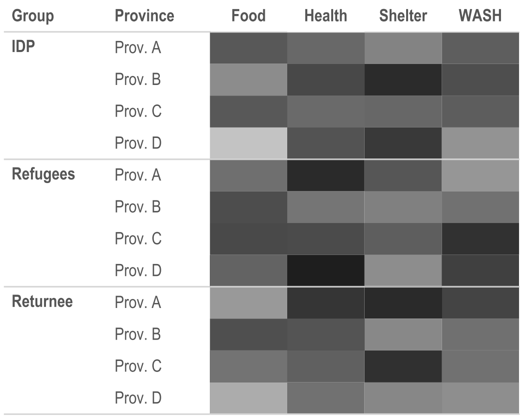
Effective use of dimensions and measures allows, for example, to quickly filter the data related to IDPs, food insecurity and province X, or to isolate issues with the highest severity score, faced by refugees in the health sector. Additionally, median severity scores can be calculated across all observations related to protection issues, in one particular provinces where several sites were assessed.

Moreover, permutations between dimensions allows for exploring data through different angles. This will later be used to present results to different target audiences with different information needs.

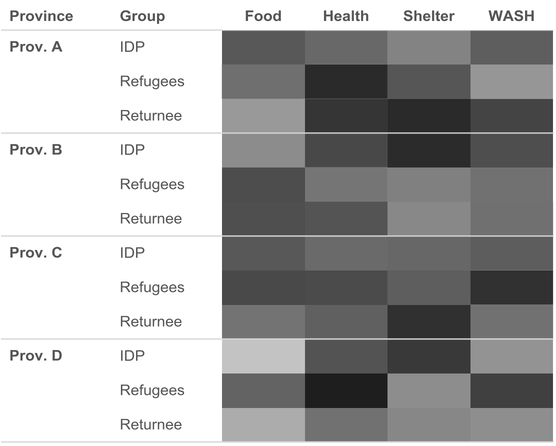
For instance, use of three dimensions (i.e. geographical areas, sectors, and affected groups) and one measure only (i.e. a severity score) allow for six possible permutations, visualizations and types of analysis, as shown in the following charts. Here, the same data is filtered, summarized, and displayed using different dimensions arrangements, so:

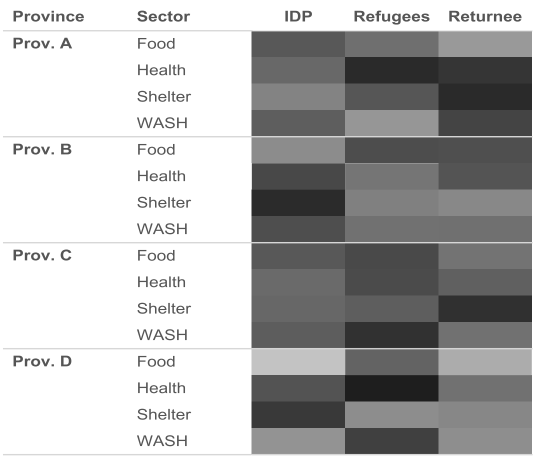
- A protection officer can use the following figures as they provide information at the group level (prioritization possible across provinces and sectors for each group).



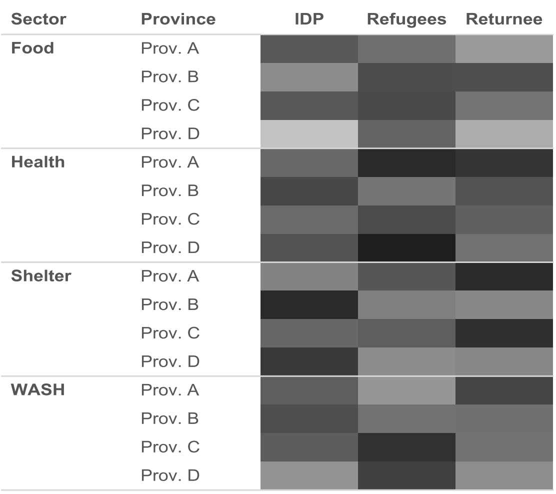


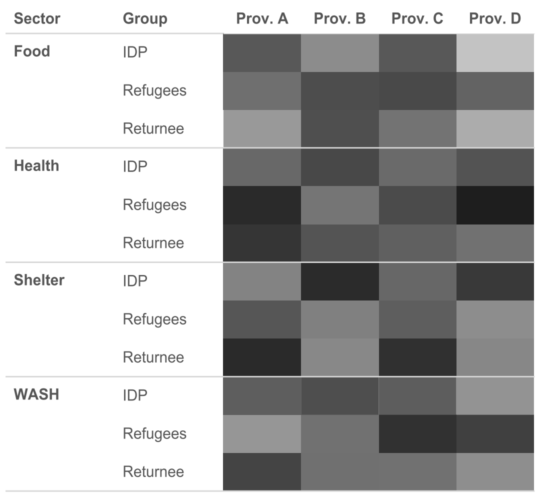
- A generalist can use the following figures as they provide information at the province level (prioritization possible across groups and sectors).





- A sector specialist will be more interested by information provided at the sector level (prioritization possible across provinces and groups).





At this point, an analyst might begin searching for patterns among the data and deriving meaning from what may seem unrelated and/or diffuse observations. This is when the real analysis work start.

Step 3 - Compare: identify patterns and relationships within the data

An observation, by itself, is not always interesting. The first question to ask about a piece of data (i.e. measure, number, judgment, perception, etc.) is: *compared to what*? , because data in analysis is only meaningful when compared to other related data.

Looking for differences or similarities between two or more observations or narratives allows for the identification of patterns or trends in the data, and ultimately the discovery of associations or relationships between observations or data points.

a. Compared to what?

The first step in comparing data is to decide what to compare. Often this will entail contrasting a figure from your own data against numbers or distributions from other sources or comparing a qualitative observation to a pre-existing baseline. For instance, you might compare the average nutritional status for children <5 years old in one country against international standards or you may compare today's bread prices against historic records for a different date at the same location.

Comparison is at the core of the analysis process. It is used to identify how things are alike and how they are different. Comparing requires thinking about the specific attributes or characteristics of the things that are observed and studied, and uses these characteristics as the basis for identifying similarities (comparing) or differences (contrasting). The following types of comparison are often used in needs assessments:

- **Humanitarian standards** can be used as the reference for measures of data, i.e. SPHERE minimum standards. If external standards don't exist or additional contrasts are required, observations or narratives within your own data can be used, for instance the average access to clean water in visited middle income neighbourhood.
- **Geographic** comparison entails comparison between different geographically delimitated areas, i.e., province A vs. province B, higher conflict intensity areas vs. lower conflict intensity area, etc. This type of comparison can be extended to other arbitrarily defined elements with spatial attributes, such as the type of setting, i.e. rural vs. urban or camps vs. non-camps.
- **Social group** comparisons identify different levels of needs and vulnerabilities between different population groups, i.e. agro-pastoralists vs. farmers, residents vs. IDPs, etc. This is especially useful to describe the variation of need between, and within, different affected groups identified in the IASC humanitarian profile [9]. Types of respondents are also a commonly used layer for comparison, i.e. male vs. female, children vs. older persons.
- **Time** comparisons are also useful and sometimes employed as a proxy measure to show the impact of a disaster, i.e. pre vs. post-disaster situation.

b. Patterns and trends

Patterns are a repeating event and demonstrate recurring themes or categories appearing in a predictable manner. They are regularities, variations or exceptions which stand out above the typical noise evident in nature or in raw data[10].

Our perception of patterns in qualitative or quantitative data is fundamental to the sense-making process. For example, certain health conditions may cluster in particular geographical areas or people from a particular group may apply similar coping mechanisms. These patterns may not be specifically what was looked for or anticipated, but they may be important in themselves and deserve increased attention, or they may shed

[9] http://cod.humanitarianresponse.info/sites/cod.humanitarianresponse.info/files/refdocs/iasc_guidelines_on_the_humanitarian_profile_common_operational_dataset_2012-08-07_0.pdf , 2011

[10] Annex 1 provides details on specific types of patterns that are most commonly found in data.

light on new areas of interest or specific elements of the data.

Here are three principles for identifying which details in the material are more worthy of attention than others:

- *What repeats? What goes with what?* Look for patterns of repetition or resemblance. In virtually all subjects, repetition is a sign of emphasis. Once apparent similarities have been located, analysts can refine their thinking by pursuing significant distinctions among the similar things (looking at differences within the similarity or similarities despite the difference).
- *What is opposed to what?* Look for binary oppositions. Sometimes patterns of repetition are significant because they are part of a contrast around which the subject matter is structured. One advantage of detecting repetition is that it will lead analysts to discover opposites which are central to locating issues and concerns.
- *What doesn't fit?* Look for anomalies, outliers [11], and things that don't fit. An anomaly is literally something that cannot be named, *a deviation from the normal order*. Anomalies help us revise stereotypical assumptions, and noticing them often leads to new and better ideas. Observations can fall outside the norms for three reasons: errors, extraordinary events or extraordinary people / institutions / organizations.

The following three groups of questions are typical of what goes on in an analyst's head as s/he attempts to understand a subject and identify patterns and trends[12]:

Define significant parts and make the implicit explicit:

- Which details seem significant? Why?
- What does the detail mean?
- What else might it mean?

Look for patterns:

- How do the details fit together? What do they have in common?
- What does this pattern of details mean?
- What else might this same pattern of details mean? How else could it be explained?

Look for anomalies and keep asking questions:

- What details don't seem to fit? How might they be connected with other details to form a different pattern?
- What does this new pattern mean? How might it cause to read the meaning of individual details differently?

c. Iteration and interaction

Because the purpose of analysis is to figure something out, analysts shouldn't expect to know at the start of the process exactly where they are going, how all of the significant parts of the data fit together, and to what end. The key to the analytic process is to be patient and confident in the knowledge that there are procedures analysts can rely on to take them from uncertainty to understanding.

The usual analysis approach is to begin with descriptive analyses, to explore and gain a *feel* for the data. The analyst then address specific questions from the assessment aims. To do this s/he explores the data in comparison with findings from the background literature, from his/her own professional experience, and from patterns suggested by the data itself.

Interaction with data allows the analyst to look deeper into meaningful patterns, trends, associations, exceptions, etc., and to filter out what is not needed, drill into details, and combine multiple variables for comparison in ways that promote a smooth flow between seeing something, thinking about it, and manipulating it. Ultimately conclusions and recommendations emerge as data is tested, refined, and retested against new information

[11] An observation or a value that is distant from the rest of the data.

[12] Adapted from Writing Analytically, Rosenwasser, Stephen, 2012

until explanations are consistent, triangulated, and defendable.

The most common iteration and interaction processes are detailed below (See also Annex 2 for more detailed description of these techniques).

Deconstruction breaks down observations into component parts and compares each part and its importance to the whole to gain a better understanding of it. Common deconstruction techniques are as follow:

- Analysis often requires examination of data at different levels of detail: aggregation involves viewing data at a higher level of summarization; disaggregation involves viewing data at a lower level of detail. Disaggregation allows for moving between high-level (the big picture) and low-level (the details) views of the data.
- Similar to disaggregation, drilling involves filtering data to a lower level in a defined hierarchy and excluding from the view all data not directly related to the specific data value that you chose to drill into. For instance, if you drill into a particular affected group category, your data shows only affected groups that belong to that category (i.e. all districts contained within a single province).

Manipulation involves the re-sorting, rearranging, and reorganisation of data (without fundamentally changing it). It is used both as a precursor to other data analysis activity as well as a means in itself of exploring data analytically. Common manipulation techniques include:

- The act of sorting data, especially by the magnitude of the values from high to low or low to high, features the ranking relationship between those values and makes it easier to compare the magnitude of one value to the next.
- Adding/removing variables: You might need to view different variables at different times during the analysis process, so it is common to add or remove fields of data as necessary to focus your analysis.
- When you want to focus on a subset of data, use filtering, the removal of everything you are not interested in at the moment or what is not immediately relevant to your question.
- Sometimes it is useful to group members of a data variable together, treating them as a single member of the variable (i.e. regrouping access and availability issues in most common problems expressed by the population). This may take the form of combining some members and leaving others as they are, or of creating an entirely new variable that combines all members of an existing variable into a group to form members of a higher level variable.

Transformation changes the data through processing to arrive at a new representation of the observations. Sometimes it is useful to express a data variable as a different unit of measure, such as re-expressing numbers as percentages.

Generalization takes specific data from observations and creates macro level general statements.

While summarization and aggregation are the most commonly used analysis techniques and are useful on their own, the real power of analysis comes from the ways in which different techniques are combined to form a rich and sophisticated process of data analysis.

d. Identify relationships

Ultimately, pattern detection also allows for the identification of specific relationships within the data and leads to the discovery of the most interesting, relevant or important stories that you will highlight for decision makers. A relationship is the correspondence, connection, or link between two or more variables of interest whose credibility can be triangulated against other data.

Identifying relationships is the centerpiece of the analytic process because it prepares the ground for moving from a simple description of the population conditions and settings to explanations of why things happened as they did in that particular setting during the interpretation stage.

Assembling data into an explanation is akin to putting together a jigsaw puzzle. With the puzzle, it is necessary to compare pieces of data and identify the relationships that individual pieces (data) have with one another. One way to do this is by grouping all pieces that look alike together, i.e. sky, and placing these pieces near the top.

Another way to put the puzzle together is to group pieces together which share a similar dimension, property or feature (e.g. color). Puzzle pieces will typically have to be rearranged many times before they emerge into a coherent pattern. When the puzzle assembler is successful, a whole structure will eventually be built, held tight by the interconnected pieces[13].

Most meaningful quantitative relationships in assessment data that are worth being explored, identified, or communicated can be classified into seven types[14]:

1. Time – how values change through time
2. Ranking – how values (associated with categorical items) are ranked according to size
3. Parts-to-whole – how values and sizes compare to one another and the whole
4. Deviation – how two or more sets of values differ (e.g. targeted beneficiaries vs. reached beneficiaries)
5. Distribution – how values relate to one another as a matter of proximity (i.e. their distribution through the entire range of values)
6. Correlation – how two sets of quantitative variables associated with a common set of entities behave in relation to one another
7. Geospatial – how the spatial positions and distribution of values (e.g. where they reside geographically) contribute to their meaning.

For qualitative data, Spradley's universal semantic relationship can be used[15]. Key informants usually express themselves by using terms that are linked together by means of semantic relationships, only the semantic relationships are hidden by the more apparent folk terms for things and actions.

Spradley classification provides a useful place to start in discovering ways to *read* stories people tell about their experiences relating to the disaster, as number of semantic relationships in any culture is quite few and appear to be universal:

Type	Form of relationship
Strict inclusion	X is a kind of Y
Spatial	X is a place in Y. X is a part of Y
Cause-effect	X is a result of Y. X is a cause of Y
Rationale	X is a reason for doing Y
Location for action	X is a place for doing Y
Function	X is used for Y
Means-end	X is a way to do Y
Sequence	X is a step (stage) in Y
Attribution	X is an attribute (characteristic) of Y

For example:

- Strict inclusion: Rice is a kind of food.
- Spatial: This village is a place in this district.
- Cause-effects: Lack of fuel results in tree cutting.
- Rationale: Insecurity is a reason for displacement.
- Function: Drugs are used to cure.

Step 4 - Interpretation: determining what the data mean

The two previous steps (organize and summarize observations, compare data to identify patterns, trends and relationships) in the analytic process have helped us know *what* is

[13] LeCompte, 2000
[14] Stephen few, 2006
[15] Adapted from J.P. Spradley, 1979

going on (what has changed, what is higher or lower, etc.). The next step is for analysts to understand what the patterns, trends, and relationships uncovered through analysis mean. Deriving this entails outlining alternative explanations and suggesting which conclusions can or cannot be drawn. This calls for identifying *why* data indicates a particular condition for one group and not another and why people behave as they do. In short, i**nterpretation is the process of attaching meaning to data.**

Data interpretation has structural approaches and logical limits. Some interpretations are more reasonable, coherent, and convincing than others. Good interpretations are neither absolute truth nor personal opinion; they are inferences, suggestions, or theses about what the data mean based on a foundation of empirical enquiry and individual expertise. When analysts have to interpret data, they draw on personal and collective knowledge, and use experience, logic, and parsimony to propose one or more plausible explanations from the data.

During interpretation, evidence supporting the claims must be weighed and findings contextualized before they acquire practical value to decision makers.

Interpretation involves:

- Ensuring that findings are useful for decision making. *What is important in the data? Why is it important?*
- Determining why particular conditions are here. *Why is this happening?*
- Putting data into context. *Does it make sense? Is it plausible?*
- Evaluating the evidence put forward in support of the explanations. *How sure we are?*
- Considering where the thesis leads and what conclusions and recommendations follow. *So what?*

a. Making critical sense of the data

Sense making is the ability or attempt to make sense of imperfect (and often incomplete) information on a situation which may be ambiguous or complex. Analysts extract cues from the context to help them determine what information is relevant and what explanations are best[16]. The points of reference link ideas to broader networks of meaning.

The following questions are among those frequently used by analysts to interpret humanitarian needs assessment data.

What is important?

- What did we find? What seems to be happening in this data? Do any interesting stories emerge from the responses? What is not happening? What should be there? What did we learn? Does this prove or disprove a hypothesis? What main points keep coming up (words, behaviours, attitudes)? What are the contradictions, dilemmas in the data? What doesn't seem to fit?
- What patterns and themes emerge in the results? Are there any systematic deviations from these patterns? If yes, are there any factors that might explain these deviations?
- What is new, what was expected, and what has changed since the emergency started (comparing pre and in-crisis data)? What has stayed the same when everything else has changed? Are the conditions being described the result of the disaster? What is surprising, perplexing or disturbing? What is not surprising and doesn't need to be presented in detail? If something didn't get worse, why not? If it is anticipated to get worse, what will trigger that?
- What is important or different about one group, one time, or one place when compared to another? Are there differences? Did different groups show different results? Are patterns consistent across different groups and sources of information? Do they make sense? How does one geographical

[16] Salancick & Pfeffer, 1978; Brown, Stacey, & Nandhakumar, 2007

area differ from another? What variations are there between locations?

Why is it important?

- <u>What does the observation imply?</u> Why does this information matter? What is known about similar past disasters or crises in the region and what does this tells us about our current findings? Does it make sense? Are they meaningful in a practical way? Are they useful for decision making?
- <u>What conclusions can we draw?</u> Where does this information get us? What theories or mechanisms might account for findings? What new hypotheses are suggested?

What do we do with the information?

- Do the results suggest recommendations for improving the humanitarian response?
- <u>What is missing?</u> What is the next level of detail required? Do the results lead to additional questions? Where do gaps in knowledge persist? Are there things you don't understand very well where further study is needed? What are the next research steps?

Making appropriate use of critical thinking should take into consideration all the parameters, constraints, values, and conditions that influence the way an analyst can acquire the best possible understanding of a situation. Using the following recommendations helps to ensure that findings are interpreted as neutrally as possible:

- Suspend judgment and stand back from the information given
- Examine it in detail from multiple angles and from varied sector viewpoints. Practice self-reflexivity: pursue significant questions, stay open, resist conclusions
- Check closely whether each statement follows logically from what went before
- Seek multiple perspectives and viewpoints, use sectoral experts, local knowledge, etc. Be able to determine and explain why different people arrived at different conclusions
- Look for possible flaws in the reasoning, the evidence, or the way that conclusions are drawn
- Be able to argue why one set of opinions, results or conclusions is better than another
- Be on guard for narrative or statistical devices that encourage an audience to take questionable statements at face value, such the use of emotionally laden case studies
- Check for hidden assumptions.

b. Correlation and causation

Pinning down the issues or underlying factors that cause an effect, or understanding why the issues caused a certain effect but not another, can explain why the data presents one conclusion over another. This involves distinguishing between *correlation* and *causation* and the role that can be played by *confounding factors* in skewing the evidence.

A *correlation* exists when one event is more likely to occur because another event has taken place**.** In a correlation although the two events may be associated, one does not necessarily cause the other, and the second event can still occur independently of the first. A correlational relationship simply says that two things perform in a synchronized manner.

For example, lessons learnt from past experience support a correlation between certain patterns of drinking water collection and the incidence of gender based violence. However, even though harmful water collection and violent behaviour may co-occur, there is no evidence that it is collecting water that *causes* violence.

A causal relationship exists when one event (cause) is necessary and sufficient for a second event (effect) to occur. The order in which the two occur is also critical. For example, for intoxication to occur there must first be consumption of alcohol. Determining cause and effect is an important function of assessments, but it is also a major challenge as causation can be complex and involve multiple factors.

To rule out that a relationship between two events has been distorted by other external factors, it is necessary to control for *confounding* (an extraneous variable that may influence the results).

Addressing possible confounders is an important part of proper interpretation of results. For instance, when assessing the impact of a crisis on a particular behaviour, e.g. an unusual and early destocking of livestock in Sahel, we must know whether this issue has coincided with new legislation or government support for destocking or an attractive change in terms of trades, before concluding it is due to a degradation of household food security.

To rule out confounding, additional information that could influence possible outcomes must be gathered and analysed. In the process of drafting conclusions, analysts should attend carefully to context as well as to the evidence they have at their disposal to support their claims.

c. Context is king

Meaning will always be context-dependent: When it comes to needs analysis, the impact of 50,000 people displaced in Burundi is different to 50,000 people displaced in China.

Data alone rarely provides complete answers, and looking at data without ascribing the proper context can yield imperfect, and often misleading, conclusions. Because facts are not always the truth, context is as important as the data itself, as it can suddenly make a problem meaningful and something worth considering.

Context locates the information you have gathered for analysis in place and time by answering the question: *does it make sense? Is it plausible? Is it important? Is it worth exploring further?* To acquire real meaning, data must be balanced with context, including experience, judgement and lessons learnt from past disasters and crises. Contextualizing results involves the association of related data points to yield the highest possible degree of understanding of the problem.

Three practical uses of context are recommended for needs analysis[17]:

1. Mainstream context in your interpretation

- Make sure that past experience (historical learning and lessons learnt) are taken into account. Know the theoretical context and base of existing knowledge for the topic at hand. This includes knowing what outcomes one can reasonably expect from a particular type of crisis in a specific area. This can complement your findings, ensure your conclusions follow a common logic and allow you to assess whether cofounding might be occurring.
- Relate new information to what is already known, i.e. contrast current bread price to last month's price, to the price in an unaffected district, or to the national price average. Use data attributes (spatial, temporal, group, etc.) to determine how the new observation relates to other historical observations.
- Recognize when two items are the same (or different) despite having been described differently (or the same), i.e. the difference between six reported cases of cholera versus one case reported six times. Analysts that cannot count or recognize unique objects can develop wrong assumptions and miss important relationships.
- Analyse data points as they come (using real time, micro-analysis), but ensure that such analyses are informed and complemented by the results or assumptions generated by meta-analysis processes where you try to understand what the whole set of information depicts or how larger pieces of assembled observations potentially fit together. Use end of the day debriefing or weekly brainstorming to depict the big picture based on the limited available information.

[17] Jeff Jonas, 2013

- Be aware of new observations that change earlier assertions, or revise or rectify invalidated assertions accordingly. *Now that I know this, are there assertions I have made that should have been made differently, in real-time and across ALL the assessment data?*
- Let disagreement and conflict coexist in the collected data; otherwise new emerging trends will not have a chance to add up to anything when interpreted. Don't over clean the data available, keep original field questionnaires and be open to the value of low-quality or not coded data.

Jigsaw puzzle solvers are often able to make correct guesses about the whole picture with remarkably few pieces in place. Having only half the puzzle pieces fitted together generally provides enough detail to show the outline of the entire puzzle image. Assembling small areas of the puzzle and studying them in depth allows one to determine the probable nature of the whole. This is good news for the analyst who may not be able to capture all the necessary data due to external constraints, but can still make a fairly good claim about the impact of the disaster by attending to context.

2. Understand how the data was generated
Observations cannot be fully understood out of the context in which they are produced.

- Acknowledge how the data was collected, the types of biases involved, the limitations of the assessment methodology, the qualifications of the enumerators, and the time they spent in the field, etc. Recognizing the strengths and limitations of the data and methods used will allow you to place your findings in the larger context of what is and isn't known.
- Recognize the strength and limitations of the evidence at hand to build your conclusions and explanations.

3. Develop context dependent explanations
Analysts must decide which possible interpretation, as seen through which plausible interpretive context, best accounts for what they think is the most important and interesting aspect of the data.

An interpretation is not a fact but a theory. Often, the best analysts can hope for with their explanations is not that the other will say *Yes, that is obviously right* but rather *Yes, I can see why it might be possible and reasonable to think as you do*.

The decision makers' willingness to accept an interpretation is powerfully connected to their ability to see its *plausibility*, that is, how if follows from both the supporting details selected by the analysts and the language used in characterizing those details.

Regardless of how the context is arrived at, an important part of getting an interpretation accepted as plausible is to argue for the appropriateness of the interpretive context you use, not just the interpretation it takes you to.

An interpretive context is a lens. What matters is that analysts share their data, show the reasons for believing that it means what they say it means, and do this well enough for a decision maker to find the interpretation reasonable.

d. Validity and evidence

Meaning must be reasoned from sufficient evidence, if it is to be judged plausible. Conclusions can always be refuted by people who find fault with your reasoning or can cite conflicting evidence. It is therefore especially important to locate the conclusions in the context of others assessments, surveys or studies who have achieved similar results, or to support your claim with good enough evidence.

Validity (in quantitative research) and trustworthiness (in qualitative research) is the degree to which conclusions about relationships or theses in our data are reasonable, credible or believable. While we do want to tell a story about the data, it is not just any story, but one which we can claim is an accurate reflection of the real situation. It means weighing up the arguments and evidence at our disposal.

To make a claim plausible, the analyst can support it in two ways:

- Corroborating evidence by using several pieces of evidence which individually support the claim.
- Converging evidence by using individual pieces of information that do not suffice to support the claim, but when linked together, constitute a robust body of evidence for supporting the claim. This type of argument needs to be highly contextualised and reasoning be made explicit.

Ultimately, the quality of the findings will depend on three distinct but related elements[18]:

1. The use of rigorous techniques and methods for gathering high-quality data.
2. The skills of data analysts[19] which depend on training, experience, and educational background, and the peer review process the interpretation has been through.
3. A fundamental appreciation of naturalistic inquiry, qualitative methods, inductive analysis, critical thinking and clarity of assumptions that underpin the analytic process.

e. From Observations to Conclusions

Asking *so what?* is a universal prompt for spurring the move from observation to *implication* and ultimately, to *conclusions.* It implies moving beyond patterns and relationships that have been observed in the data to suggested meanings, and finally tentative explanations about what the observations suggest.

By making explicit what is implicit, analysts address such questions as *what follows from this* and *if this is true, what else is true*? The process of drawing out implications is also known as making inferences[20].

Conclusions are your explanation of why the data look the way they do and should relate back to the research question defined in the assessment objective. Recommendations are based on your findings and conclusions. They generally take three forms:

1. Most severe problems and key priorities
2. Action that should be taken
3. Further information that should be gathered.

Analysts should state all findings clearly and unambiguously so that results are easy to interpret. This entails:

- Focusing on value added for decision making
- Articulating results and keeping the information as simple as possible but ensuring that important information is not omitted.
- Translating conclusions into easily understandable results
- Summarizing the main conclusions briefly and clearly in an executive summary which draws together the main findings from the needs assessment into a few coherent messages.
- Avoiding repeating information.
- Using accepted terms and standards, e.g., SPHERE, etc. and avoiding jargon and technical language. Key terms should be clearly defined to avoid misunderstanding and different interpretations. For example, what does *affected* or *damage* mean? What is meant by *site*?

Limitations in analysis will emerge from the interpretation phase and should be reported, either in written form (be explicit and honest about limitations) in the final report or in verbal presentations (be prepared to discuss limitations). Know the claims you cannot make and help readers understand the limitations of

[18] See also ACAPS Technical Brief: How Sure Are You?
[19] See Annex 3 for more information on analyst's skill sets.
[20] However, implication describes something suggested by the data itself, when inferences describe your thinking process: you infer what the subject implies.

the data and analysis so they do not misuse the results.

When developing conclusions or findings, analysts should clearly differentiate facts from judgements or assumptions and interpretation from data. Potential confounders should be openly acknowledged in the assessment results.

Communicate your audience with the *who, what, when and where* of your data. Without them, the audience cannot interpret conclusions, locate the data in space or time, and can, as a result, misuse the data.

Finally, identify clearly information gaps, i.e., the known unknowns, and information needs for further assessment phases.

6. References

Baty, S. (2009). *Patterns in UX Research.* http://uxmatters.com/mt/archives/2009/02/patterns-in-ux-research.php

Brennan, M. (2005), *Qualitative data analysis.*

Brown, A. D., Stacey, P., and Nandhakumar, J. 2007. *Making sense of sensemaking narratives.* Human Relations, 61(8): 1035–1062.

Coffey, A. and P. Atkinson (1996). *Making sense of qualitative data*, Thousand Oakes: Sage.

Creswell, J. (1998). *Qualitative inquiry and research design*, Thousand Oaks: Sage.

Creswell, J. (2002). Analyzing and interpreting qualitative data *(pp256-283)*. In J Creswell, *Educational research*, Thousand Oaks: Sage.

Dey, I (1993) *Qualitative data analysis*, London: Routledge.

Egger, A.E and A Carpi (2008). Data: Analysis and Interpretation. http://www.visionlearning.com/library/module_viewer.php?mid=154

EnglHUM1 (2013). *Five Analytical Moves.* http://englhum1.wikispaces.com/Five+Analytical+Moves

Few Stephen, 2004, *Eenie, Meenie, Minie, Moe: Selecting the Right Graph for Your Message* http://www.perceptualedge.com/articles/ie/the_right_graph.pdf

Few Stephen, 2008, *Now you see it*

Jonas Jeff, 2012, http://jeffjonas.typepad.com/

Holland, J (2009). *Deconstructing Analysis Techniques.* http://johnnyholland.org/2009/02/deconstructing-analysis-techniques/

Holland, J (2009). *Manipulating Data.* http://johnnyholland.org/2009/08/manipulating-data/

ICAP (2013). *Data Analysis and Interpretation.* http://www.icap.org/PolicyTools/Toolkits/EvaluationToolkit/4DataAnalysisandInterpretation/tabid/446/Default.aspx

Kloprogge, P. and J. van der Sluijs, A Wardekker (2007). Uncertainty Communication, *Issues and good practice.*

Kirk Andy, *Data Visualization: A successful design process*, 2012

LeCompte, M. D. (2000). Analysing qualitative data. Theory Into Practice, 39(3), 146–154.

Leek, J (2013). *Data Analysis,* Johns Hopkins. https://www.coursera.org/course/dataanalysis

Marshall, C. and G. Rossman (1989). *Designing qualitative research*. Newbury Park: Sage.

Maykut, P. and R. Morehouse, (1994). Qualitative data analysis: using the constant comparative method , In P. Maykut and R. Morehouse, *Beginning qualitative research*, London Falmer Press

Miles, M and A Huberman (1984). *Qualitative data analysis,* Newbury Park: Sage.
Miles, M and A. Huberman (1994). *Qualitative data analysis : An expanded source book (2nd edition), Thousand Oakes*: Sage.

Nelson, D and M. C. Hughes (2013). *How do I perform a data analysis?* http://www.wisegeek.com/how-do-i-perform-a-data-analysis.htm

Newton Suter, W. (2012) Introduction to Educational Research, A Critical Thinking Approach *Second Edition, chapter 12,* http://www.sagepub.com/upm-data/43144_12.pdf

Paul, R. and L. Elder (2013). *The Miniature Guide to Critical Thinking Concepts and Tools.* http://www.criticalthinking.org/files/Concepts_Tools.pdf

Salancick, G., and Pfeffer, J. 1978. *A social information processing approach to job attitudes and task design*. Administrative Science Quarterly 23: 224–253.

Scriven, M., and R. W. Paul, (1987). *Critical Thinking as Defined by the National Council for Excellence in Critical Thinking*.

Social Research Methods (2013). *Threats to Conclusion Validity.* http://www.socialresearchmethods.net/kb/concthre.php

Spradley, J.P. (1979). *The Ethnographic Interview*. Harcourt, Brace, Janovich.

Tesch, R. (1990). *Qualitative research,* New York: Falmer Press.

Tukey, J. W. (1962). The future of data analysis. *Ann. Math. Statist.*, Vol. 33, 1-67.

Tukey, J. W. (1977). *Exploratory Data Analysis*. London: Addison-Wesley.

Wilson, T (2009). *The most meaningful insights will not come from web analytics alone* in Analysis, Metrics, Social Media Analytics. http://www.gilliganondata.com/index.php/2009/09/14/the-most-meaningful-insights-will-not-come-from-web-analytics-alone/

7. Annex 1 Patterns and trends

A. Patterns

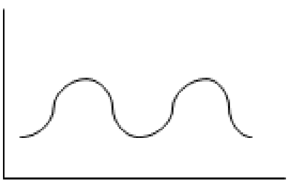
Meaningful patterns in quantitative data fall into three general categories:

- Large-scale patterns (a.k.a. trends). These are patterns that reveal what is going on in general, that is, as a whole (i.e. cereal prices have trended downwards over the course of the year).
- Small-scale patterns. These are patterns that reveal what is going on in specific subsets of data. For example, high incidence of water borne diseases correlates exclusively among the people of a particular district.
- Exceptions or outliers. These are values that appear outside of what is normal, standard, expected, regular or acceptable. For example, out of all visited areas, only one particular village showed a complete lack of food product in the local market.

Most common patterns in data are presented in the following paragraphs.

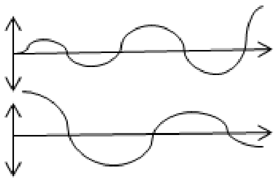
Repetitions: Often, in data, series of numbers or values repeat themselves. In a repetition, one value might consistently follow another or, when a value occurs, it might repeat three or four times before shifting to another value.

Repetitions can indicate either that a process is stuck or that there is a relationship between one event and another. For example, a longer task completion time (i.e. firewood or water collection) might be followed consistently by task abandonment periods. Repetitions are slightly, but significantly, different from cycles in that the entire sequence does not recur.



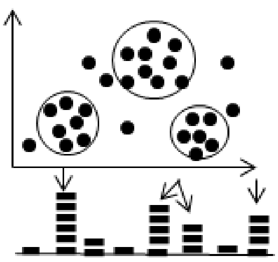
Cycles: Cycles are easily recognisable because each segment of the data looks similar. In a cycle, there is a regularly recurring pulse reminiscent of a heartbeat or the ebb and flow of the tides. Cycles indicate an underlying rhythm to an event you are observing and measuring. Examples might include the rise and fall of cereal prices in an area with various harvests or the peak of admissions in nutritional centres during the hunger gap period.

Recognising the presence of a cycle and understanding the driving forces behind it can help you plan ahead and gain deeper insights into your information. Cycles can also highlight the presence of negative forces acting against growth. For example, the incidence of malaria or cholera cases can be correlated to temperature and climate, while prices rises can be associated to the rainy season leading to increased time of transport in remote areas.



Feedback Systems: Feedback systems are like cycles that get bigger and bigger, or smaller and smaller, because some influence gives the system a small kick each time around. Variations become more and more accentuated as one event exacerbates the next.

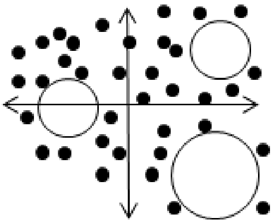
Feedback patterns can also indicate that a process is out of control. Small improvements in road infrastructures can lead to greatly increased traffic, resulting in the system becoming overwhelmed once more.



Clusters: Knowing how many concentrations are present is just as important as knowing where they fall. When clustering occurs in data, it may appear as a concentration of objects in just one small area, or data might group in several areas, as shown in the drawing, depending on what you are testing or researching. A cluster might represent something as simple as the distinguishing characteristics of different livelihood groups or geographical areas (urban vs. rural).

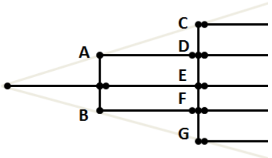
Depending on the complexity of your research, approaches to identifying clusters will vary. In simple cases, when dealing with just one or two characteristics, you can use a two-dimensional visualisation to highlight each concentration. For more complex cases, identifying clusters may require statistical analysis. In this last case, it is important to use a technique that is flexible, in terms of the number of clusters it generates.

Gaps: Gaps in the data represent the absence of any observable data, which can be just as informative as actual observations. Gaps in data are the opposite of clusters.

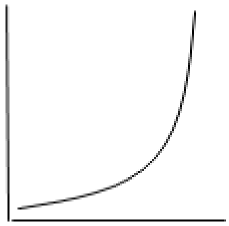


For example, looking through the demographic data gathered about the affected area may highlight an untapped population segment or reflect that targeted beneficiaries are not showing up at the distribution point or perhaps highlight that the health centre shows a significant drop in visit frequency during dry season. Whatever the scenario, gaps like those shown in the drawing tell us about issues and/or opportunities.

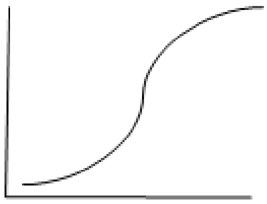
Pathways: The aim in analysing pathways is to present the data's branches and progression from node A to node C or D and so on. You can use sequential data (e.g. use of specific coping mechanisms) to identify major pathways. Higher-use paths receive a higher value, and you can use a thicker line or a different colour to identify the track most users are following. Analysing pathways isn't really a case of seeing a pattern as much as it is about recording, manipulating, and visualising your data so that it clearly illustrates a pattern.



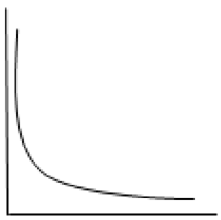
Exponential Growth: A constantly increasing rate of growth characterises exponential growth, as illustrated in the drawing. Exponential growth rates are typical of early adoption stages in a technology lifecycle, the presence of network effects, or the viral spread of a vaccination campaign.



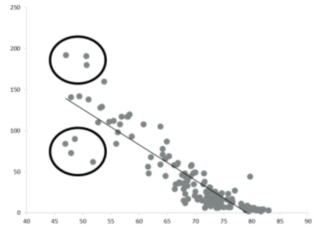
Diminishing Returns: Following an initial period of rapid growth, diminishing returns occur when the growth curve flattens out—still rising, but at a much slower rate, as shown in the drawing enclosed. It is clear that the curve is reaching some limit, possibly because of increasing competition or market saturation. This pattern is typically associated with mature products or strategy (e.g. vaccinations or a hygiene promotion campaign). The presence of a diminishing-return pattern can serve as a trigger for a more creative approach to product or activity enhancement, i.e. you might pare away sub-activities to refocus your strategy rather than adding more and more features or you might completely re-evaluate the way a product (malaria treatment, therapeutic feeding product) addresses a problem.



The Long-tail: In a long-tail pattern like that illustrated in the drawing, the data rises steeply, then falls off sharply, and levels off over a large range of low values. The long-tail is an example of a power law distribution that is common in nature and works typically for sales or a new product. The presence of a long-tail pattern might simply indicate that things are working normally, but it can also highlight deviations from the expected patterns in your data.



Exceptions: Also called outliers or anomalies, exceptions refer to abnormal values in a set of data. They can be described as data elements that deviate from other observations by so much that they arouse suspicion of being produced by a mechanism different than that which generated the other observations. Every abnormal value can and ought to be explained. Three possible reasons can create outliers: Errors (caused by inaccurate data entry, measurement or bias), extraordinary events (storm, earthquake, etc.) or extraordinary entities (richest person in the village, etc.).



We are so adept at recognising patterns that we sometimes detect ones that aren't there. When one pattern is found, especially one that we were prepared to find, we stop looking and can miss others that are unfamiliar and unexpected.

It is important at times to disregard familiar patterns and view data with fresh eyes. New patterns can emerge if we let ourselves look without pre-conceptions and drill down to specifics as well as scanning the big picture. Examine details and see what might be there that you cannot anticipate. Let yourself get to know the trees before mapping the forest.

B. Trends

A trend is a general direction in which something is developing or changing. They refer to the changes or movements in facts and figures over a period of time. They are usually used to describe the difference between two or more points on a graph, to compare two or more columns on a bar chart and to show the difference between information in a chart.

Basic trends can be categorised as:

- Upward (➚) or downward (➘) movements
- Stability (➔ no change or movement)
- Change in direction (➘ or ←, ∩ or **U**).

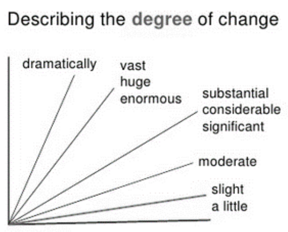
When data show a clear trend, all data progress in the same direction. In an upward trend, each subsequent piece of data is higher than the last. In a downward trend, each subsequent piece of data is lower than the last. Trends can appear in a variety of types of data such as site visits, price lists, transactions, time series, etc.

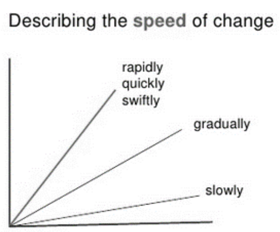
In a trend, the progression of data up or down is rarely completely smooth. Data will regularly dip down or shoot up against the general trend. When plotted as a graph, the lines representing the data will look jagged and rough.

Recognising trends is often a matter of looking at the data at the appropriate level of scale. If looked at too closely, data is simply a series of peaks and troughs lacking any real sense of direction. However, when one zooms out and views a greater range of data, the overall shape of the data becomes clearer.

Identifying trends, particularly when viewing data collected over a long period of time, can be difficult when the length of time each data point represents is short. Because data constantly shifts up and down, an upward trend can appear to be heading downward or vice versa. This is noticeable when looking at, for example, nutritional admissions in therapeutic feeding centres. Looking at only one or two days' data makes it difficult to identify a trend. However, looking at admissions over months or years makes trends clearer.

The degree and the speed of change also need to be considered.

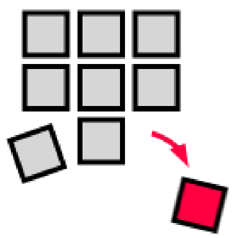




This annex was adapted from Steve Baty at http://johnnyholland.org/ and Stephen Few at www.perceptualedge.com/ .
Illustrations were adapted from Steve Baty.

8. Annex 2 Insight through interaction

a. Deconstruction



Breaking down data into component parts is a standard technique for analysis. This is in fact a classical definition of analysis. One example of deconstruction is turning an interview transcript into a series of separate comments or answers to questions. Deconstruction serves a dual role in analysis work: as a preparatory technique to get research data ready for other analytic processes such as manipulation or summarization, or even abstraction; and as a method of isolating, exposing, and testing assumptions deeply embedded in our mental models.

Deconstruction is one of the most frequently used and fundamental techniques in the analysis toolkit. Its' aim is to distinguish each component so as to allow inspection of each separately. In other disciplines, this process is used as a device for critical thinking, bypassing the potentially misleading image conveyed by the whole. Deconstruction can often be used in close association with other analytic techniques.

Examples of deconstruction include:

- Diagnosing causes through the identification of the system components and their interactions
- Quality control, i.e. testing the functionality of the health system by first identifying its components
- User interviews: identifying individual behaviour or opinions
- Task analysis: breaking down complex activities into individual tasks and their components.

There is a wide range of other examples of how deconstruction occurs, but the aim is always to reach a state where the smallest unit is defined by the assessment objectives. Smaller, more granular data provides for greater flexibility, and separating items, ideas or objects into their own data elements allows for greater control over how elements are treated and positioned with respect to other elements.

It should be noted that there are typically more things going on during the deconstruction process than merely breaking down the data. In the case of a key informant interview, the individual words, phrases or opinion may be tallied, grouped, manipulated or otherwise worked with to form new insight. To begin drawing connections and identifying themes between interviews, we need to break down – or deconstruct – the interview to the level of individual ideas or concepts, feelings, thoughts etc. Once the data is in this more granular form, we can carry out further analysis on the interview.

Deconstruction can be, and often is, built into the design of the research. Assessment results are a typical example of data where deconstruction is built into the research, through introduction of key variables that allow for further stratification and filter use (camp vs. non camps, rural vs. urban, etc., IDPs vs. residents, etc.).

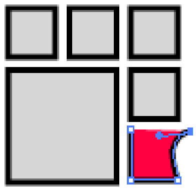
Dangers in deconstruction

It requires extra effort to break data down and then to store it in more granular form. It also takes effort to request and record extra data during the research process itself. So, whatever level of data granularity is used should be for specific reasons, and to address specific research questions.

Ultimately, analysis should lead to something substantively new. This can be difficult if we lose sight of the macro-level problem in pursuit of an understanding of the data in finer and finer detail. In studying the fine detail of our data, we can miss patterns in our data that help drive insights and accelerate the transition to design concepts. However, some patterns only become visible or apparent when we reach a level of granularity appropriate for the data.
Deconstruction can also generate noise in our data which obscures our sense-making abilities.

This noise may be the result of data overload, simply having too much information to allow for processing. Or it may be that small-scale, natural random variations are masking higher-level trends or patterns. In these cases, the use of summation and aggregation techniques might be an appropriate contrast to the deconstruction technique.

b. Manipulation



Manipulation involves re-sorting or rearranging the research data without fundamentally changing it. This is used both as a preparatory technique, as a precursor to other activities, and as a means of exploring the data as an analytic tool in its own right. The ability to *play* with the data is a critical process in analysis. We utilize this technique in many situations: searching for patterns or trends in our observations; or as a preparatory stage for further analysis. For example, sorting data in some way - be it through alphabetic, chronological, complexity or numerical – is a form of manipulation.

Manipulation can be seen as one of many low level analysis techniques with which we work every day. We've all encountered it in one form or another, and probably spent little time considering it. Yet it is one of the major workhorses of any analysis, and one which we should understand.

Many analysts talk about the need to immerse themselves in the data before they can make any kind of sense of it. Manipulating the data is a way of gaining that immersion – that familiarity – through direct engagement. Analysts undertake this process in a number of ways, depending on the format in which data has been stored. One common form of manipulation is to write out key concepts, observations, and ideas onto post-it notes and stick these to a wall. The analyst team them actively moves the physical post-it notes around, rearranging and grouping concepts and observations to help trigger creative ideas. This type of exploratory analysis can be powerful and is a key tool in the analysis arsenal. Exploratory analysis facilitates process of looking for connections within the data that make you think *hey, that's interesting* or that show patterns of behaviour.

One key characteristic of a manipulation technique versus related techniques like transformation is that the underlying data remains unchanged. The main thing being done is changing the relationship – logical or physical – that one piece of data has with another. Reorganizing the data helps highlight patterns that may otherwise not be apparent.

Let's start by taking a more detailed look at some of the processes that contribute to the manipulation of data.

Re-sorting is a technique aimed at changing the order of the data. Re-sorting is most often carried out on numerical or quantitative data, but can just as easily be applied to text content. Numerical, alphabetical, chronological ordering are the most common types of re-sorting. Sorting data helps to isolate significant individual values – the highest or lowest, most-frequent or least-frequent, first or last; and can also be a way of highlighting the shape of the data.

Re-arranging is an activity that typically involves the physical or digital repositioning of a data element so that it sits in closer proximity to another. This might be to organize the data into a narrative or to juxtapose contrasting ideas for discussion.

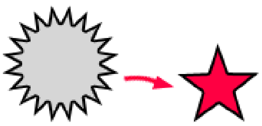
Much of the rearranging done is exploratory, although at times it will be more directed. In these cases, we might try to present a new configuration for our data to better support some activity (i.e. rearranging drivers for scenario building). Some of this manipulation will be more purposeful. For example, we might seek to reorganize data from similar affected group into similar piles or draw out common themes in community group discussions.

Re-sorting and rearranging data allows for the identification of patterns. There are different types of patterns one might seek to find and identify in research data, including[21]:

- Trends: the gradual, general progression of data up or down
- Repetitions: a series of values that repeat themselves
- Cycles: a regularly recurring series of data
- Feedback systems: a cycle that gets progressively bigger or smaller because of some influence
- Clusters: a concentration of data or objects in one small area
- Pathways: a sequential pattern of data
- Gaps: an area devoid of observations
- Exponential growth: rapidly increasing rates of growth
- Diminishing returns: decreasing rates of growth
- Long tail: a pattern that rises steeply at the start, falls sharply, and then levels off over a large range of values.

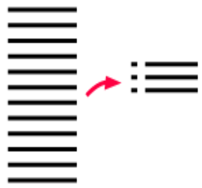
Like its' counter-part in analysis, deconstruction, techniques of manipulation are easy to undertake and require little or no preparation. Perhaps more importantly, manipulation encourages exploration. It works well as an unstructured activity, and therefore works well as an entry point into those vast collections of messy data points we're so often faced with early in the analysis. If you're not sure where to begin, start with manipulation – the more tangible and tactile the better.

c. Transformation



Transformation entails processing the data to arrive at some new representation of the observations. Unlike manipulation, transformation has the effect of changing the data and turning it into something else. For example, you can rescale results from an assessment so they fit a certain distribution, so you end up with10% A, 15% B, 25% C, 25% D, etc. Another example might be to convert raw data into a logarithmic form to reduce the impact of extreme values or to demonstrate power laws in the data.

d. Summarization



Collating similar observations together and treating them collectively is a standard technique in many quantitative and qualitative analysis methods. The goal of summarizing data is to generate an additional set of data, typically more succinct, that encapsulates the raw data in some way. This may be a short sentence that highlights the essential point from several minutes of an interview transcript, for example, *the participant reported difficulty in accessing primary health care*.

We can also summarize the data quantitatively using summary or descriptive statistics such as frequencies, means, and standard deviations. Unlike the process of abstraction, where specificity is sacrificed for the sake of clarity, or aggregation, where several data sets are *rolled up*, summarization seeks to characterize the underlying data.

Once again, spread sheets are a useful tool, especially when dealing with quantitative data. An equally useful medium for capturing summaries, particularly of qualitative data, is the post-it or sticky note. This medium is also highly suited to manipulation and exploration of the resulting data. One advantage post-it notes have over a spread sheet is that you can arrange and re-arrange them in two or more dimensions, so you can further manipulate and explore the summaries.

Index cards share many of the same advantages as sticky notes. They can be an excellent tool for capturing and working with summaries and have the added advantage of being relatively robust and can sustain a greater degree of handling.

[21] These are detailed in Annex One.

e. Aggregation

Closely related to summarization, aggregation draws together data from multiple sources. Such collections typically represent a *higher-level* view made up from the underlying individual data sets. Aggregate data is used frequently in quantitative analysis.

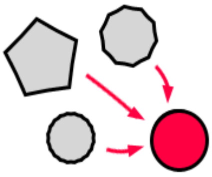
In one respect, aggregation is simply the process of bringing together data from a variety of sources and adding it together. In an analytic context, it also carries with it the connotation of combining those sources into something new.

A good example to highlight aggregation in action is the creation of a (fictional) severity index using data from:

- Direct observation of assessment teams
- Interview transcript with key informants
- Available secondary data.

Combining data from each of these sources enables the analyst to calculate a single figure. That single figure is the aggregate. Unlike a summary, which characterizes a single piece of data, an aggregate is a composite value.

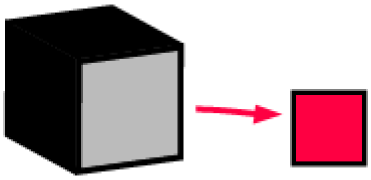
f. Generalization



Taking the results of a specific assessment and drawing general inferences about the broader population is one of the most common, but perhaps least understood analytical technique. Generalization draws a great deal of its strength from the discipline of statistics and the particular techniques of statistical inference.

Generalization is similar to abstraction in that it reflects a move from the specific to the general. It is a way of describing the common characteristics of the assessment data. An example of generalization might be: *security is important, especially for IDPs in self-settled camps*. This is a general statement based on an analysis of key informant interviews.

g. Abstraction

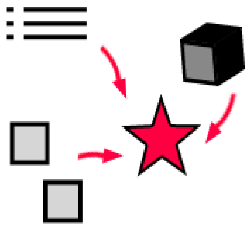


Abstraction is the process of stripping out the particulars, information that relates to a specific example, so that more general characteristics come to the fore.

The process of abstraction involves the progressive removal of specific data to retain only the essential information needed to communicate particular characteristics. For example, *affected population* is a more abstract form than *IDPs*, *affected residents* or *host population*.

Abstract representations are useful because they remove a lot of visual noise from the analysis process. What is left is a *high-leve*" depiction devoid of specific detail, but with a focus on those elements key to the assessment at hand.

h. Synthesis



Synthesis is the process of drawing together multiple concepts, ideas, objects and other qualitative data into new configurations. It is also referred as to the act of putting back together again, of integrating pieces, seeing the data set as a whole, or changing the lens of inquiry to wide angle. Similar in some respects to aggregation, synthesis typically deals with non-numeric data.

Synthesis is often undertaken towards the end of an analytic process as the reverse of deconstruction. So where one might begin by breaking data down into its component parts and examining them, synthesis recombines these components in new ways. If deconstruction allows for the critical examination of assumptions by isolating individual components, synthesis allows for exploring new configurations for the whole.

i. Final considerations

Not all of these happen simultaneously during the analysis process. You generally need to deconstruct before you can aggregate, and so on.

Deconstruction and perhaps manipulation are the only two techniques that address the *dirty* side of analysis – getting into the data, letting it sink in (reflection), and developing hypotheses about how to make sense of it, in relation to the needs of the decision makers you would like to inform.

There are other techniques that assist in that phase/time of the process, things like pulling together the themes and major learning (like deconstruction but less top-down/structural), mapping relationships, systems and processes to flush out the data, finding analogies to help you think differently about the data, and so on. These can be manipulated or transformed later, but initially in the analysis process it is important to see the data in multiple ways that might not be refined sufficiently to share with others.

This annex was adapted from the work of Steve Baty at http://johnnyholland.org/ and Stephen Few at www.perceptualedge.com/ .
Illustrations are from Steve Baty.

9. Annex 3 Quality of a good analyst

The following table provides an overview of the different skills and characteristics of a good analyst, followed next page by analyst's key principles.

Good analysts are

- Analytical, rigorous and precise
- Use context and evidence through observation
- Present a well-thought through argument
- Balance differing points of view
- Raise important questions and formulate them clearly and precisely
- Gather and assess relevant information, using abstract ideas to interpret it effectively
- Come to well-reasoned conclusions and solutions, testing them against relevant criteria and standards
- Think open-mindedly within alternative systems of thought, recognizing and assessing, as need be, their assumptions, implications, and practical consequences
- Communicate effectively with others in figuring out solutions to complex problems, without being unduly influenced by others' thinking on the topic.
- Sensitive to ways in which critical thinking can be skewed by egocentrism, socio-centrism, wishful thinking, etc.
- Intellectually honest with themselves, acknowledging what they don't know and recognizing their limitations.
- Listen open-mindedly to opposing points of view and welcome criticisms of beliefs and assumptions
- Base their beliefs on facts and evidence rather than on personal preference or self-interest.
- Are aware of the biases and preconceptions that shape the way they perceive the world
- Think independently and are not afraid to disagree with group opinion
- Able to get to the heart of an issue or problem, without being distracted by details.
- Have the intellectual perseverance to pursue insights or truths, despite obstacles or difficulties
- Have the intellectual courage to face and assess fairly ideas that challenge even their most basic beliefs
- Love truth and are curious about a wide range of issues.

Poor analysts are

- Subjective
- Poorly structured
- Descriptive (i.e. structured around narrative, rather than analysis)
- Generalised
- Vague
- Base analysis on unfounded or unacknowledged assumptions
- Based on value judgements
- Prolix (i.e. wordy)
- Circular in its argument
- Often think in ways that are unclear, imprecise, inaccurate, etc
- Often fall prey to egocentrism, socio-centrism, wishful thinking, etc.
- Pretend they know more than they do and ignore their limitations
- Are close-minded and resist criticisms of beliefs and assumptions
- Often base their beliefs on mere personal preference or self-interest.
- Lack awareness of their own biases and preconceptions
- Tend to engage in 'group think', uncritically following the beliefs and values of the crowd.
- Are easily distracted and lack the ability to zero in on the essence of a problem or issue
- Fear and resist ideas that challenge their basic beliefs.
- Are often relatively indifferent to truth and lack of curiosity
- Tend to preserve when they encounter intellectual obstacles or difficulties.

The analyst's craft

Analyst's job is to…	• Define intelligence problems and issues clearly. • Anticipate trends and developments. • Provide our end user's with judgments and insights. • Tell our consumers what is *really* happening in a situation. • Be responsive to our end user's. • Evaluate raw information critically to determine its relevance, reliability, and weight as evidence. • Extract key points from raw information or otherwise identify what is important in a sea of detail. • Make meaningful characterizations about data by "synthesizing" them into judgments that are greater than the data they're based on. • Deal with ambiguity, uncover and test assumptions, reconcile conflicting information, and guard against bias, subjectivity, deception, and "politicization." • Consider the views of others. • Evaluate alternative scenarios. • Assess implications for our end user's.
Analyst's job is not to know everything	• We make judgments on the basis of information that is incomplete, conflicting, and of varying degrees of reliability. • We provide the best possible answer given the time and information available. • We do not pile up detail. Data dumps are not the way to show our expertise. • And we are not historians.
Analysts interpret and synthetize	• We interpret, not describe/narrate • We render the complex clearer, not simpler. • We read, weigh, and assess fragmentary information to determine what it means, to get the "big picture." • We see the forest, not just the trees • We conceptualize, focus, frame, and advance defendable judgments. • We write or speak so clearly and simply that the reader cannot possibly misunderstand our message. Everyone who reads what we have written or hears what we have said comes away with exactly the same message. Our job is not done until that is accomplished.
Analyst ethics	• We promote and protect objectivity: without objectivity, our products have no value, and we have no credibility. • We have the courage to press our opinions where the evidence warrants, no matter how unpopular our conclusions might be, and courage to recast our findings when our thinking changes or when we find new evidence. • We must not allow our products to be distorted by motivations that could range from individual biases and misplaced assumptions (those of others or our own) to implicit or explicit pressures to twist analysis for policy or operational reasons. • Primary responsibility clearly rests with the analyst(s) concerned and with the appropriate layers of management. • Responsibility for encouraging analytic objectivity must be shared across a wide spectrum of people. Pursuing objectivity requires a team effort and special vigilance to prevent bias from affecting analysis. A number of people can become involved, including colleagues from other parts of the organization, from different components of the humanitarian community, and, finally, the decision maker's. • We must submit the best draft we can, a draft that shows we've spent a great deal of time up front thinking through the problem logically and planning the product before we started drafting. Provides sound substantiation for our judgments. Is written in a clear, concise, precise, and well-structured style. Demonstrates we've considered other outcomes, rejected them, and why.

Adapted from *Analytic Thinking and Presentation for Intelligence Producers*