place—something to measure against in the future. A baseline will allow you to set reasonable targets and measure the FBCOs' progress against them. For organizations that have not collected data on indicators in the past, the first round of data collection will establish the baseline.

Similarly, comparisons are useful when you want to measure your performance against something—either another time period, another organization, or an established set of standards. Although there are comparative standards in some fields (e.g., health—where infant mortality rates and other morbidity rates have been established nationally), there are few standards in human services that organizations can use for comparison purposes. Some national human service organizations are beginning to establish standards specific to their own programs. More likely, however, you will use your previous performance as a standard, comparing your success last year to this year, or this year to next year.

At the outset, it's important to stick to the basics of outcome measurement—defining outcomes and the indicators that make them measurable. You'll have time to move on to targets and comparisons later. A worksheet to assist you with identifying indicators is included in the appendices.

**PERFORMANCE INDICATORS CHECKLIST**

- ❑ Are the outcomes related to the "core business" of your program?
- ❑ Do your indicators make sense in relation to the outcomes they are intended to measure?
- ❑ Are your indicators directly related to the outcome? Do they define the outcome?
- ❑ Are your indicators specific?
- ❑ Are your indicators measurable or observable? Can they be seen (i.e., observed behavior), heard (i.e., participant interview), or read (i.e., client records)?
- ❑ Is it reasonable that you can collect data on the indicators?
- ❑ Is it likely within your resources to collect data?

## PHASE 2: CREATE AND IMPLEMENT A DATA COLLECTION PLAN

This section discusses how to plan for and implement the collection of data. You will consider what data collection methods are most appropriate and get the best data, and how to manage and ensure the integrity of the data you collect.

### Data Collection Methods

Your data collection will include attention to all the elements of your logic model—what resources you had available, what activities you actually provided, how many of each output you delivered, and to what degree you accomplished your outcomes. For all of these elements, except the outcome and indicator data, your organization's records should provide the information you need. However, in collecting indicator data, you have other options. You are likely to use one or more of four methods for collecting data—surveys, interviews, observation, and record or document review. In selecting the best method for data collection, you will need to consider the advantages and disadvantages regarding:

- Type of information needed — some outcomes and indicators lend themselves to one particular method; in other cases, data could be collected in more than one way
- **Validity** and **reliability** — some methods generate more accurate and consistent information than others
- Resources available —including staff availability and expertise, time, and money
- Cultural appropriateness — using the broadest definition of culture to ensure that the methods fit language, norms, and values of the individuals and groups from whom you are collecting data

**Surveys**

Surveys are standardized written instruments that can be administered by mail, e-mail, or in person. The primary advantage of surveys is their cost in relation to the amount of data you can collect. Surveying generally is considered efficient because you can include large numbers of people at a relatively low cost.

There are two key disadvantages to surveys. If a survey is conducted by mail, response rates can be very low, jeopardizing the validity of the data collected. There are mechanisms to increase response rates, but they will add to the cost of the survey. Later in this section, we include some tips for boosting response rates. Written surveys also provide no opportunity for respondents to clarify a confusing question, again creating a problem with the validity of responses. Good pre-testing of a survey instrument can reduce the likelihood that problems will arise.

*Examples:*

- Survey to track grassroots organizations' use of and satisfaction with technical assistance services you provide
- Survey of all organizations receiving technical assistance to learn about changes in their fundraising tactics and the results of their efforts to raise more money

**Interviews and Focus Groups**

Interviews also use standardized instruments, but they are conducted person-to-person either in person or over the telephone. In fact, an interview may use the same instrument created for a written survey, although interviewing generally provides the advantage of having the opportunity to explore questions more deeply than with a written survey. You can ask more complex questions in an interview, since you have the opportunity to clarify any confusion. You also can ask respondents to elaborate on their answers, eliciting more in-depth information than a survey can provide.

The primary disadvantage of interviews is their cost. It takes considerably more time and, therefore, more money to conduct telephone and in-person interviews. This usually means that you are able to collect information from far fewer people. The reliability of interviews can also be problematic if interviewers are not well-trained, since they may ask questions in different ways or otherwise bias the responses.

*Examples:*

- Interviews at different grassroots organizations to learn about the way in which they are applying new knowledge of partnership development
- Interviews with individuals within an organization to explore their perceptions of changes in capacity and ability to deliver services

A focus group is a particular type of interview conducted with a small group of people to obtain information in a defined area of interest related to your outcomes and indicators. While interviews with individuals are meant to solicit data without any influence or bias from the interviewer or other individual, focus groups are designed to allow the participants to discuss the questions and share their opinions. This means that people can influence one another in the process, stimulating memory or debate on an issue. The advantage of focus groups lies in the richness of the information generated. The disadvantage is that you can rarely generalize or apply the findings to your entire population of participants or clients. Focus groups are often used prior to creating a survey to test concepts and wording of questions. Following a written survey, focus groups are used to explore specific questions or issues more thoroughly.

*Examples:*

- Structured meetings with staff in a faith-based or community organization to learn more about their grants management practices, what worked during the year, and what did not
- Structured meetings with staff from several organizations to explore their use of computer technology for tracking financial data

**Observation**

Observations can be conducted of individual behaviors or interactions among individuals, of events, or of physical conditions within a site or facility. They require well-trained observers and detailed guidelines about whom or what to observe, when and for how long, and by what method of recording. The primary advantage of observation is its validity. When done well, observation is considered a strong data collection method because it generates first-hand, unbiased information by individuals who have been trained on what to look for and how to record it. Observation does, however, require time for development of the observation tool, training of the observers, and the data collection, making it a more costly data collection method than some of the others.

*Examples:*

- Observation of individuals participating in training to track the development of their skill in the training topic
- Observation of community meetings sponsored by grassroots organizations to learn about their partnership-building techniques and collaborative behavior

**Record Review**

Record or document review involves the systematic collection of needed data from internal, organizational records or official records collected by other groups or institutions. Internal records available to a capacity builder might include financial documents, monthly reports, activity logs, purchase orders, etc. The advantage of using records from your organization is the ease of data collection. The data is there and no additional effort needs to be made to collect it—that is if the specific data you need is actually available and up to date.

If the data is available, record review is a very economical and efficient data collection method. If the data you need is not available, it is likely well worth the time to make improvements to your data management system so that you can rely on internal record review for your outcome measurement work. Just a few changes to an existing form can turn it into a useful data collection tool. And just a small amount of staff training can increase the validity and reliability of internally generated data.

*Examples:*

- Sign-in logs from a series of workshops to track attendance in training, measuring consistency of attendance as an indicator of organizational commitment to learning
- Feedback forms completed by workshop participants to learn about satisfaction with training provided

Official records can include Federal, state, or local government sources such as the U.S. Census, health departments, law enforcement, school records, assessor data, etc. If the data available in official records are relevant to your outcomes and the indicators are accessible, then official record review is a very low-cost method for collecting data; but the questions about relevance and accessibility are major ones.

**Checklist for Selecting Data Collection Methods[8]**

This checklist can help you decide which data collection methods are most appropriate for your outcome measurement.

| **SURVEYS** | | |
|---|---|---|
| 1. Do I need data from the perspective of the participant, client, beneficiary, or customer? | Yes | No |
| 2. Do I have a systematic way to get it from these individuals? | Yes | No |
| 3. Do I need data that are standardized so that statistical comparisons can be made? (For example, will I need to report percents or other statistics?) | Yes | No |
| 4. Will participants be able to understand the survey questions? (Consider age, cultural backgrounds, etc.) | Yes | No |
| 5. Do participants have the necessary knowledge or awareness to *accurately* answer questions about the outcomes? | Yes | No |
| *If you answered Yes to questions 1 through 5, surveys may be appropriate for collecting data on your outcomes and indicators*. | | |

| **INTERVIEWS** | | |
|---|---|---|
| 6. Are more in-depth answers necessary to adequately measure the indicators or to get information on what is needed or what should change? | Yes | No |
| 7. Will it be necessary for someone to personally ask participants questions (either on the phone or in person) in order to collect the information related to this outcome? (Consider age, cultural background, as well as state of mind or receptivity of participants.) | Yes | No |
| *If you answered Yes to questions 6 and 7, interviews may be appropriate for collecting data on your outcomes and indicators*. | | |

8 Reisman, Jane and Judith Clegg. *Outcomes for Success!* (Seattle, WA: The Evaluation Forum, 2000), 94–97.

| OBSERVATION | | |
|---|---|---|
| 8. Is it difficult to accurately measure the indicators by asking people questions about opinions or perceptions? | Yes | No |
| 9. Can this outcome or indicator be assessed accurately by someone trained to observe it in action (i.e., can something actually be observed)? | Yes | No |
| 10. Do you have the staff resources for someone to observe events, conditions, interactions, or behaviors? | Yes | No |
| *If you answered Yes to questions 8, 9, and 10, observation may be appropriate for collecting data on your outcomes and indicators.* | | |

| INTERNAL RECORD REVIEW | | |
|---|---|---|
| 11. Do you have individualized records, reports, logs, or other systematic ways that you track things in your program or services? | Yes | No |
| 12. If an information system exists, are the data consistently entered into it in a timely way? | Yes | No |
| 13. If a system exists, can information be extracted from it easily? | Yes | No |
| *If you answered Yes to questions 11, 12, and 13, internal record review may be appropriate for collecting data on your outcomes and indicators.* | | |

| OFFICIAL RECORD REVIEW | | |
|---|---|---|
| 14. Do official records exist that track the data you need on your outcomes and indicators? | Yes | No |
| 15. Are the data accessible to you (i.e., will it be possible to get the cooperation of outside agencies or institutions in order to get access to official records)? | Yes | No |
| *If you answered Yes to questions 14 and 15, internal record review may be appropriate for collecting data on your outcomes and indicators.* | | |

## Validity and Reliability

Validity and reliability are two critical concepts in implementing effective outcome measurement systems. Validity is the accuracy of the information generated. The validity of a data collection instrument is how well it measures what it is supposed to measure. Putting in the time to create good instruments, carefully considering what is being measured, and pre-testing the instruments will help increase their validity.

Reliability refers to consistency, or the extent to which data are reproducible. Do items or questions on a survey, for example, repeatedly produce the same response regardless of when the survey is administered or whether the respondents are men or women? Bias in the data collection instrument is a primary threat to reliability and can be reduced by repeated testing and revision of the instrument.

You cannot have a valid instrument if it is not reliable. However, you can have a reliable instrument that is not valid. Think of shooting arrows at a target. Reliability is getting the arrows to land in about the same place each time you shoot. You can do this without hitting the bull's-eye. Validity is getting the arrow to land on the bull's-eye. Lots of arrows landing on the bull's-eye means you have both reliability and validity. Pre-testing is critical to ensuring that your data collection instrument will produce valid and reliable results.

## Instrument Development and Pre-Testing

This guidebook does not include information on how to create the instruments you'll need for data collection, although there are many resources available to help with this task (see the appendices). The appendices also feature a checklist you can use to assess survey and interview instruments and two examples of surveys related to organizational capacity building.

You need to pre-test each data collection instrument you create. To pre-test, you try out your tools with a small group of people who are representative of those you intend to include in your data collection. For a survey or interview, you must include people who are like those who will complete the survey or interview during your actual data collection. For observation, you must try out your form with individuals who will complete it during data collection. For record review, you must test the record extraction form with the individuals who will do the work to pull information from your records.

During the pre-test phase, you want people's help in checking the wording of questions, the content, the clarity of instructions, and the layout and format. You'll also want to know how long it took and whether any problems arose. Most importantly, you want to see if you get the responses you expect and the information you need.

## Deciding When and How to Collect Data

Once you have identified the data collection methods you intend to use, you need to decide when you will collect the data and how often. Then consider the procedures you need to put in place to ensure that your outcome measurement system is sustainable and produces quality data.

### Frequency and Scheduling of Data Collection

The following table describes the five approaches or designs you are likely to use for your data collection.

**Data Collection Designs**

| TYPE OF DESIGN | COLLECTION FREQUENCY | EXAMPLE |
|---|---|---|
| **Post-only Measures** | Data are collected once — at the end of the program, service, or activity | *Level of participant knowledge on a survey after a training workshop* |
| **Pre/Post Measures** | Data are collected twice — at the beginning, to establish a baseline, as well as at the end of the program | *Comparison of an organization's documented fundraising success before and after receiving technical assistance* |
| **Time Series** | Data are collected a number of times — during an ongoing program and in follow-up | *Monthly observations of an organization's collaboration meetings to track changes in partnership development and communication* |
| **Measures with a Comparison Group** | Data are collected from two groups — one group that receives the intervention and one that doesn't | *Comparison of data on skill development from individuals who participated in training and those who have not yet taken your workshop* |
| **Measures with a Comparative Standard** | Data are collected once — at the end of the program, service, or activity, and are compared with a standard | *Comparison of this year's data on organization's success in fundraising as compared to last year's data* |

The first three approaches in the table are fairly self-explanatory; the last two approaches need a little discussion. "Comparison groups" can be very useful in demonstrating the success of your intervention. The main question is whether you can find a group of people or organizations that is just like the group with whom you are working. In order to provide a valid comparison, the two groups must have the same general characteristics. A similar group may be difficult to find. However, if you are working with different groups at different times, and the groups are similar, this approach may work for you. Or if, as in the example of using a comparison group shown above, you have people on a waiting list for training who are similar to the people who have trained already, they may make a good comparison group. Or perhaps you might compare FBCOs receiving subawards with those who applied but did not receive them.

"Comparative standards" are standards against which you can measure yourself. There are standards of success in some fields, e.g., health mortality and morbidity rates, student achievement scores, teen birth rates. For intermediaries, however, there are unlikely to be many standards regarding your program outcomes or indicators. But you could compare your results for one time period to an earlier one, as shown in the example of using comparative standards in the table above. You would collect data for the first time period as your baseline and use it as your standard in the future.

**CHECKLIST ON DATA COLLECTION DESIGN**

When considering which approach is best, these questions may help you make a decision:

- ❑ When would change occur—immediately, gradually, or over a long period of time?
- ❑ Are there milestones that can be measured along the way to the outcome you are trying to achieve?
- ❑ What is the frequency of contact with the organizations with which you are working—once, weekly, monthly, or at some other interval?
- ❑ When will data be available? (When are some of your internal records available?)
- ❑ Are there any groups that might serve as comparison groups?
- ❑ Do you have baseline data you can use as a standard?
- ❑ Are you accountable for short-term outcomes or longer-term impacts?

### Data Collection Procedures

To finish your planning work, a number of procedural issues must be considered. You and your team will need to think about the following questions as you formulate and implement your data collection plan:

1. **Who will collect the data?** How will you recruit these data collectors? What steps will they need to take to collect the data? How will you train them? Finally, who will be responsible for monitoring the data collection process to ensure that you are getting what you need? It's important to answer each of these questions during your planning. You don't want to discover halfway through the process that your three-month follow-up surveys were not mailed out because you didn't identify who would do so!

2. **Have you prepared your clients for data collection?** Do they know that you will be collecting data, either at the time of service or in follow-up? Do they understand why it is important to you and how you intend to use the data? Organizations often have outcome reporting requirements themselves, so they are usually responsive if they have been alerted to your needs ahead of time. Advising them in advance about your data collection plans will help increase their willingness to participate during implementation.

3. **How will you ensure confidentiality and informed consent?** Anonymous and confidential do not mean the same thing. **Anonymous** means you do not know who provided the responses. **Confidential** means you know or can find out who provided the responses, but you are committed to keeping the information to yourself.

   You must ensure that you protect the confidentiality of any individual's data or comment. It is easy to make your surveys anonymous, but if you want to track people over time, you'll likely need to attach ID numbers to each person from whom you collect, keeping a list of the names and numbers in a locked file.

   It is important to inform people that you are measuring your program's outcomes and may use data they provide in some way. You must let them know that their participation is voluntary and explain how you will maintain the confidentiality of their data.[9]

---

9 A sample consent form can be found in United Way of America's *Measuring Outcomes: A Practical Approach* http://www.liveunited.org/outcomes/resources/mpo.

4. **How will you ensure quality of data?** As data is collected and entered into the storage mechanism, checking for errors and data quality is an important step that is easily overlooked. Build in time to review data and follow up about discrepancies in your overall timeline; the more data you collect, the more time you will need to assure its quality. Here are some typical strategies for assuring the quality of data:
   - **Double entry.** This entails setting up a system to collect data twice, and then comparing for discrepancies. This can be costly and time-consuming, but it is the most thorough method for quality control.
   - **Spot checking.** This entails reviewing a random sample of data and comparing it to the source document for discrepancies or other anomalies. If discrepancies are found, the first step is to identify any patterns (data entered during a particular time period or by a specific staff person; data associated with a particular beneficiary organization; a specific type of data that is incorrect across many records, e.g., if all data for additional persons served at an organization was formatted as a percentage instead of as a whole number). The capacity builder may need to review all the data entered, especially if there is no discernible pattern to the errors.
   - **Sorting data** to find missing, high, or low values. If you are using a database or spreadsheet function, identifying outliers—those pieces of data at either extreme—is very easy, whether through the use of formulas or sorting functions.
   - **Use automation**, such as drop-down menus. Automating a data collection provides a uniform way to report information and makes sorting and analyzing data much easier. For example, organizations reporting the number of additional persons served will all use the same language to report the outcome, whereas without such automation the language could vary significantly from report to report. Additionally, more sophisticated forms can pre-populate performance goals from an existing database, which reduces data entry errors made by those filling out the forms.
   - **Format database** to accept only numbers. Whether organizations are filling out forms directly or your staff is entering data from a handwritten form, formatting your data fields to accept only numbers reduces errors related to typos.
   - **Review data for feasibility.** This strategy requires that a staff person who is familiar with the organization's capacity building interventions and who has a good eye for detail review the data collected and identify anomalies. Some of these anomalies may not appear with general sorting.
   - **Discuss data discrepancies** with the organization. If discrepancies remain unexplained after implementing any of these quality assurance mechanisms, take the data back to the organization for discussion and clarification.

5. **How can you boost response rates?** Low response rates can threaten an outcome measurement effort. Following up with organizations that are no longer actively participating in a capacity building program is especially difficult. Listed here are some strategies to increase response rates:
   - **Tie data collection to project milestones.** Throughout the course of the capacity building relationship, it is relatively simple to require organizations to report desired data. For example, an evaluation could be due as a requirement to moving on to the next phase of the project, such as releasing funds for a capacity building project or approving a consultant to begin work. However, once the organization exits the capacity building program, the capacity builder loses this leverage.
   - **Conduct an exit interview.** Once the entire engagement is complete, the capacity builder can schedule an exit interview, on the phone or in person, to collect data. Participation in this interview can be mandated in a memorandum of understanding. The organization will need to know what data to prepare for the interview. The advantage of this method is that an exit interview is close

enough to the intervention that the organization may still be invested in maintaining its relationship with the capacity builder and follow through on the commitment. However, the organization may not have realized all its possible outcomes, and therefore the data may not capture some of the ripple effects, where outcomes are realized after the data has been collected.

- **Stay in touch.** By holding monthly meetings or conference calls with organizations after they exit the program, the capacity builder can maintain more informal connections. The organizations have access to advice and support and may be more likely to participate in a follow-up data collection effort. Establishing a community of practice among organizations so that they have even more reason to be in touch with each other and your organization is one way to implement this strategy.
- **Provide the outcome data to the organization.** Offer organizations a short summary report card of the data you collect from them and demonstrate how it can be used as a marketing tool. This summary can prove invaluable to a program, and it may increase the number of responses you get to your data surveys. If you can use the merging functions available in most productivity software like Microsoft Word and Outlook, generating report cards for tens or even hundreds of organizations may take just a few hours.
- **Offer multiple collection methods.** Be available to complete the survey on the phone with the organization. Be available to go to the organization's headquarters and conduct the survey in person. Be prepared to offer language translation services if necessary, offer the survey electronically (at least as an attachment to an e-mail), or mail the survey with a stamped envelope. Response rates will increase the easier it is for an individual to complete the survey, and preferences will vary across individuals.
- **Be culturally competent.** Capacity builders may take great steps to ensure that training and technical assistance is culturally appropriate, and this should be done for data collection efforts as well. Moreover, if you are engaging a third party to collect data—a consultant or a team of interns, for example—remember that being a third party means that they have not had the benefit of getting to know an organization and its staff through the course of the capacity building engagement. Language barriers, cultural differences, and individual preferences can influence whether you are likely to get a response.
- **Introduce your external data collectors.** If you are working with those third parties, introduce them to the organizations you are working with early on. If maintaining a relationship helps improve response rates, then so too will the lack of a relationship hurt response rates. As a caveat, be sure to maintain confidentiality about the results, especially if the third party is collecting direct feedback about your services.

Finally, when you put together your outcomes, indicators, data collection methods, frequency and schedule of data collection, as well as procedures, you will have created your plan for outcome measurement. Worksheets to assist you with your evaluation planning are in the appendices of this guidebook. The bottom line is that how you collect your data and your attention to those you're collecting it from will ensure that you have optimum quality data to work with in the analysis phase.