

METHODS

This chapter contains:

1. Overview of basic concepts
2. Standard data collection methods
3. Cautionary note about data collection
4. Guiding questions to aid in selecting methods
5. Instrument development and testing
6. Disaggregated data
7. Data analysis
8. Unique peacebuilding tools
9. Sound basis for generalization
10. Record maintenance systems

INTRODUCTION

“All men by nature desire knowledge.”

-ARISTOTLE

This chapter introduces peacebuilding practitioners to the options and considerations for selecting the means of data collection for baselines, monitoring, and evaluation. In logical frameworks, methods are referred to as means of verification (MOV), while in other circles they are called research instruments. In this chapter, the term “research” will be used to mean any data collection done for baselines, monitoring or evaluation.

Data collection methods have been developed over a period of many decades and now have well-established standards and techniques. A comprehensive introduction to methods requires a manual or two of its own. For the purposes of the practitioner, one needs to understand the core concepts and terminology as well as when to use which method for the best results. The techniques of designing and implementing those methods are beyond the immediate needs of the average practitioner, hence the scope of this manual.

It should be made clear what this chapter is not. This chapter is not trying to make researchers out of practitioners. It does not have enough information on how to select, design, and implement methods nor does it cover how to analyze the resulting data such that a beginner could do so effectively. Instead, this chapter is intended to prepare a practitioner to have knowledgeable conversations with professional evaluators in order to make informed choices.

What are the basic concepts I need to know about data collection?

Mastering a few core concepts and their associated terminology is the first step in understanding data collection.

Key Terms

Fundamental to research is the notion that data and conclusions are two different things. Data is the building block of information and is often thought of as statistical or quantitative, although it may take many other forms, such as transcripts of interviews, maps, photographs or videotapes of social interactions. Conclusions are drawn from data through analysis.

Data collection methods have been developed over a period of many decades and now have well-established standards and techniques.

Fundamental to research is the notion that data and conclusions are two different things.

Methods are the means of acquiring the necessary data.

Methods are the means of acquiring the necessary data. Once gathered, the data is analyzed in order to generate conclusions.

There are a number of other terms used in research that are important to understand. These terms are listed below in the Methods Key Terminology table, and their accompanying definitions have been made informal and simple.

Methods Key Terminology

Bias	To be inclined toward a particular way of looking at or understanding something
Causality	The direct effect of one event on a future event
Correlation	The extent to which two or more things are related to one another
Disaggregate	To separate into component parts
Generalization	The extent to which one can come to broad conclusions about a group or phenomenon based on information gathered from a set of representatives of that group or phenomenon
Reliability	Do repeated applications of the method, even when different people apply it, result in the same outcome?
Sample	Representative members of the entire client base for the activity
Statistically Significant	Meaningful, measurable relationship or level of change
Unit of Analysis	The primary entity under evaluation, such as individuals, groups, artifacts, geographic units or social interactions
Validity	Does the method measure what it is supposed to?

Quantitative & Qualitative: Methods and Data

Quantitative and qualitative are important concepts for practitioners to understand. Quantitative methods are used to gather data to be analysed in numerical form. They pose questions of who, what, when, where, how much, how many, and how, generally in the form of surveys and questionnaires. These methods are designed to produce data that tells us how many people do or think something, and which is statistically reliable. Quantitative data typically is in numerical form such as averages, ratios or ranges.

Qualitative methods have greater flexibility and pose questions in a more open-ended manner. They give an in-depth understanding of why people hold particular views. They also explore how people make judgments, in a way that structured quantitative research cannot. Qualitative

Quantitative methods are used to gather data to be analysed in numerical form.

methods are not intended to be statistically reliable, but findings can — if participants (those who provide data to the study) are broadly representative — be strongly indicative of the population as a whole. Standard qualitative methods include interviews and focus groups. Qualitative data is typically words or text, though it can include photographs, video, or sound recordings.

The descriptions provided thus far cover quantitative and qualitative data and data collection methods. It is important to understand, however, that there are also quantitative and qualitative analysis approaches. See the question about analysis on page 50 for more information.

There has been an ongoing debate about which method — quantitative or qualitative — is better suited for baseline, monitoring, and evaluation purposes. Advocates for a quantitative approach argue that their data is hard, rigorous, credible, and scientific. On the other hand, proponents of the qualitative method contend that their data is sensitive, nuanced, detailed, and contextual. For peacebuilding baseline, monitoring, and evaluation purposes, this debate is needless because both approaches are necessary.

Mixed methods — the use of both qualitative and quantitative approaches — are becoming the new norm because they produce a richer set of information to meet the needs of conflict transformation projects. Numbers alone rarely answer the questions that peacebuilders have regarding why and how social change occurred. Perceptions and feelings that cannot be generalized to a greater population do not provide the complete picture either, necessitating the use of quantitative methods.

What are the standard data collection methods?

All of the standard social science methods, such as surveys or interviews, as well as participatory techniques like mapping, can be utilized in baseline, monitoring, and evaluation of conflict transformation. Each of these methods can be implemented individually or in combination with each other depending on the research needs. This chapter provides short overviews of some of the more commonly utilized methods and their strengths and weaknesses. How to choose the right method for research follows.

- **DIRECT OBSERVATION:** Watching, taking notes, and recording specific actions within a target community, such as communications, spatial interaction, or exclusion. The observation can be focused on a project process in which the people participating in the intervention are observed. It can also be focused on changes, such as in people's behaviors and attitudes, which involves watching people go about their daily business at home, in the community, or in the fields.

Quantitative methods are used to gather data to be analysed in numerical form.

Qualitative methods have greater flexibility and pose questions in a more open-ended manner. They give an in-depth understanding of why people hold particular views.

All of the standard social science methods can be utilized in baseline, monitoring, and evaluation of conflict transformation.

- **INTERVIEWS:** One-on-one contact with stakeholders, either in person or by telephone. These can be formal structured exercises, where a strict interview protocol is followed, or semi-structured meetings that are partially structured by a flexible interview guide. For comparability purposes, a minimum degree of commonality must exist in unstructured interviews.
- **FOCUS GROUPS:** Small-group conversations that seek to understand how people feel or think about an issue, product, service or idea. Focus groups have a specific purpose, size, composition, and process. They are best conducted with 6-8 people who are selected because they have something in common. Leading focus groups requires a skilled moderator and is best done in a comfortable, permissive environment. Such groups are a compromise between participant observation and more in-depth interviews.
- **PARTICIPANT DIARIES:** These are narrative descriptions of a personal experience. They may be open-ended to allow individuals to capture what was of importance to them each day or week, for example. Participant diaries can also be structured so that individuals take note of specific attitudes, events, behaviors that they have experienced in the allocated timeframe (e.g., daily, weekly).
- **PHOTOGRAPHY/VIDEO:** Utilizing photographs or video to collect visually represented information
- **PROJECT DOCUMENT REVIEW:** Collecting, reading, collating, and analyzing key documents such as proposals, donor reports, annual reports, case studies, etc.
- **QUESTIONNAIRE:** A set of specific, targeted questions to which stakeholders respond in writing. The questions must reflect cultural awareness and be language sensitive in addition to fitting within a set of formal methodological standards. Questionnaires can be distributed electronically, by post or by hand. For a statistically significant conclusion, the number of responses needed depends on the total population size. However, to do statistical applications like developing the mean or plotting charts, one needs a minimum of 30 responses for answers to be valid.
- **SECONDARY DATA REVIEW:** An examination of existing data. This type of review is often the initial inquiry that precedes data collection with stakeholders. It is also called a desk review. Sources include academic theses, annual reports, independent studies by NGOs or researchers, and census data.
- **SURVEY:** A sequence of focused, targeted questions posed to stakeholders in a fixed order by a surveyor. As with questionnaires, survey questions must reflect cultural awareness and be

language sensitive in addition to fitting within a set of formal methodological standards. Surveys are generally utilized for large-scale efforts, though they may be used on a smaller scale. To reach a statistically significant conclusion, the number of responses needed depends on the total population size. However, statistical applications like developing the mean or plotting charts one require a minimum of 30 responses.

- **TESTING:** This is usually a series of questions or exercises – oral or written – for measuring the skills, knowledge, capacities, or aptitudes of an individual or group. Testing is generally used before and after training as a way to measure change.
- **PARTICIPATORY LEARNING AND ACTION TECHNIQUES (PLA)**²⁷: The application of Participatory Rural Appraisal techniques in a mutual learning process utilized on broader issues than the original rural development focus. There are many techniques within PLA that are useful for baseline, monitoring, and evaluation purposes such as:
 - **Venn diagrams:** These are made with circular cards of different sizes and colors placed in relation to one another with each card representing an issue. The size of the card represents the issue's importance to the conflict, with a larger card indicating greater importance, and the degree of overlap between cards represents the intensity of interaction of those issues. Men and women, wealthy and poor, young and old, may well produce different diagrams whose differences are often instructive.²⁸
 - **Pairwise ranking:** This technique helps to determine the relative importance of various options. The participants compare only two options at a time, and the reasons for preferring one option over the other are made clear. Going through all the possible combinations finally results in a list of criteria by which villagers can assess options.
 - **Conflict Mapping:** This is a technique used to represent a conflict graphically by placing the parties in relation both to the problem and to each other. When people with different viewpoints map their situation together, they learn about each other's experience and perceptions, and their differences and commonalities become clear.
 - **Drawing:** This technique is often called mapping as well. It is a visual depiction, generally in pictorial form, of the focus of the discussion. This can be a geographic map, an emotion, or a situation.

²⁷ Adapted from Participatory Rural Appraisal (PRA), part of the "Working instruments for Planning, Evaluation, Monitoring and Transference into Action" series (Swiss Agency for Development and Cooperation [SDC], Strategic Unit, January 1997)

²⁸ Adapted from Participatory Research for Sustainable Livelihoods: A Guide for Field Projects on Adaptive Strategies. <http://www.iisd.org/casl/CASLGuide/RepRel.htm>.

- **Role Playing:** Taking on a role enables people to creatively remove themselves from their usual roles and perspectives to portray a situation.

Since every context and application of a method can vary, it is difficult to give definitive guidance on what method to use in every situation. Some of the more common strengths and weaknesses of each method are described in the table below. Because cost is almost always a variable in selecting methods, a separate column has been added to illustrate if the method is of high, low, or average expense. Of course, this is also a question of scale: the larger the scale, the higher the cost. Thus, the cost column depicts comparative costs by assuming the same project scale is used in each method listed.

Selecting the Most Appropriate Methods

Method	Strengths	Weaknesses	Cost
Direct Observation	<ul style="list-style-type: none"> • Minimal preparation required • May enable the experience of minorities or women to be captured, particularly in situations where speaking out against the norm is dangerous 	<ul style="list-style-type: none"> • Must be done at the right moment in the right place • Does not provide information on why things occur • Presence of the observer may influence behaviors 	Low
Focus Groups	<ul style="list-style-type: none"> • Can identify issues that need probing through another method • Allows one to observe various perceptions on an issue • Enables more people to be involved in less time than individual interviews 	<ul style="list-style-type: none"> • Difficult to manage multiple opinions • “Group think” may occur • Individuals may not feel comfortable to dissent 	Low/ Average
Interviews	<ul style="list-style-type: none"> • Good for small numbers • Allows for exploration into how and why • Generates data on needs, expectations, attitudes, perceptions, beliefs, and feelings 	<ul style="list-style-type: none"> • Time-consuming • May be difficult to differentiate between those who are telling you what they think you wish to hear from those telling the truth 	Low/ Average

Selecting the Most Appropriate Methods

Method	Strengths	Weaknesses	Cost
Participatory Learning and Action Techniques	<ul style="list-style-type: none"> • Can offset the biases of the evaluator • Empowers the local people because their views are taken seriously • Excellent methods for working with illiterate communities • Useful if the purpose is to determine whether needs are being addressed by the evaluation (Appropriateness) 	<ul style="list-style-type: none"> • No hard quantified data is produced • For comparative work replication is difficult as each situation has its own unique situation 	Low/ Average
Photography/ Video	<ul style="list-style-type: none"> • Less open to interpretation • Easily disseminated • Can be a rapid technique • Easily led and completed by participants • Easily preserved 	<ul style="list-style-type: none"> • Must be done at the right moment in the right place • Can be one-dimensional information that does not explain how or why 	Low/ Average
Project Document Review	<ul style="list-style-type: none"> • A low-cost way of learning the history and background of a project • Provides insight into the perceptions of the practitioners 	<ul style="list-style-type: none"> • May be limited in the degree of detail • May be tailored to donor or other needs and requirements but omit key information • May have gaps in time that reports do not cover 	Low
Questionnaire (by post or e-mail)	<ul style="list-style-type: none"> • Good if the intervention affects large numbers of people • Good if statistically significant results are needed 	<ul style="list-style-type: none"> • Requires literacy • Time-consuming • Is difficult to utilize in contexts with multiple languages • Requires a distribution system (e.g., postal system, Internet) for large numbers of people 	High
Secondary data review	<ul style="list-style-type: none"> • Fast means of gathering key background information • Offers a variety of perspectives and insights • Can save the evaluation team time since they do not need to collect the data 	<ul style="list-style-type: none"> • May not be tailored to the needs of the project • Data may be flawed 	Low

Selecting the Most Appropriate Methods

Method	Strengths	Weaknesses	Cost
Surveys	<ul style="list-style-type: none"> • Good for interventions that affect large numbers • No literacy requirements for respondents • Good if statistical comparisons are required 	<ul style="list-style-type: none"> • If not well trained, surveyor bias may affect responses • Requires greater resources than a questionnaire does • Does not explain how or why something happened • Not very good if the purpose of the assessment is to study complex processes • Time-consuming 	High
Tests	<ul style="list-style-type: none"> • Good for knowledge acquisition; less reliable for skills 	<ul style="list-style-type: none"> • Does not reveal whether or not the new knowledge or skills will be retained or applied in the future • May require literacy 	Low

Data collection is based on standards of practice that are essential to the quality and, hence, the credibility of the data collected.

Data collection is based on standards of practice that are essential to the quality and, hence, the credibility of the data collected. The manner in which questions are asked and formulated, the behavior of interviewers, and identity of the interviewer (in terms of gender, nationality, race, etc.) can influence responses. The most common error made by novices is to allow their personal bias to influence the situation. The type of questions researchers ask can introduce bias as can the choice of who they talk to and when data collection is conducted. In addition, the way that data is analyzed or presented can introduce bias.

When considering how bias can affect the types and form of questions, consider, for example, a project working with French-speaking citizens of Quebec on changing their perspective from that of seeking political separation to remaining part of Canada. One of the attitudes to be changed is animosity toward the federal government. If the baseline asked, “How much do you dislike the federal government?”, there is clearly a negative bias that assumes all francophones dislike the federal government and it is simply a matter of how much. A very different response would be expected if the question were phrased as follows: “Tell me about Quebec’s relationship with the federal government.”

The manner in which questions are asked can also affect the answers given. This becomes particularly important when the evaluator and those giving information are from different cultures. Moving to the other end of Canada, the Blackfoot nation is one of the First Nations (indigenous populations) of Canada living in Alberta. If an evaluator were collecting data from the Blackfoot people, she/he would need to understand their use of silence and not rush to fill it with more questions or answers. The

evaluator who does not allow for silence and the processes of reflection within that community will return with data that is superficial at best and probably significantly flawed.

Finally, who asks the questions also influences responses. The attributes of the questioner are all noticed by the respondent and subtly interplay with the answers given. Formal education, use of special terminology, clothing, being in a paid position, being perceived to be in a position of authority, and being perceived to be sympathetic to the other side due to one's nationality are all examples of the things that may affect the responses received. Consider an older, well-dressed, male French professor going into the immigrant communities of the Paris suburbs following the November 2005 riots to gather baseline data for a project involving disaffected immigrant youth who participated in the violence. The differences between the data collector and the respondents are not fatal to the project; those differences simply need to be taken into account when designing the data collection process and designating who should collect the information.

Ways to minimize bias and errors in data collection include the careful training of researchers, setting of objectives and indicators, and the triangulation of information.

Ways to minimize bias and errors in data collection include the careful training of researchers, setting of objectives and indicators, and the triangulation of information. Because the field of social science research is well-established, there are many ways to obtain training in methods implementation ranging from university degrees in the subject to short courses and practical trainings.



How are methods selected?

Having a clear understanding of what the research exercise (e.g., baseline, monitoring or evaluation) is to explore how the resulting conclusions are to be utilized is essential to good method selection. The information required should drive the selection of methods. Use the following questions to further guide the method choice.

- **HOW COMPLEX IS THE PROJECT?** If the project is extremely complex, qualitative methods are likely to be better suited to handling its intricacies than other methods. As project complexity increases, so too does the need for triangulation.
- **WHAT IS ALREADY KNOWN ABOUT THE PROJECT?** The more data available, the more the evaluators can focus on gathering information to fill the gaps. This may be exploring the why behind a fact in which case qualitative means. On the other hand, it could be that generalized conclusions are missing; hence, quantitative means would be better suited.

The information required should drive the selection of methods.

Determining if a certain design tool was used to plan the project and whether or not the project was launched with clear assumptions and an articulated theory of change provides a sound starting point for an evaluation team. If this information is not available, time should be allotted to finding it through project document review and possibly participatory learning and action techniques with the project team.

- **WHAT INFORMATION IS NEEDED?** Does the team need information on attitudes, knowledge, perceptions, behaviors or skills? If data on knowledge and skills is required, testing might be a good approach. If behaviors are the focus, direct observation or interviews may be a good choice. If it is important for the answers to be generalized, quantitative methods are preferable.
- **HOW MUCH TIME IS AVAILABLE?** Some methods require more time to design, implement, and analyze than others. A large questionnaire, for example, takes far more time than do focus groups. A questionnaire needs to be developed, tested, redesigned, and distributed to respondents, who must be given time to fill it in. Once the questionnaires are returned, the data is generally entered into a database and then analyzed. All of these tasks not only require enough days for each to be performed, but also a sufficiently lengthy evaluation duration to allow for the data collection method to be completed. For information on time and duration estimates, see page 138 of the Evaluation Management chapter. An evaluation for which data collection must be completed in two weeks would not permit enough time for the development of a questionnaire.
- **WHAT IS THE EVALUATION APPROACH?** Although most methods can be applied within most approaches, the approach selected may lend itself better to one method over another. If the approach is self-evaluation and no one on the team has survey or questionnaire expertise, that is probably not the best option. If utilization-focused evaluation is the approach and the project team wants results that can be generalized, surveys or questionnaires would be a good method choice. However, the use of a different utilization-

focused evaluation may be desirable in order to explore why some people drop out of a program and this would best be accomplished through interviews or focus groups.

- **WHAT IS THE EVALUATION SCOPE?** The scope of the evaluation is an important factor in methods selection. If the evaluation is to cover a large number of people, questionnaires and surveys should be considered. For a medium-sized group, participatory learning and action techniques may be a good choice. If the geographic scope is large or the conclusions need to be generalized, time-intensive methods such as interviewing and direct observation would not be the best options.
- **HOW DIFFICULT WILL IT BE TO ACCESS THE DATA?** Also called data availability, the degree of difficulty involved in accessing information is always a consideration in method selection. Are the data sources out in the bush fighting a war and therefore hard to access or are they based in the city centre and accessible through the Internet? The latter may offer the opportunity to use an Internet-based questionnaire tool, while evaluation of the former would require direct observation, interviews or potentially focus groups. Do those you seek to engage speak many different languages? If so, questionnaires and surveys require translation and testing in each language, which will have cost and time implications.
- **WHEN IS THE INFORMATION NEEDED?** If there is a restricted timeframe for the evaluation, rapid methods like secondary data review, project document review or Participatory Learning and Action Techniques may be helpful. If a longer time period is available, more time-intensive methods can be used, such as surveys.
- **WHAT LEVEL OF RELIABILITY IS REQUIRED?** If a high degree of reliability is essential, Participatory Learning and Action Techniques are often not the best options, whereas questionnaires and surveys may be a better choice. The importance of method reliability increases with the emphasis on quantitative results and the ability to generalize to populations.
- **WHAT IS THE AVAILABLE BUDGET?** As outlined in the Selecting the Most Appropriate Methods table on page 207, different methods have different cost structures associated with them. Direct observation and project document review are very low in cost, while questionnaires and surveys can be expensive.
- **WHAT IS THE CAPACITY OF THE DATA COLLECTORS?** The difficulty of the method should be considered in relation to the capacity of the data collectors. If project staff members are gathering the data, selection of a simple and easily applied method is important to the quality of the data collected. Try to build on the skills

that the staff already have acquired through their project implementation work by choosing, for example, such methods as conflict mapping or photography/video. In some cases, the method may appear simple, but actually gathering data that contributes to sound decisionmaking may be more complex. The greater the degree of complexity required by the method (also called the method difficulty), the greater the level of knowledge and skills required to implement it effectively.

Triangulation is simply using different methods to research the same issue and then analyzing all of the results.

- **WOULD TRIANGULATION MAKE THE CONCLUSIONS MORE RELIABLE?** Triangulation is simply using different methods to research the same issue and then analyzing all of the results. For instance, if examining the results of a project that sought to change laws on citizenship in Ivory Coast, one could conduct in-depth interviews with members of the government, use conflict mapping with groups of non-citizens in their communities, and gather feedback forms from events. Here, the evaluation is cross-checking one result against another (i.e., triangulating), which increases the reliability of the conclusions.

Triangulation is useful in many ways. Contradictory results produced through different methods often indicate important problems with question design and/or fundamental issues surrounding the researcher's understanding of a topic. Triangulation is essential when using Participatory Action and Learning Techniques and helpful when the researcher, in exploring sensitive issues, is uncertain if the data source is able or willing to provide the full story.

The conflict and its volatility affect the choice of method.

- **ARE THERE CONTEXTUAL ISSUES SUCH AS CULTURE OR LEVEL OF DEVELOPMENT THAT WOULD MAKE SOME METHODS BETTER THAN OTHERS?** The context should always be considered in method selection. If women will not speak openly in front of men or youth cannot speak their mind in front of elders, then focus groups or surveys conducted in open environments may not be the right choice. Individual interviews or questionnaires may allow those who cannot speak their mind in all situations better opportunity to express their views.
- **ARE THERE CONFLICT ISSUES THAT MAKE SOME METHODS BETTER THAN OTHERS?** The conflict and its volatility affect the choice of method. There may be conflict situations where someone is unable to state a dissenting opinion without putting themselves in danger. In such a case, methods should be selected that ensure the anonymity of sources by not requiring the disclosure of names or documentation. Any method that requires experiences to be documented, through the use of participant diaries or photographs, for example, deserves extra consideration in conflict contexts. If discovered by the wrong people, such as a paramilitary group or the army, these participants and sources might be in danger.

Take, for instance, three selection criteria mentioned earlier – data availability, method difficulty and method reliability – and compare their implementation in stable versus volatile environments. Recall the definitions for each:

- **Data availability** reflects the ease or difficulty of obtaining data. High data availability means that it is relatively easy to access the data from the source. Low data availability would mean that is difficult to obtain the information needed.
- **Method difficulty** considers the complexity of appropriately developing and implementing the method. High method difficulty implies that there is some intricacy involved in developing and implementing the instrument, while low method difficulty means that it is relatively simple.
- **Method reliability** refers to the degree to which the method produces the same results when used by different people. High method reliability means that the instrument can be used many times and the same responses will be generated. Low reliability means that, if different people utilized the method, the answers would likely be different.

In the table on page 215, the left half of each column represents a stable environment in which a non-sensitive issue is discussed. The right half of each column represents a highly contentious conflict situation dealing with a sensitive issue.

For example, consider a project in the Ukraine that seeks to decrease negative attitudes of Ukrainians toward Russian nationals living in the Ukraine. Using focus groups would provide high data availability since accessing this information would not be difficult. In comparison, consider a similar project in Iraq in 2005 that seeks to decrease negative attitudes that Sunnis hold toward Americans. The use of focus groups would likely have low data availability because Sunni respondents would be fearful of speaking out against the accepted norm of “hatred” in front of their community.

Data Availability, Method Difficulty and Reliability²⁹

Methodology	Data Availability		Method Difficulty		Method Reliability	
	<i>Stable</i>	<i>Volatile</i>	<i>Stable</i>	<i>Volatile</i>	<i>Stable</i>	<i>Volatile</i>
Direct Observation	High	High	Low	Low	High	High
Focus Group	Average	Low	Average	High	Average	Average
Interview	Average	Average	Average	High	Average	Average
Participatory Learning & Action Techniques	Average	Low	Average	Average	Low	Low
Photography/Video	High	High	Low	Average	High	High
Project Document Review	High	High	Low	Low	High	High
Secondary Data Review	High	Low	Low	Low	High	High
Survey/Questionnaires	Low	Low	High	High	High	Average
Tests	Average	Average	Average	Average	Average	Average

What do I need to know about instrument development and testing?

Once the methods have been selected, the next step is to design the instrument to be used, such as the questionnaire or the interview protocol. The methodological standards for instrument development are based on a well-researched body of literature, and it is critical to carefully follow those standards. The result of not using these standards will be flawed or biased instruments that produce unreliable data. That flawed data will then create false conclusions. Since the conclusions of an evaluation are what inform program decisionmaking, the consequences of using flawed instruments can have significant negative effects on the project and the people it is meant to assist.

At this point in the process, the average practitioner should seek technical assistance if developing the instrument on her/his own or assign it as one of the tasks for the external consultant.

Once the instruments have been designed, they must be tested. Both qualitative and quantitative methods of data collection require validity and reliability tests. These check for clarity, accuracy and whether the

²⁹ Adapted from *Managing for Evaluations: A Manual for PVOs*. PVC Grants Managers and Evaluation Team Members by A. Willard (USAID/PVC/Matching Grants, 1998).

tool is appropriate to the data needs. In essence, you want to know whether or not the instrument accesses the type of data required in the way that is intended. This step is commonly omitted in peacebuilding evaluation to the detriment of the quality of results. When external consultants are hired, they should be told that they will be expected to perform validity and reliability testing and this expectation should be included in the terms of reference (see page 138 for more information on terms of reference development.) If instruments are not tested and refined, the quality of the resulting data will be compromised.

Once the instruments have been designed, they must be tested.

Testing can be as elaborate or as minimal as is needed by the project at hand. Assuming that your project staff members are broadly representative of both the local population and the conflict, a cheap and easy strategy is to use the project staff as subjects for the test. In other words, try the instrument on project staff, by having them fill out the questionnaire or respond to an interview. This will not only enable the evaluators to have input into refining their instrument, but it will also provide the project team with insight into what is being asked and how. Of course, if the instruments are intended to gather data from Israelis and Palestinians but the staff members of the evaluating organization are all Israeli, this would not be an effective test. Similarly, if a consensus-building project in Washington, DC worked primarily with community members who have limited educational backgrounds, testing the instruments on highly educated staff would not be appropriate. Always test the instruments in the language and culture in which they will be used.

Testing can be as elaborate or as minimal as is needed by the project at hand.

Data can be disaggregated by many different factors or subgroups.

What is disaggregated data?

In thinking about the data to be sought from the baseline, monitoring or evaluation, consider whether the conclusions would be more useful if they were broken down according to different factors such as ethnicity. These factors correspond to key groupings within the project's target population. This is called disaggregated data.

It is also important to consider the state of the conflict and the sensitivity of the factor when making these decisions.

Take, for instance, a dialogue project in Northern Ireland that brings together Protestants and Catholics. Attitudinal information that represents all participants gathered through interviews will be informative, but it could hide important differences between the two communities. Potentially, if the data was broken out by Catholic participants and Protestant participants, it could reveal differences in attitudes that would be key to informing future decisions about the project.

Data can be disaggregated by many different factors or subgroups. The importance of each factor is dependent upon the focus of the intervention being explored and the evaluation objectives and lines of inquiry being chosen. Disaggregating the data within a line of inquiry should provide more useful information than if it were not disaggregated.

If the information is not more useful when disaggregated, it is not worth doing it. Common factors to disaggregate against for conflict transformation projects include:

- Gender
- Residence
- Ethnicity
- Religion
- Age
- Tribe
- Ex-combatant
- IDP or Refugee

Remember that, as the number of factors increases, so too does the amount of data needed and the amount of time required to analyze it, which affects the cost of the research.

Like instrument development and implementation, data analysis is subject to strict standards of practice.

It is also important to consider the state of the conflict and the sensitivity of the factor when making these decisions. If asking someone about her/his religion or ethnicity is highly sensitive or even, in some instances, dangerous, there might be proxies that can be used to substitute for such questions as a way to avoid endangering the respondents. For instance, in some conflicts, neighborhoods are strictly segregated, so that asking where someone lives could substitute for ethnicity or tribe. The project team plays a key role in alerting external professionals to these sensitivities and to potential proxies.

Quantitative analysis can also be used on qualitative data.

Once the factors have been chosen, the appropriate questions need to be included in the research instrument in order to gather the data. For instance, if the gender and religion of participants were important to the project, these two questions would be added to a questionnaire or interview as factors to be collated. Or, if a feedback form is used to monitor a workshop, these two factors (gender and religion) would be added to the form.

How is data analysis done?

Like instrument development and implementation, data analysis is subject to strict standards of practice. Even though high-quality data collection instruments have been developed and tested appropriately, and the process of data collection may have followed all the best practices, it all will be for naught if the proper analysis techniques are not utilized.

Quantitative data analysis is usually called statistics. Generally speaking, quantitative data is processed through statistical computer software packages. The most popular of these is the Statistical Package for Social Sciences (SPSS); however, the analysis software could be as simple as an Excel spreadsheet or an Access database. With the proper safeguards for confidentiality, processing quantitative data in a public or transparent way helps build credibility in communities where people distrust the process and/or technology.

There are a number of approaches to qualitative data analysis, although none have the same precision in the rules as statistical analysis. The abundance of approaches is an asset because any set of qualitative data can be approached from a variety of different perspectives. In other words, different techniques can be applied to the same data, which may highlight new aspects of that information. Hence, the approach taken in the analysis affects the conclusions drawn.

Quantitative analysis can also be used on qualitative data. This means that qualitative data, such as transcripts of interviews, can be analyzed to produce statistical conclusions.

Identifying the means of analysis is often forgotten when discussing baselines, monitoring, and evaluation, yet it is a central part. Peacebuilding practitioners should question evaluators about their chosen means of analysis and why they believe it is the best option. For all forms of research, the author needs to be able to explain how she/he arrived at the conclusions from the data. In other words, the means of analysis must be described and scrutinized.

Analysis should not be the exclusive domain of the evaluator. The people under scrutiny in an evaluation frequently offer very insightful analysis and, at times, bring out dimensions that only they can perceive. Whoever participates in the analysis needs to be aware of her/his own biases and assumptions.

Are there unique peacebuilding tools for data collection?

In many professional fields, such as education, basic social science data collection methods are blended together or refined to meet specific needs of that field. When a standard data collection method is refined or blended with core features of a field it is called a tool. These “tools” often become the accepted techniques for baseline, monitoring, and evaluation in that field.

Few tools have been developed specifically for conflict transformation, although there are some that were developed for other fields that could be adapted. None of the tools in this latter group have taken root in the peacebuilding field, none have been deemed more or less effective than others, nor is there a norm or standard in terms of application.

Selecting a tool is not essential for an evaluation to be effective. Evaluators and project teams should consider what information is being sought and then select the best way of obtaining it. That selection may include the use of straightforward data collection methods, such as surveys or interviews, or there may be another tool that is more effective in obtaining the data.

There are a number of approaches to qualitative data analysis, although none have the same precision in the rules as statistical analysis.

When a standard data collection method is refined or blended with core features of a field it is called a tool.

Each tool needs to be adapted to the context and purpose of the intervention for which it is being used to evaluate and, as with methods, tools can be combined. Regardless of the tool used, it is necessary to have an understanding and ability to implement the data collection methods from which they are developed.

As of 2005, the Office of Transitional Initiatives (OTI/USAID) is developing a monitoring and evaluation toolkit for transition and conflict transformation projects.³⁰ A sample of the tools in this toolkit, along with several others, is listed below. Note that this is not a comprehensive list of every tool that has been developed – others are available.

ACTIVITY INTERVIEW³¹: An activity interview seeks to identify the views participants hold about an intervention’s activity, such as a Good Neighborliness Seminar, regarding the effects of that activity and/or to obtain process feedback. This input is added to the observations and opinions of the staff responsible and compiled in a short report. Using a semi-structured format, activity interviews take place a few weeks after the activity. Not all participants need to be interviewed, nor does it need to occur after each activity; rather, a sample of each may be selected. This tool would most commonly be used in monitoring.

COGNITIVE SOCIAL CAPITAL ASSESSMENT TOOL (CSCA): CSCA is “a quantitative method for collecting basic information about cognitive social capital quickly. Cognitive social capital refers to people’s perceptions of the trustworthiness of other people and key institutions that shape their lives, as well as the norms of cooperation and reciprocity that underlie attempts to work together to solve problems.”³² CSCA uses a questionnaire, which can also be utilized as a survey. It can be implemented in small- to large-scale applications. The quantitative nature of the tool may fail to capture some of the complexities of social change. This tool is most appropriate for a baseline and evaluation.

MEDIA CONTENT ANALYSIS TOOL: This tool allows project managers to “evaluate media coverage, placement of stories, tone, and visual images, prominence of quotes/personalization, and reach of a media outlet. It can be used to track how different media cover topics such as conflict, human rights, reintegration of ex-combatants, and local governance reform.”³³

To utilize this tool, the evaluator identifies the media outlets to include, determines their circulation/reach, and then scores each against seven measures: prominence, headline, visuals, quotes, tone, column inches, and political ideology. The score for each outlet is then multiplied by

³⁰ At the time of this manual’s publication, the *OTI Performance Monitoring and Evaluation Toolkit, Fast Learning for Program Improvement, Local Empowerment and Results* [hereinafter *OTI Toolkit*], is not yet available publicly. Inquiries should be made to Social Impact at <http://www.socialimpact.com/>.

³¹ Adapted from *Strategic and Responsive Evaluation of Peacebuilding: Towards a Learning Model, Report of the Second Action-Reflection Seminar Convened by NPI-Africa and the NCCK-CPBD Project*, March 2001.

³² *OTI Toolkit*, forthcoming, pp. 146.

³³ *OTI Toolkit*, forthcoming, pp. 111.

the outlet's individual ranking in terms of circulation/reach. The scoring measures may need to be adapted to those issues of importance to conflict transformation. This tool is best used for baseline and evaluation studies, though it could be modified to contribute to monitoring efforts.

CASE STUDY: This tool investigates a contemporary event within its real-life context. Case studies are a way to learn from past experience since they explore how something happened. A case study results in a report that contains a rich narrative of the phenomenon detailing how it came about.³⁴ It is based on a particular worksite the boundaries of which need to be clearly defined to allow the study to be focused. If the case is to be illustrative of the wider context, selection of a site that is not unique is important.

Case studies rely on multiple data sources because the data needs to converge in a triangulating fashion. A case study involves the use of a variety of data collection methods, predominately interviews and project document review, but can also include direct observation or focus groups. Case studies can be done on a rolling basis to monitor the changes that occur over time or as part of an evaluation both of which result in brief, reflective snapshots of complex and dynamic situations.

CAPACITY ENHANCEMENT NEEDS ASSESSMENT (CENA): CENA "is a participatory assessment [tool] designed to evaluate existing capacity within key community and local government stakeholder groups, identify capacity gaps and weaknesses and recommend possible remedies."³⁵ Based on interviews and focus groups, information from a CENA is then plotted against each indicator on a scorecard. The tool was created for community-based development and, as such, the indicators would need to be adapted for conflict transformation programming. This tool would be best used in baseline and evaluation studies, though a streamlined version might be possible for monitoring efforts.

FOUR LEVELS OF TRAINING EVALUATION³⁶: The four levels approach - reaction, learning, transfer and results - is a systematic way to assess the quality and outcome/impact of training. Information from the previous level serves as a base for the next level as one works through all four. Reactions (level one) should never be the only level utilized and can often be blended with learning (level two). Generally, reaction (level one) utilizes a questionnaire and learning (level two) involves a pre- and post-test, while transfer (level three) can either use interviews or surveys several months after the training has taken place. Levels one and two should be included in monitoring systems while levels three and four might be a monitoring or evaluation tool.

³⁴ Robert Yin, *Case Study Research: Design and Methods*, Third Edition, Thousand Oaks, Sage Publication, 2003.

³⁵ *OTI Toolkit*, forthcoming, pp. 55.

³⁶ Adapted from the *OTI Toolkit*, forthcoming.

Can I draw a conclusion for the entire population from this data?

How to draw conclusions for an entire population, be it a community, tribe, or geographic area, is beyond the scope of this introductory manual. Conclusions are drawn using quantitative methods and should be based upon “enough” of a representative or random sample of the population. What constitutes “enough” requires a calculation based on established practices.

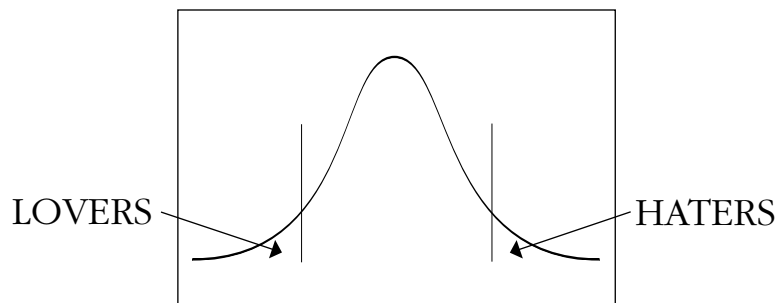
In other words, they are not representative of the average – they are lovers or haters.

As practitioners who will be gathering data through monitoring practices, however, there is one related concept that is important to understand. It can be called informally *Lovers, Haters, and Everybody Else*. As depicted below, the bell curve represents the average population of people, with the majority located in the middle and the extremes located at either end. Extreme means not being representative of the average or norm within the group. Although it appears that the extremes are the same size at either end, this is not necessarily the case.

Lovers, Haters, and Everybody Else relates to conflict transformation monitoring because it explains the implications of basing our monitoring data collection only on participant-initiated measures such as personal narratives, testimonials, fan mail, individuals who call into radio shows, or informal comments from participants. As a general rule of thumb, individuals who are willing to step forward and to take their own time to volunteer information fall at either end of the bell curve. They either love what is being done and want to praise it so it continues or they hate what is happening and feel compelled to intervene to stop it. In other words, they are not representative of the average – they are lovers or haters.

Classifying such individuals as lovers or haters does not discount their input. What it does tell us is what those two ends of the spectrum think about the work. We cannot take that work to the next step and assume that our information is representative of everyone involved.

The Lovers and Haters and Everybody Else



What record maintenance systems are necessary for collected data?

Record maintenance for baseline, monitoring, and evaluation data is quite straight-forward. For baselines and evaluations, the raw data collected, such as completed questionnaires or interview notes, should be kept in its hard-copy form. Future evaluators and researchers may wish to go back to the original data and do a different analysis or verify the previous one. If the data is sensitive or confidential, it should be stored appropriately, either in a locked filing cabinet or an inaccessible room. How long raw data is kept is dependent upon the project, organization, and potential future uses. If a formative evaluation was performed and a summative evaluation will occur later, the records should be saved until the summative evaluation is done.

Evaluators often do not return raw data to the project team unless it is requested. Particularly in the case of baseline and formative evaluation, retaining the raw data should be part of an organization's good practice norms.

For the material itself, one can use well-labeled filing cabinets, computer floppy disks, or CD-ROMs to preserve the information. Generally, it is good practice to keep a back up file of all electronic documents. If the data was processed using Excel spreadsheets or SPSS, keep those electronic files because it will save data entry time for future efforts.

When considering record maintenance systems for monitoring data, simple is always better. If most project staff members are comfortable with using an Excel spreadsheet, select that as the medium for data analysis rather than a statistical package that requires a specialist to operate.

What are the ethical obligations of feeding back the results to the people involved?

When people hear nothing about a study or evaluation in which they participated, they are less inclined to contribute to future data collection efforts. This is even more the case in highly participatory exercises where people feel they have an investment in the outcome of the research. Even the briefest feedback on the general findings and the use of those findings is often appreciated very much.

Future evaluators and researchers may wish to go back to the original data and do a different analysis or verify the previous one.

If the data is sensitive or confidential, it should be stored appropriately, either in a locked filing cabinet or an inaccessible room.

Further Reading:

Chronic Poverty Research Centre Methods Toolbox

<http://www.chronicpoverty.org/CPToolbox/toolboxcontents.htm>

William M. Trochim, *The Research Methods Knowledge Base*, Atomic Dog Publishing, 2001. <http://www.socialresearchmethods.net/kb/>

Keith Punch, *Introduction to Social Research Quantitative and Qualitative Approaches*, SAGE Publications, 2000.