

**TRAINING PACKAGE FOR USING SOCIAL SCIENCE IN COMMUNITY ENGAGEMENT AND/OR COMMUNICATIONS ACTIVITIES**

**SESSION 4.6:** Qualitative data analysis in operational social science research

SESSION CONTENT

**Learning approach:** Real-time presentation, individual and group exercises, case examples

**Delivery mode:** Online and offline, 110 minutes approx.

**Essential sessions to have completed before this session:** 4.1, 4.2, 4.4

**Summary:** This session aims to improve knowledge around the tools and steps to analyse qualitative data in operational social science research.

**Learning outcomes:**

* Understand the steps to analyse data collected with qualitative methods like interviews, observations and focus group discussions

FACILITATING THE SESSION



**TRAINING PACKAGE FOR USING SOCIAL SCIENCE IN COMMUNITY ENGAGEMENT AND/OR COMMUNICATIONS ACTIVITIES**

Introduction: (5 minutes total)

Talk through session summary and learning outcomes.

Position this session in the question flow.

1. How to ensure that this information goes back to communities? To inform community-level actions and decision-making of the broader response?
2. What methodology and tools should be used to collect and analyse this information?
3. How to track the information used to ensure that it effectively contributes to operational and strategic priorities?
4. Who can collect this information?
5. Does this information already exist? Is there a related needs assessment or study?
6. What information is needed?

**DATA TO ACTION:**

Key questions in social science research

1. Who needs this information?
2. How to ensure that the information is used to make operational and/or strategic decisions?

Qualitative methods (10 minutes total)

|  |  |
| --- | --- |
|  | Question to participants (5 minutes):  Can anyone remind us from the (previous session) what qualitative interviews, focus group discussions and/or observations are?  Online: Invite the participants to write the answers in the chat function and share it with everyone  Offline: Ask one participant to give a recap for each method |

Recap of interviews, focus group discussions and observations

Key informant interviews (KIIs) or semi-structured interviews (SSIs) or in-depth interviews (IDIs) are conducted with people who can provide detailed information about the topic you want to know more on. They use a set of predefined questions or a topic guide covering the area of interest. They allow interviewees/key informants to provide both first-hand experiences and personal opinions. Key informant interviews can be a helpful method when seeking information on sensitive issues that are difficult to discuss in a wider group. They are also suitable when engaging vulnerable individuals who are difficult to reach and/or who require extra time to build trust on a one-to-one level. You might face both of these scenarios in your activities. KIIs can be conducted with a range of people – in practice it often starts off with visible community leaders to better understand the situation, start to identify social dynamics, vulnerabilities and risks, etc., but then can be expanded to include other members of the community who can share information to help answer your research question.

Observations are when the researcher observes participants ongoing behaviour in a natural situation and takes very detailed notes. This allows for thorough documentation of what people do in a given scenario and can provide direct information on their actions and behaviours. Observation can be used to verify information collected verbally from participants – is what they say they do actually what they do? For example, if a health worker interviewee describes giving detailed information on vaccine side effects to patients, then additional observations would help verify this.

Focus group discussions involve in-depth discussion with a group of people about their experiences and beliefs on a certain topic. Again, these are based on a pre-set group of questions or a topic guide covering the area of interest. This method allows for members of the group to agree or disagree, which gives insight into the range of ideas and opinions and levels of agreement and consistency. This is important as many of the issues that community engagement/communications activities address involve a range of beliefs, attitudes and related actions, all of which should be understood – e.g. contraception use.

We will discuss qualitative data analysis considering information generated from these three methods.

*Although we provide an overview here, to apply these data analysis approaches you may need the support of someone with  
 a technical background in social science research*

Steps of thematic analysis (70 minutes total)

There are many ways to analyse qualitative data. Today we will cover the basic steps of thematic analysis of qualitative data.

Let us first consider the data:

* The analysis of interview, observation and focus group and data can be quite complex
* It involves the analysis of both verbal and nonverbal communication
* The words and behaviours of participants will provide answers to the research question

Step 1 – Transcription and translation (10 mins)

* If you recorded the discussion sessions or interviews, you will be able to produce a word-for-word transcript. Note that transcribing takes time. It can take four to eight hours to accurately transcribe one hour of audio, depending on how fast you can type. So, make sure to include time for transcription in your workplan. If time is short, you can choose to listen to the recording in real time and just take notes or note the most important points, rather than typing each word. Make sure you state clearly on the document that these are notes, and not a word-for-word transcription. You may need to go back and check the audio again if you want to include an exact quote in a report.
* Translate the transcripts if the discussions were recorded in a language different from the language in which the analysis will take place. Compare the transcript with the handwritten notes taken by the note-taker to fill in any gaps due to a sound problem, and add in any other observations of nonverbal behaviours. If time is short, you may choose to transcribe directly into the target language. The downside of this is that it does not allow for any quality assessment of the translation, or for easily checking back on exact quotes in the original language.
* If it was not possible to record the discussions (for example, if the participants did not agree to be recorded), you will need to work from your notes. It is a good idea to sit down with your notes as soon as possible after the session to check they are legible, to complete sentences you did not have time to finish writing, and to add things from memory, including visual observations that you did not have time to record during the session. If you leave it too long, you will not remember these details and you may even find parts of your notes too hard to understand, which means lost data.

|  |  |
| --- | --- |
|  | Question to participants (5 minutes):  What are some challenges or top tips from your experiences with transcription?  Online: Invite the participants to write the answers in the chat function and share it with everyone  Offline: Collect answers from one or two participants and summarize  Show **Handout 1** for an example interview transcript (explaining that FGD data can be handled in the same way), and **Handout 2** for an example of observation fieldnotes. Ask participants to read these. |

**Step 2 – Coding (15 mins)**

* Once the transcribing and translation is done, the next step is reading and coding the data in the transcripts or notes. A code is usually a word or a phrase that represents a recurring theme or idea in the data. For example, in **Handout 1** and **Handout 2**, a key theme might be ‘low motivation’. These can be converted to initials or colours to make the coding easier, for example ‘LM’. Coding involves reading the data, assigning different bits of data (the text) to categories and labelling them with the word, phrase, initials, or colours chosen.
* There are many ways this can be done, and the best way will depend on the amount and type of data you need to analyse and how specific the research question is, as well as the preferences of the analyst. We will give two examples here, using the common software programs Word and Excel. Coding can also be done using qualitative data analysis software (as described in the final section), or even using pen and paper. Whatever platform you decide to use for your analysis, you will first need to decide whether to use deductive or inductive coding.

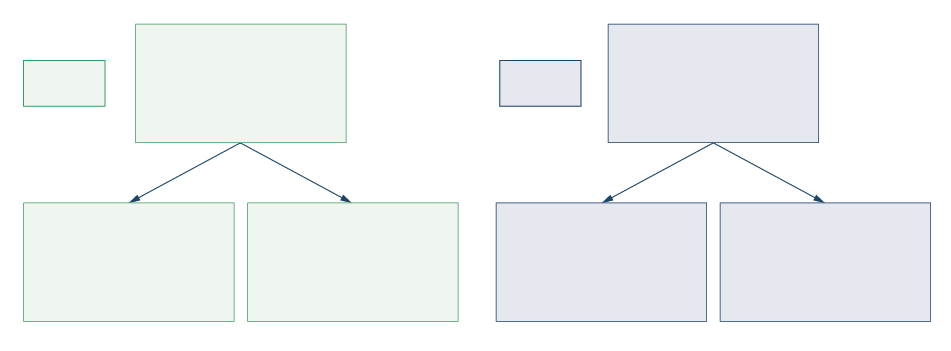
***Deductive vs. inductive coding***

* Deductive coding means you create a predefined set of codes before starting your analysis. This can be useful if you already have a good idea of the research questions you want to answer and the types of themes that you want to explore. You might also have an idea of the types of themes that will emerge in your data after reviewing previous research. For example, in the research in Handout 1 and 2 a predefined code might be ‘A-NGO = Attitudes towards NGO services’. The downside of deductive coding is that it might create some bias and prevent you noticing new themes emerging that you hadn’t expected.
* Inductive coding means you do not have any predefined codes, but create the codes as you read through the data. This method can be useful if your research question is broader and you do not want to be influenced by preconceived notions of what might be the most important themes. But it can be more time-consuming as codes emerge and change throughout the process and you may need to go back and re-code data as new insights emerge.
* A good compromise is to use a mix of deductive and inductive coding, starting out with a predefined codebook, but being prepared to modify it, adding new codes, blending codes or splitting codes as new themes emerge.

***Developing a codebook***

* If you have chosen to use deductive coding, you can prepare a ’codebook’ before you start your analysis. If you choose inductive coding, you can develop the codebook as you go. This can be as simple as a list of words or short phrases in a Word or Excel document to guide you. For example, for the research discussed in **Handout 1** and **Handout 2** you would include: ‘LM=Low Motivation’. In this document you can also create a hierarchy of codes, with codes and sub-codes, and should include a description of what each label means.

Figure 1: Example hierarchy of codes for a coding book. You can see how the sub-codes relate to the codes.

**

**HC**

(Role of host   
community

**RV**

(Risk of violence)

**R**

(Rejection)

**PRB**

(Perceived risk of borrowing money)

**LM**

(Low motivation)

**PB**

(Psychological   
barriers to work)

**CODES**

**SUB-CODES**

* Codebooks can be especially important if you are working with a *team of coders* and need to make sure everyone is using the same codes. If this is the case, the team should be in constant communication throughout the analysis phase, to ensure all coders have a shared understanding of what the different labels mean, and to agree on any changes made or new codes added throughout the process.

***Coding using basic word processing software (e.g. Microsoft Word) or on paper:***

* If all your data is in Word or a similar word processing software (e.g. typed transcripts), you can place all the text in a table with a column to one side.
* As you read through the text, type the relevant codes next to each piece of data in the column to the side.
* If you do not have too many codes, it can be useful to use different colours for each code, to help you to visually process the different emerging themes and their frequency.
* You can also do this straight onto handwritten notes, by drawing a margin for you to write the codes in.
* Note that some bits of text can have multiple codes attached to them.

Figure 2: Example of coding done on Word (**Handout 3**). Figure 3: Coding book extract –   
used to guide coding in Figure 2 (**Handout 3**).



***Coding using basic number processing software (e.g. Microsoft Excel)***

* Set up an Excel spreadsheet (or similar) with your predefined codes in a row along the top or create new columns as you go.
* As you read through each transcript, copy and paste relevant pieces of text under the column with the relevant code, so that all pieces of data relating to that code fall into the same column and can be easily scanned later.

Figure 4: Example of coding done on Excel

|  |  |
| --- | --- |
|  | Individual exercise (15 minutes)  Ask participants to look through the interview transcript in **Handout 1**. Using either Microsoft Word or Excel, code the transcript. Use the hierarchy of codes from Figure 1 as a starting point. Also feel free to identify new codes and themes.  Online and offline: Ask two or three participants to present their themes to the wider group |

Step 3 – Identifying and grouping themes (10 mins)

* The next step is to identify themes that have emerged from your data, see if there are links between the themes and categories of themes, and generate conclusions to your research questions.
* You may choose to select an existing model (e.g. the behavioural drivers model – see session 2.2 for more information) to guide your analysis and help to illustrate the relationships between themes. In this case you can draw themes from the model (i.e. deductive coding). Otherwise, you can look back at your data and name the themes as they emerge (i.e. inductive coding).
* As with coding, if you are working with a team of researchers, you will need to be in constant communication throughout the process to agree on the themes and to add new themes.

**WHAT IS THE DIFFERENCE BETWEEN CODES AND THEMES?**

Codes are basic analytic units. You can think of them as labels (words or short phrases) given to a piece of text about a particular topic. They are really just tools for organizing your data. (E.g. ‘Gender differences’).

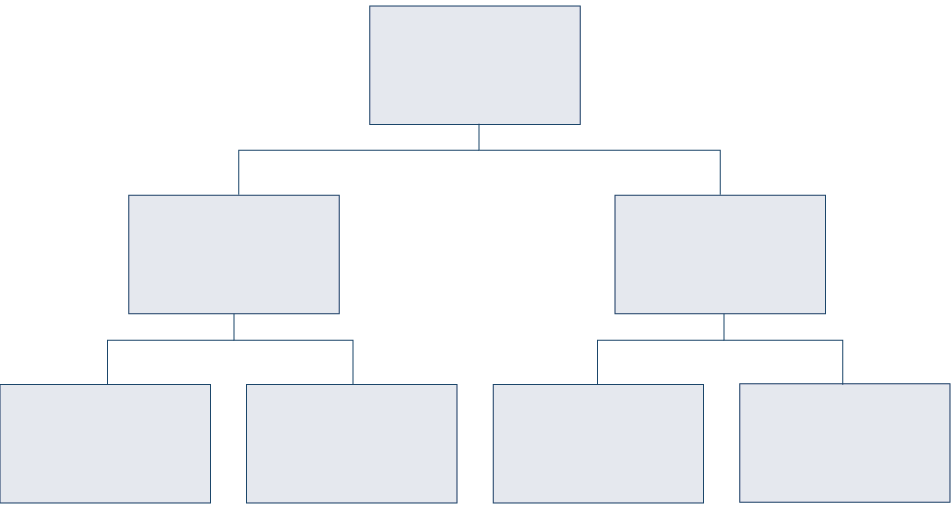
Themes begin to emerge when we look over our codes and start to group them into categories that are relevant to our research questions. They tend are broader and more inclusive than codes, and might be expressed in longer phrases or sentences. (E.g. ‘Gender differences influence access to healthy foods’).

* When we start to identify and group themes, we think about how they relate to the research question. So, keeping with the examples shown above, let’s say our research question is:

*What factors influence access to healthy food for adolescents in rural Kenya?*

* In both of the examples above, we have started to identify themes and how they are related through the coding process, by creating a hierarchical code book.
* As you can see in Figure 3, we have already placed the sub-codes ‘Growing independence (GI)’, ‘Restricted movement (RM)’ and ‘Gender
* differences (GD)’ under the category of ‘Social position’.
* If we look back over our codes and how frequently they appear, we might realize that an overarching theme that relates directly to our research question is: ‘Social position influences access to healthy foods for adolescents in rural Kenya.’
* Beneath this theme, we might have other related themes, such as: ‘Gender influences social position’; ‘Education influences social position’; ‘Gender influences access to healthy foods’; ‘Education influences access to healthy foods.’

Figure 5: Identifying themes



Social position influences access to healthy food

Social position

Access to food

Education influences access to food

Gender influences access to food

Education influences social position

Gender influences  
social position

We will now give examples of how we can categorize coded data under themes using Word and Excel.

**IDENTIFYING AND GROUPING THEMES USING BASIC WORD PROCESSING SOFTWARE   
(E.G. MICROSOFT WORD) OR ON PAPER**

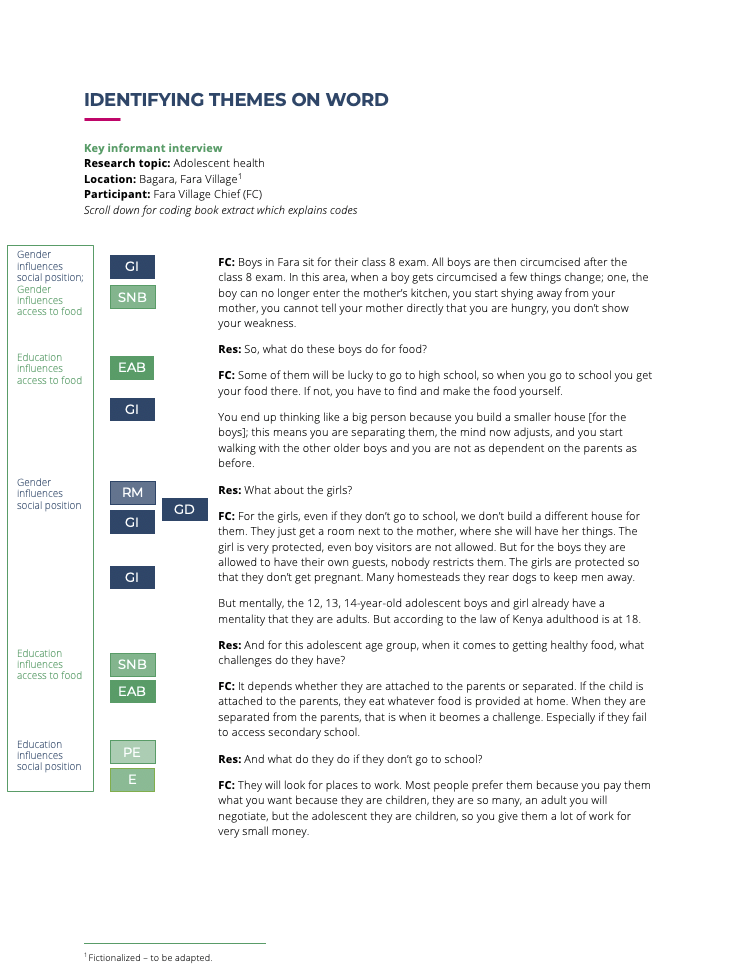


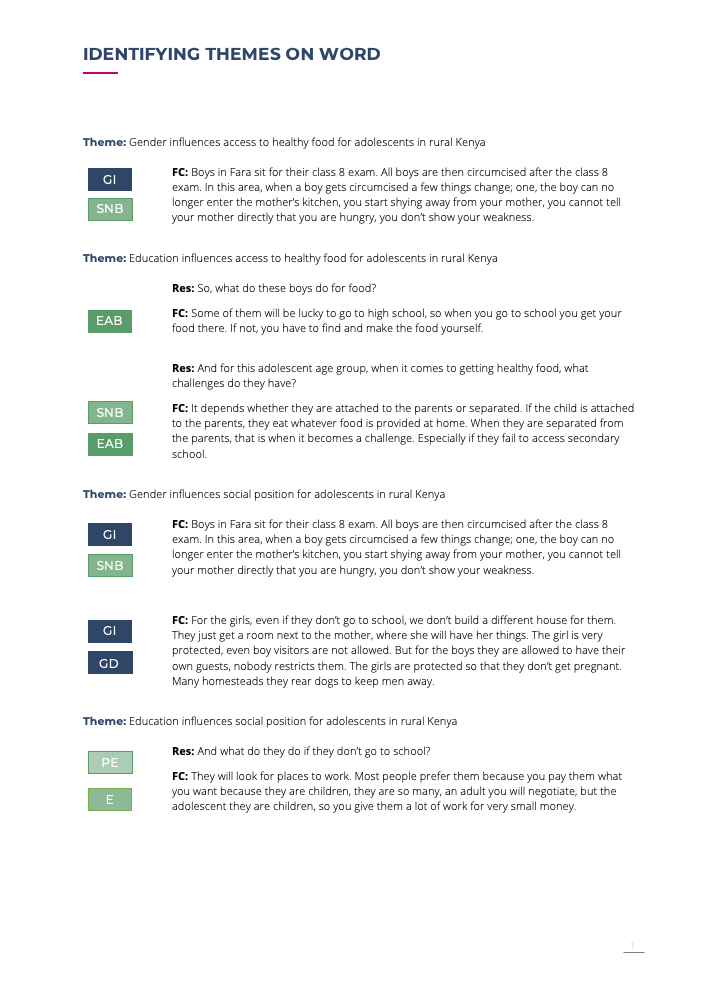
This example includes two steps, but you may only need to complete one or the other depending on the complexity of your research project.

Step 3A:

* Read through your coded transcripts again and you will start to see broader themes emerging that may encompass several codes.
* Add another column to the left of the codes and add descriptive themes in this column.

Figure 6: Identifying and grouping themes on Word (**Handout 4**).



******

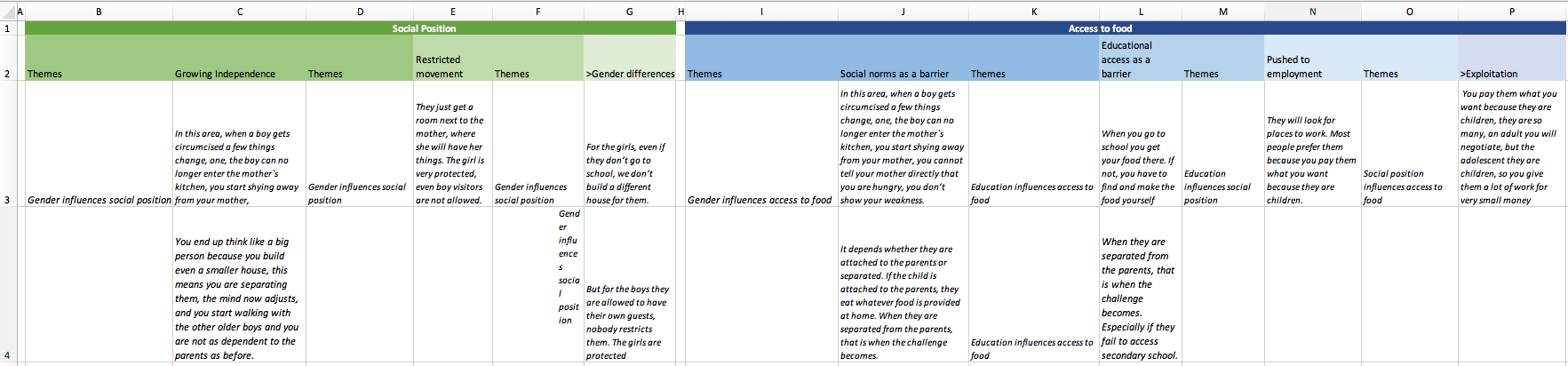
Step 3B:

* Write each theme as a heading in a new Word document.
* Paste relevant pieces of data under the relevant theme heading.
* You can colour-code your headings to make it easier to find the right theme quickly.
* Once you have done this, you will have all the data relevant to a theme in one place, and can look back over it to draw conclusions from the available data.
* Note that this method can become unwieldy if you have a lot of data. It is best for smaller data sets that need to be analysed quickly.
* This can also be done on paper, by literally cutting out relevant sections of text and placing them in theme piles.

***Identifying and grouping themes using basic number processing software (e.g. Microsoft Excel)***

* If you have done your coding using Excel, you can add another blank column alongside each ‘Code’ column, with the heading ‘Themes’.
* As you look down the column where you have grouped data relevant to a code, you will see themes emerging. Label these themes in the new blank column.
* This will help you to understand the different themes that have emerged relevant to a given topic, and their frequency.

Figure 7: Identifying and grouping themes on Excel



Step 4 – Key Findings (10 mins)

Having identified the different themes, you can then draw conclusions from your data to answer your research questions. For example, if our research question was ‘*What factors influence access to healthy food for adolescents in rural Kenya?’*, after our analysis we may be able to conclude that*: social position influences access to healthy foods for adolescents in rural Kenya*, and *gender and education influence social position*.

There are a few things to consider in selecting key findings:

Consider response frequency and salience

* How many times did a similar theme or response emerge? How much importance should we give to it? Is it representative to all participants/certain groups/a few individuals?
* In qualitative research we can also place value on something that only one person or a small number of people have said. It may be something that others also agree with but are choosing not to say for some reason. Or perhaps we didn’t ask the question. This would lead to further enquiry. If it is something important, we could include it in our reporting, but be clear as to how representative it is of all participants, and note the need for further research.
* Note the intensity of the respondents’ comments, as well as how specific they were with the details, to assess how much weight to give to a response.

Be aware of response bias

* If findings are unexpected, or if everyone is saying the same thing, you may wonder if there is some response bias. For example, if you as the researcher are from a humanitarian organization and all the participants are saying how wonderful your organization is, you might wonder if they are saying it to please you rather than because they believe it. This would call for further research, this time trying to mitigate the potential biases you have become aware of. In the meantime, the best way to present your findings is to:
* Be transparent, and present potential limitations openly.
* Triangulate with other data and sources.

Triangulation

* We can’t always take people’s words at face value. This is not because they are lying. There are many reasons why someone might say one thing and do something else (e.g. they are not aware of it, it might be rude or embarrassing to say it, they are saying what they think will be most helpful for the researcher). This is where triangulation is important. This means comparing and cross-checking your data with other data to see if there are any discrepancies, and if there are, trying to understand why.
* You should triangulate your data with:
* Your own observations of what people actually do, as compared to what they say.
* Other data sources.
* You should also validate your findings by sharing them with the participants and other experts in the area, and asking for feedback. That way you can check if you have misinterpreted any of the data or placed too much emphasis on things that may not be so important for the participants themselves. Be sure you are including all the relevant voices in this process and remember it may be in some people’s interests not to make certain information public.
* We cover triangulation in greater detail in session 4.8.

Presenting findings

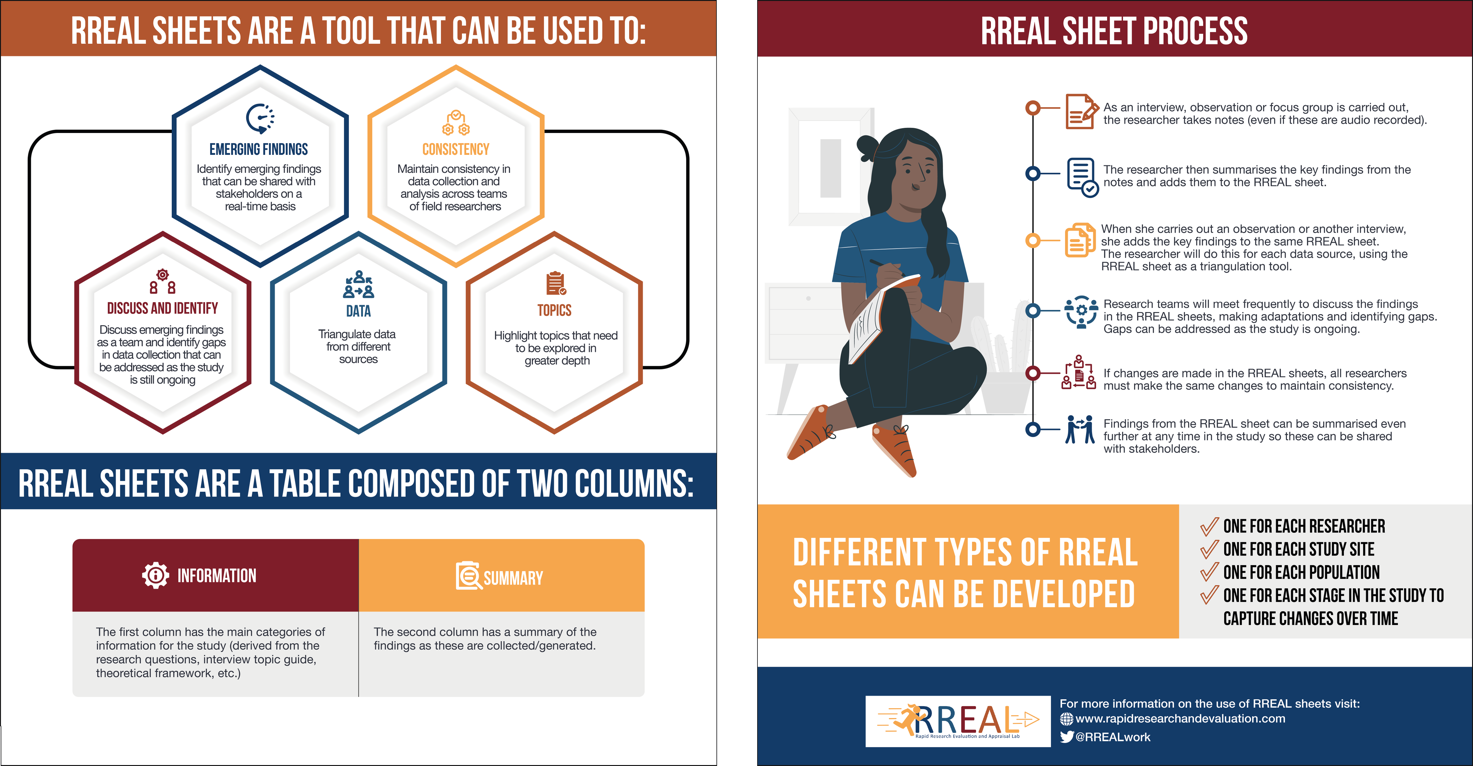
* When presenting findings, it is important to use direct quotes to illustrate the different ways responses were expressed. These should be reported anonymously.
* Be sure to consider and present findings in context – don’t use a quote just because it backs up what you want to say, whereas the participant intended it to have a different meaning.
* Make sure you make every effort to share the final product with the participants so that they can benefit from it too.
* Session 5.3 provides further information on effectively communicating findings and Session 5.4 covers how to feed findings back to communities to inform community-level action.

RAP sheets (10 minutes total)

Another emerging approach to more rapidly analyse qualitative data is the use of Rapid Assessment Procedure (RAP) sheets developed by the Rapid Research, Evaluation and Appraisal Lab (RREAL).

Play 2-minute animation available [here](https://www.rapidresearchandevaluation.com/resources).

Figure 8: RAP sheet overview



See **Handout 5** and the [RREAL website](https://www.rapidresearchandevaluation.com/resources) for more information.

Qualitative data analysis software overview (10 minutes total)

We have just covered steps to a thematic analysis not using software. Computer software packages that can analyse qualitative data are used increasingly by social scientists.

These packages include:

* [NVivo](https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home)
* [Dedoose](https://www.dedoose.com/)
* [MAXQDA](https://www.maxqda.com/)
* [ATLAS.ti](https://atlasti.com/)
* [QDA Miner Lite](https://provalisresearch.com/products/qualitative-data-analysis-software/freeware/)

For smaller sample sizes (e.g. less than 20) most people analyse ‘by hand’ as shown above, but for larger sample sizes software should be considered. Use of software may also be most useful when working with larger research teams to analyse and code data.

|  |  |
| --- | --- |
|  | Question to participants (5 minutes):  What might be the benefits of using software? What might be the downsides?  Online: Invite the participants to write the answers in the chat function and share it with everyone  Offline: Ask two or three participants to share their thoughts |

Some of the acknowledged pros and cons of using software include:

Pros:

* Less time-consuming.
* Easier to manage, organize, and move around when you have lots of data.
* Easier to work with multiple coders (if everyone has the software).
* Some packages can also manage photos, audio, videos, and spreadsheet data (e.g. Dedoose, ATLAS.ti).
* The use of software can be perceived as more credible, which can make findings more convincing.

Cons:

* Many of the better options are expensive (e.g. NVivo, MAXQDA, ATLAS.ti) although these have ‘free trials’ available. There are some more reasonably priced options (e.g. Dedoose) and some free options (e.g. QDA Miner Lite).
* You will need training on the specific packages to be able to use them, whereas most people are already familiar with Word and Excel.
* All coders would need access to the software, whereas they likely already have Word and Excel.

The facilitator can provide links to the different software we have mentioned to allow you to read more (see hyperlinks above).

Wrap-up/summary (5 minutes)

* There are many ways to analyse qualitative data, and thematic analysis is one of these. Thematic analysis is possible to do simply on Word or Excel software.
* Steps to thematic analysis include:
* Translation and transcription
* Coding – including decided whether to use inductive or deductive coding and developing a coding book
* Identifying and grouping themes from the data
* Reporting key findings
* Rapid Assessment Procedure (RAP) sheets developed by RREAL are another approach to rapidly analysing qualitative data as it is being generated.
* Computer software packages dedicated to analysing qualitative data are used increasingly by social scientists, and there are a number of pros and cons related to their use.

ACKNOWLEDGEMENTS:

Nadia Butler and Theresa Jones (Anthrologica) developed the content of this session.   
It was reviewed by Ginger Johnson (Collective Service).