

Monitoring and Evaluation: Showing How Democracy and Governance Programs Make a Difference

By

The International Republican Institute's

Office of Monitoring and Evaluation



©2013 The International Republican Institute

Monitoring and Evaluation: Showing How Democracy and Governance Programs Make a Difference

Copyright © 2013 International Republican Institute

All rights reserved.

Permission Statement: No part of this publication may be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system without the written permission of IRI.

Requests for permission should include the following information:

- A description of the material for which permission to copy is desired.
- The purpose for which the copied material will be used and the manner in which it will be used.
- Your name, title, company or organization name, telephone number, fax number, e-mail address and mailing address.

Please send all requests for permission to:

ISBN: 978-0-9915133-0-7
International Republican Institute
1225 Eye Street, NW Suite 700
Washington, D.C. 20005
Email: evaluation@iri.org
Fax: 202-408-9462

Disclaimer: The author's views expressed in this publication do not necessarily reflect the views of the International Republican Institute and the United States Agency for International Development (USAID).

About This Handbook

This handbook was developed by the staff of the Office of Monitoring and Evaluation at the International Republican Institute (IRI). IRI is a non-partisan, non-profit organization founded in 1983 with the goal of promoting freedom and democracy worldwide. This handbook was originally developed to help IRI program staff understand standards and practices from the field of monitoring and evaluation (M&E) and apply them to their work. As such, this handbook was informed by many experts and ideas within the M&E community, but focuses on their applicability to democracy assistance programs specifically.

IRI's Office of Monitoring and Evaluation would like to thank the many people that have contributed to this handbook, particularly the IRI staff as well as staff from EnCompass, LLC and Social Impact, Inc., whose expertise is also reflected in this handbook. We would also like to acknowledge the contributions of the National Endowment for Democracy (NED), the United States Agency for International Development (USAID) and the United States Department of State in supporting IRI's monitoring and evaluation efforts.

IRI would like to acknowledge the following sources that informed this handbook, and encourages those who wish to know more to refer to them directly:

- Bamberger, Michael, Jim Rugh, and Linda Mabry. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage Publications, 2006.
- Creswell, John W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th Edition. Washington D.C.: Sage Publications, 2002.
- Davidson, E. Jane. *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: Sage Publications, 2005.
- Davies, Rick, and Jess Dart. *The Most Significant Change (MSC) Technique: A Guide to Its Use*. Cambridge, UK: Davies and Dart, 2005.
- *Design Monitoring & Evaluation For Peacebuilding*. Web. (<http://dmeforpeace.org/>).
- Doucette, Anne. "Applied Measurement for Evaluation." Lecture at The Evaluator's Institute, George Washington University. Washington, D.C.: July 2012.
- Earl, Sarah, Fred Carden, and Terry Smutylo. *Outcome Mapping: Building Learning and Reflection into Development Programs*. Ottawa, Canada: International Development Research Centre, 2001.
- "Evaluation Tips: U.S. Agency for International Development." *U.S. Agency for International Development*. Web. (http://transition.usaid.gov/policy/evalweb/evaluation_resources.html).
- Fetterman, David and Abraham Wandersman. *Empowerment Evaluation Principles in Practice*. New York, NY: The Guilford Press, 2005.
- Fowler, Floyd J. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications, 1995.

- Funnell, Sue C. “Developing and Using a Program Theory Matrix for Program Evaluation and Performance Monitoring,” *Program Theory in Evaluation: Challenges and Opportunities*. Spec. issue of *New Directions for Evaluation*. Ed. Patricia J. Rogers, Timothy A. Hacsí, Anthony Petrosina, and Tracy A. Huebner. 2000.87. (Fall 2000): 91-101.
- Funnell, Sue C. and Patricia Rogers. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: John Wiley and Sons, Inc., 2011.
- George, Alexander L., and Andrew Bennett. *Case Studies and Theory Development in the Social Sciences*. Cambridge, MA: MIT, 2005.
- Kumar, Krishna. *Evaluating Democracy Assistance*. Boulder, CO: Lynne Rienner Publishers, 2013.
- Miles, Matthew, Michael Huberman, and Johnny Saldana. *Qualitative Data Analysis: An Expanded Sourcebook*. 2nd Edition. Thousand Oaks, CA: Sage Publications, 1994.
- Morra Imas, Linda and Ray C. Rist, *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, D.C.: The World Bank, 2009.
- Patton, Michael Q. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford, 2011.
- Patton, Michael Q. *Qualitative Research & Evaluation Methods*. 3rd Edition. Thousand Oaks, CA: Sage Publications, 2002.
- Patton, Michael Q. *Utilization-focused Evaluation: The New Century Text*. 4th Edition. Thousand Oaks, CA: Sage Publications, 2008.
- Preskill, Hallie S., and Tessie Tzavaras. Catsambas. *Reframing Evaluation through Appreciative Inquiry*. Thousand Oaks, CA: Sage Publications, 2006.
- *Principles for Evaluation of Development Assistance*. Paris: OECD-DAC, 1991.
- Scriven, Michael. *Evaluation Thesaurus*, 4th Edition. Newbury Park, CA: Sage Publications, 1991.
- Scriven, Michael. “The methodology of evaluation”. Stake, R.E. *Curriculum Evaluation*. No. 1. Chicago: Rand McNally, 1967, pp. 39-89.
- Stewart, David, Prem Shamdasani, Dennis Rook. *Focus Groups: Theory and Practice*. Cambridge, MA: Harvard University Press, 2006.
- Vondal, P. *Performance Monitoring and Evaluation Tips: Using Rapid Appraisal Methods*. Washington, D.C.: USAID Center for Development Information and Development Evaluation, 2010.

M&E for democracy assistance programs is an emergent field, and this handbook has benefited from the shared experiences of other programs, organizations and experts. In this spirit, IRI invites you to send comments and feedback to: evaluation@iri.org.

Please note that IRI does not warrant that any of the content of this handbook are accurate, complete, or current. IRI may update this handbook periodically: please contact evaluation@iri.org for information on recent updates. IRI does not, however, make any commitment to update the materials. The content of this handbook is provided "as is." The published content is being distributed without warranty of any kind, either express or implied. The responsibility for the interpretation and use of the content lies with the reader. In no event shall the IRI be liable for damages arising from its use.

This handbook is made possible by the generous support of the American people through the United States Agency for International Development (USAID) under Award No. DFD-A-00-08-00350-00. The opinions expressed herein are those of the author(s) and do not necessarily reflect the views of IRI, USAID or the United States Government. Any errors or omissions are the sole responsibility of the authors.

Table of Contents

Chapter 1: An Introduction to Monitoring and Evaluation (M&E)	8
What is M&E.....	8
Why M&E.....	8
M&E at the International Republican Institute	8
IRI's Office of Monitoring and Evaluation.....	9
M&E 101 Cheat Sheet.....	10
Chapter 2: M&E at the Program Design and Planning Stage	12
Why It Is Important to Think About M&E When Designing Your Project.....	12
What is a Result?.....	12
M&E at the Program Design Stage: What You Need to Know	14
<i>Step 1: Defining Your Problem</i>	14
<i>Step 2: Identifying Your Objective(s)</i>	14
<i>Step 3: Describing Your Program Theory</i>	17
<i>Step 4: Mapping the Results Chain</i>	19
<i>Step 5: Developing Effective Indicators</i>	20
<i>Step 6: The M&E Indicator Matrix: What Is It For and How do You Fill It Out</i>	22
<i>Step 7: Including a Graphical Framework</i>	25
<i>Step 8: What to Say in the Proposal Narrative about the M&E Plan</i>	26
Consider a More Rigorous Program Design	27
Sample Results Chain, M&E Indicator Matrix, Results Framework and Graphical Framework	29
Needs Assessment Guide	34
Formative Country Program Assessment.....	35
Outcome Mapping Guide	37
Program Theory Framework Guide.....	39
Chapter 3: M&E in the Field: Measuring Success with a Proper M&E System	41
Established Purpose and Scope	41
Data Collection.....	42
What Do I Need to Consider?	42
Making Sure the Data is Actionable.....	43
Examples and Guides to Common Data Collection Methods	44
Data Collection Tools.....	48
<i>Public Opinion Research</i>	48
<i>Self-Administered Surveys (such as Training Questionnaires)</i>	49
<i>Focus Group Discussions</i>	53

<i>In-Depth Interviews</i>	54
<i>Scorecards and Checklists</i>	57
<i>Observation and Visual Evidence</i>	60
<i>Content Analysis (Document Review)</i>	61
Data Collection Approaches for Special Circumstances	63
Program and M&E Activity Workplan	65
Increasing Data Collection Rigor	65
Data Collection Tips.....	67
Data Analysis	67
Qualitative Data Analysis.....	68
Quantitative Data Analysis.....	69
Data Use	73
Data Storage	74
Evaluation Ethics.....	74
Chapter 4: Evaluations: Taking Stock of Your Programs	76
Why a Formal Evaluation?.....	76
Formal Evaluations: Who Conducts? Who Commissions?.....	77
Designing and Implementing an Evaluation: What to Consider	80
<i>Step 1: Determine the Evaluation Need and Purpose</i>	80
<i>Step 2: Design the Evaluation</i>	83
<i>Step 3: Collect the Data</i>	86
<i>Step 4: Analyze the Data</i>	87
<i>Step 5: Use the Results of Evaluations</i>	87
<i>Step 6: Disseminate Evaluation Results</i>	88

Chapter 1: An Introduction to Monitoring and Evaluation

What is M&E?

Monitoring and Evaluation (M&E) is a conjoined process. **Monitoring** is the systematic collection of data on specified indicators to help you know if you are on track toward achieving your desired results. **Evaluation** is the objective assessment of the relevance, efficacy or efficiency of a program. With both, a program can determine results and lessons learned about its efforts.

In order to work, M&E must be an integrated system, one that is rigorous and objective but also reflective of individual program approaches and needs. Only by inculcating M&E at every stage of a program's lifecycle can the pistons of this system work effectively.

Why M&E?

Succinctly put, M&E is the process of learning about a program's implementation and results, and using that knowledge to make decisions about a program. Done well, M&E can help you set goals and design an effective program, help you adapt the program to changing circumstances, and improve the program along the way. This ensures that activities themselves are the "right" activities to address the problem you are trying to solve in the environment in which you operate, and ensures that those activities are effective and efficiently implemented.

In other words, you carry out M&E to improve your programs continuously. Of course, you also conduct M&E because it is an important part of being accountable to your donors, your beneficiaries and other stakeholders. M&E helps you justify your rationale for your program, demonstrates that you are doing the right thing and doing it well, and ensures that you can point to the broader impact of your efforts.

Lastly, you do M&E because it is critical to the overall goal of your work! M&E is an important tool in helping to develop cutting edge programs, adapting programs to meet changing circumstances, and advancing the field of democracy and governance (DG) by learning from and utilizing evaluation results.

M&E at the International Republican Institute

The International Republican Institute's (IRI) rigorous and innovative approach to M&E reflects its commitment to continuously improving its programming, capturing and communicating its results, ensuring resources are invested in programs that produce results, and helping advance the field of DG work.

IRI's evaluation philosophy is guided by the following four principles:

- M&E efforts reflect the highest standards of objectivity, rigor and ethics, including guidelines established by the American Evaluation Association.
- M&E efforts are participatory and focused on use, resulting in relevant findings that show the impact of IRI's work, and inform and improve program implementation and design.

- M&E efforts respect the interests of key stakeholders, in particular in-country partners, and avoid unnecessary risks and disruptions to ongoing programming.
- M&E efforts embrace the fact that DG programs are implemented within complex and often quickly changing environments, and should therefore explore program results in the context of the broader system of actors, relationships and processes that affect and are affected by IRI's work

IRI's Office of Monitoring and Evaluation

The Institute's M&E efforts are overseen by its Office of Monitoring and Evaluation. The Office offers a one-stop shop through which program staff can get advice on M&E needs at any point in the program lifecycle, including: developing M&E plans, infusing evaluative thinking into program design, developing data collection tools and assisting with analysis, compiling data into meaningful reporting, and designing formal evaluations.

M&E 101 Cheat Sheet

What is Monitoring & Evaluation?

M&E takes place over the lifecycle of a project – from the proposal stage to the final report, which is then used to inform the next program. **Monitoring** is the systematic collection of data on specified indicators to help you know if you are on track toward achieving your desired results. **Evaluation** is the systematic and objective assessment of the relevance, efficacy or efficiency of a program. Evaluation can take place at any time, and any part of the program can be evaluated: from program needs to program outcomes and impact.

Why is M&E Important?

- It helps keep you on track to achieve your desired results – helps identify program weakness and allows you to take corrective action
- It shows others that what you're doing makes a difference – or if not, why not
- It helps you learn and improve your program and the field of DG

M&E is Logical!

When starting a new program, ask yourself these questions:

- What is the problem?
- What are the desired results?
- What resources will be necessary to achieve these results?
- What activities will take place to achieve these results?
- What will these activities produce?
- What will be the benefits of these products?
- What change will all this make?

Then all you have to do is put your answers in “M&E-speak”:

<i>M&E Term</i>		<i>Definition</i>	<i>Example</i>
Objective	=	Goal, Desired Result	Political parties in Country X select party leaders through an inclusive, participatory process.
Inputs	=	Resources	Staff time; transportation, etc.
Processes	=	Activities	Trainings on internal democracy.
Outputs	=	Products, Yields	Party members trained on internal democracy.
Outcomes	=	Benefits	Political parties elect their leaders.
Impact	=	Change	Parties are more representative of their members.

For visual people, it might help to think about it this way:



What Makes a Good Objective?

Specific – What are you trying to achieve, where and with whom?

Measurable – Will you know when you've achieved it?

Achievable – Can you do it, with the money you have and the people you have, in the time you have?

Relevant – Is it actually a solution to the problem you've identified?

Time-bound – Is it clear when it will take place?

Problem: I'm not in shape.

- Bad objective: To get in shape.
- Good objective: I qualify for the next Boston Marathon.
 - I know what I'm trying to achieve; I'll know when I've achieved it; I can realistically do it; it will get me in shape; timeframe is clear.

An objective is the highest level result you want to achieve through your program; thus, it can be at the output, outcome, or impact level, depending on the circumstance.

What is an Indicator?

- An indicator is a signpost – it visually shows the condition of a system.
 - Example: To see if you're sick, you take your temperature – your *temperature* is an *indicator* of whether or not you are sick.
- An indicator helps you know if you're on track to reach a goal.
 - Example: If you want to lose 15 pounds by August, weighing yourself regularly helps you see if you're on track – your *change in weight* is the *indicator* of whether or not you're on track.
- You can have all kinds of indicators:
 - Process Indicator: Measure of activities
 - Output Indicator: Measure of products/yields
 - Outcome Indicator: Measure of the benefits (often behavioral change)
 - Impact Indicator: Measure of systemic and sustainable change
- F-indicators are general foreign assistance indicators developed by the U.S. government. These grants should include F-indicators in their M&E plan.

What Makes a Good Indicator?

Direct – Is it actually measuring what you're trying to measure?

Clear – Is it clear what kind of change is taking place? Is it clear where and with whom the change is taking place?

Quantitative – Is it quantifiable? Can you count it or otherwise quantify it? If not, can you definitively answer yes/no as to whether it has been achieved?

Feasible – Can you realistically conduct this measurement with the money/people/time at your disposal?

Remember: An *objective* is a goal. An *indicator* is a measurement of progress toward reaching that goal.

Example Indicators

Output: Party leaders and members have increased knowledge. *Indicator:* Number of party members demonstrating increased knowledge of internal democracy.

Outcome: Political party leaders adopt internal election law. *Indicator:* Number of parties that institutionalize internal elections.

Impact: Party leadership is more representative of its members. *Indicator:* Increase in party members' opinion about leaders as measured by poll.

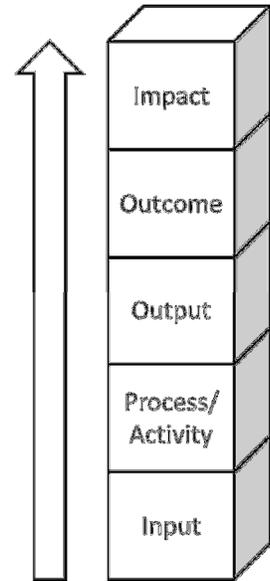
Chapter 2: M&E at the Program Design and Planning Stage

Why It Is Important to Think About M&E When Designing Your Project

Successful programs have one thing in common: they are led by a team that thinks analytically about the cause and effect of the program, adapts the program based on evidence and focuses on achieving results and documenting the achievement of those results.

In the end, the impact of your work depends heavily on what your participants do with the support given to them, and how these actions combine with other external factors to produce systemic change. For this reason, influence is strongest over activities (you control how they are implemented) and weakest when it comes to impact (you can't control whether a parliament passes legislation). This influence can also be thought of in terms of sustainability. The immediate change that results from activities (such as knowledge gained) is short-lived. Systemic change (a law is passed) is much more lasting.

These different degrees of influence and change are described in terms of different levels of results. Results demonstrate the progression of change from inputs (whether you succeed in utilizing your resources) to impact (whether systemic and sustainable change has taken place).



What is a Result?

A **result** is whatever happens as a consequence of something else that happens. You include inputs and activities/processes in a results chain to illustrate all the steps from the beginning to the end of a program.

Please note: In M&E-specific terminology, **result** is also used to refer to a significant result at the outcome or impact level that indicates a milestone progress point in achieving the objective. Often, these are referred to as intermediate results or interim results.

Input: Inputs are what you put into a program: staff time, cars, opinion surveys, etc.

Process/Activity: Processes are basically activities; it is whatever process a program undertakes as part of its intervention – what the program does.

Output: An output identifies the immediate **product** or **yield** of an activity. If the activity is a training, then the output is that participants acquire knowledge or skills. The knowledge and skills are the activity yields. A program has a lot of control over outputs. You provide the training (input) → people are trained and gain knowledge or skills (output).

Outcome: An outcome measures the **benefit of an activity**, such as a **change in behavior, relationship or action**. In a training program, the outcome would be what participants do with the

knowledge or skills they have gained from the training. If a participant holds a town hall meeting after receiving constituent outreach training, that is an outcome. Outcomes are a step beyond your control, but are linked to the service you provided. The service you provided may not necessarily have caused the outcome, but it facilitated the process. The participant gained skills from the program (output) → the participant goes out and uses those skills (outcome). Note that there can be multiple levels of outcomes (immediate, intermediate, long-term).

Impact: An impact measures **permanent or sustainable changes in a system or condition.** Examples include substantive policy changes, increased voter support for a political party that campaigned based on issues, changes in political party operating procedures or strategies, or an increase in voter turnout. This is the ultimate goal of your work: to see impact. Impact is rarely (if ever) a direct or exclusive result of your work. Often, there are complementary developments, over many years, that together lead to change. With the training program example, if the participant holds a town hall meeting, receives feedback from constituents, and then goes on to achieve something based off that feedback (such as a budget allocation, new law or regulation drafted, or a service is delivered), that constitutes impact. The participant uses the skills gained from trainings to do something else (outcome) → as a result of using that skill over and over, the participant gains information that (s)he uses to promote a new regulation that provides permanent or sustainable change (impact). The key to impact is sustainability of the change.

As you can see from these examples, influence in the change process (from inputs to impact) is strongest at the outset, and then gradually reduces the further you move up in the results chain toward impact.

This can be illustrated in the following chart (see figure one). On the horizontal axis you can see the progression of results, which is essentially the progression of change. On the vertical axis you can see the influence or relative control programs have. The solid line denotes your program’s influence and the dotted line denotes the influence of your partners.

Figure One: Attribution vs. Contribution Chart

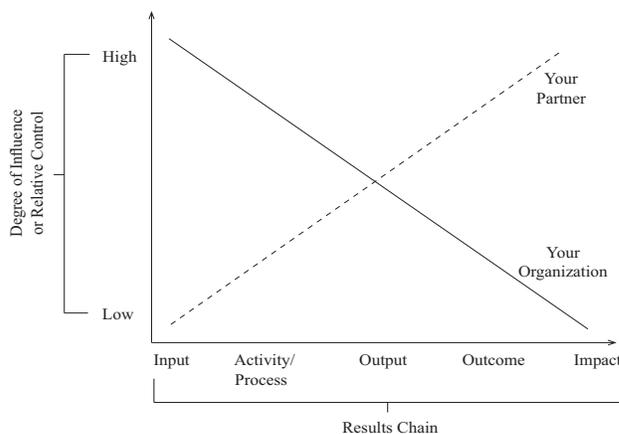


Figure one shows that the program has the most control over activities it implements and increasingly less control over what follows. Why? Because things like the political environment or the literacy level of participants influence whether or not a training program will have an effect.

For this reason, it is difficult to **attribute** an observed impact directly to the program's work, or to say that the program **caused** something to happen. So, *attributing* something to the program is often spoken about less versus the program's **contribution**. What did the program do to support that change?

M&E at the Program Design Stage: What You Need to Know

Most donors require you to submit a monitoring and evaluation plan together with your proposal. The M&E plan outlines your program objectives as well as how you will measure progress toward those objectives. But M&E thinking can, and should, inform your proposal from the very first stages. A well-designed proposal is easy to monitor, and is more likely to achieve your goals in an efficient and effective fashion. Using M&E thinking, you can identify what about a situation needs to change, how that change will happen, and how your program will contribute to that change. M&E also helps you define realistic goals for your program and develop a plan to monitor progress toward your goals.

The following section explains how M&E thinking should inform proposal design, and then how that proposal design can be linked to a robust M&E plan.

➤ ***Persuasive and focused writing is critical: make sure the reader understands why the problem you discuss is the root of the overall problem and that your program addresses it. Too often the objective does not match the problem!***

Step 1: Defining Your Problem

The first step is to clearly explain the purpose of your program by defining the problem you are trying to address. This is your problem statement.

The problem statement should include:

- Information about the situation that needs to change
- Who the problem affects
- The problem's causes
- The broader effects of the problem; its impact on society

Step 2: Identifying Your Objective(s)

Identifying the Level of the Objective(s)

With your problem defined, you are ready to write objectives. Remember that, while your *goal* is the longer-term *change* you hope to influence, your objectives are more narrowly focused and measurable.

An objective is the highest-level result you promise to achieve through your project. An objective is frequently set at the outcome level, but can also be set at the impact or output level. In rare cases, you may decide to set your objective at the input level. The level or ambition of your objective will depend on the country in which you work, the size of the grant, the donor's priorities and the history of your work in that country.

An objective can, in theory, be at any level along the results chain:

Process/activity-level objective: In some countries, accomplishing an activity, such as a training activity, is in and of itself an accomplishment; think about a closed society program.

Example: Party agent coordinators from all registered parties have access to training programs in the lead up to the 2015 parliamentary elections.

Output-level objective: In some programs, your goal is for participants to gain knowledge and skills, or to develop something, such as a document.

Example: Political party research units develop five-year plans that address the party's research needs and lay out clear timelines for completion.

Outcome-level objective: In situations where you are more confident that you can achieve a goal over which you have less direct control, you can write outcome-level objectives. An outcome-level result depends on a chain reaction influenced by the program's work: you cannot guarantee that an outcome will occur, but you are confident – because you know the situation, you have enough time and funds – that you can influence it to happen. Often, you can think of this as encouraging changes in behavior.

Example: Political parties develop issue-based platforms targeting youth in expectation of the 2011 local elections.

Impact-level objective: An impact-level objective is much more ambitious, and depends on a more complex chain reaction influenced by the program. This type of objective is most likely appropriate for situations where the political environment is such that you have confidence you can encourage more systemic movement in a positive direction, or change in a larger system or process.

Example: Parliament passes legislation aligning domestic processes with international labor conventions in advance of the 2014 international summit.

➤ ***Decide what your objective is before starting on the program theory. A program theory lays out how the objective will be achieved – if you don't know the level of change you wish to achieve or even what you want to change, then writing the program theory is like grasping at straws.***

➤ ***Remember to refer to the donor solicitation to ensure that your objectives align with what the donor wants to achieve.***

It is important to consider what level objective you want to achieve. Half the battle of a proposal is understanding what you are trying to achieve! Don't change or reword objectives at the last minute; doing so may significantly alter the scope and scale of the program.

Writing S.M.A.R.T. Objectives

Now that you have decided what degree of change your objective is focusing on, you can write S.M.A.R.T. objectives.

A good objective is:

Specific – What are you trying to achieve, where and with whom?

Measurable – How will you know when you've achieved it?

Achievable – Can you do it, with the money, people and within the time frame of the program?

Relevant – Is it actually a solution to the problem you've identified?

Time-bound – Is it clear when it will take place?

Example of a S.M.A.R.T. Objective:

Problem: My program needs funding.

Weak Objective: To write a good proposal.

This is a weak objective because it's vague about what defines success (what is a good proposal?), making it impossible to measure or achieve. There is also no timeframe indicated.

Strong Objective: I will write a proposal that responds to all request for application (RFA) requirements in 2015 that wins funding.

This is a strong objective because you know what you are trying to achieve, you will know when you've achieved it, you can realistically do it, and the timeframe is clear. An underlying assumption is that by winning funding through a proposal, you are writing a good proposal. Note that you have defined good by (1) winning funding and (2) responding to all RFA requirements.

Here are some examples of objectives for a defined problem, with explanations of how to make them S.M.A.R.T.:

The Problem: Political parties do not offer voters meaningful campaign platforms that parties can be held accountable to achieving while in office. Political parties generally campaign using famous names rather than discussing and promoting specific issues of interest to citizens. Voters tend to support parties based on what famous person joins rather than the important issues.

➤ ***Objectives are what you promise your donors you will achieve. Be careful not to overpromise!***

Weak Objective 1: Voters join political parties based on issues.

This is a weak objective because it is not specific (voters from where?), it is not achievable during the life of the grant (you cannot force voters to join parties), it is not a solution to the root of the problem (it's not relevant because the focus is on the wrong target audience for your interventions – voters rather than parties), and it is not time-bound. It is probably measurable.

Weak Objective 2: Voters in city X join political parties in anticipation of the 2011 parliamentary elections based on identification with party support of issues.

This objective is now time-bound (2011 parliamentary elections), more specific (voters in city X), and probably measurable, but it is still not achievable or relevant to the problem.

Strong Objective (output level): Political parties in city X are better prepared to perform issue-based constituent outreach in anticipation of the 2011 parliamentary elections.

This objective is now S.M.A.R.T. You have shifted focus to parties and off the voters, recognizing that voter decisions are limited by the options presented to them; this makes the objective relevant. With the change to “strengthen capacity...” the objective is now achievable. You can’t force parties to campaign a certain way or voters to vote a certain way, but through your activities, you can impart skills and knowledge that strengthen the capacity of parties to perform issue-based outreach.

Strong Objective (outcome level): Political parties in city X campaign on issues of concern to voters, in anticipation of the 2011 parliamentary elections.

This objective is an aspirational objective because it is far more ambitious: stepping beyond increasing skills, you are aiming to achieve an outcome – a campaign in city X that is based on a substantive dialogue.

Step 3: Describing Your Program Theory

After identifying what level of change you expect to work toward, you can define your program theory.

A program theory lays out the approach or framework for your program. You may also hear a program theory referred to as a **development hypothesis** or **change model**; they are all similar concepts. The theory basically explains how and why the desired change is expected to come about: What has to happen for goals to be met? How do your activities achieve results? Thinking through the program theory first is essential to guiding the remainder of the proposal: from determining and defining activities to designing the evaluation plan. A side benefit of the program theory is that it explains why you selected your activities, so when someone reads the program description (s)he already understands the need for those activities, given your approach.

A program theory statement can be about a paragraph or more in length, and must answer the following questions, all stemming from the problem as you have defined it.

To create a program theory, Sue Funnell and Patricia Rogers suggest considering a program theory as comprised of two parts: a **theory of change** (what change do you want to see and how will change happen?), and a **theory of action** (what action must be taken to encourage that change?).¹

¹ Funnell, Sue C. and Patricia Rogers. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: John Wiley and Sons, 2011.

What is your theory of change?²

What will change look like, and which main actors will bring about this change?

What is your theory of action?

- What community/need are you addressing? (e.g., what is your goal?) [***Question I***]
- What is your strategy/approach? (e.g., what kind of program are you proposing, a local governance support program?) [***Question II***]
- What are your desired results? (output, outcome, impact) [***Question III***]
- What are influential factors? (e.g., support from party leadership) [***Question IV***]
- What are your assumptions? (e.g., that electoral regulations will be passed) [***Question V***]

² These checklist questions should provide a good starting point for drafting a program theory, but for more guidance please refer to: Funnell, Sue C. and Patricia Rogers. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. San Francisco, CA: John Wiley and Sons, Inc., 2011.

Example:

Problem Statement: Political parties do not offer voters meaningful campaign platforms that parties can be held accountable to achieving while in office. Political parties generally campaign using famous names rather than discussing and promoting specific issues of interest to citizens. Voters tend to support parties based on what famous person joins rather than the important issues.

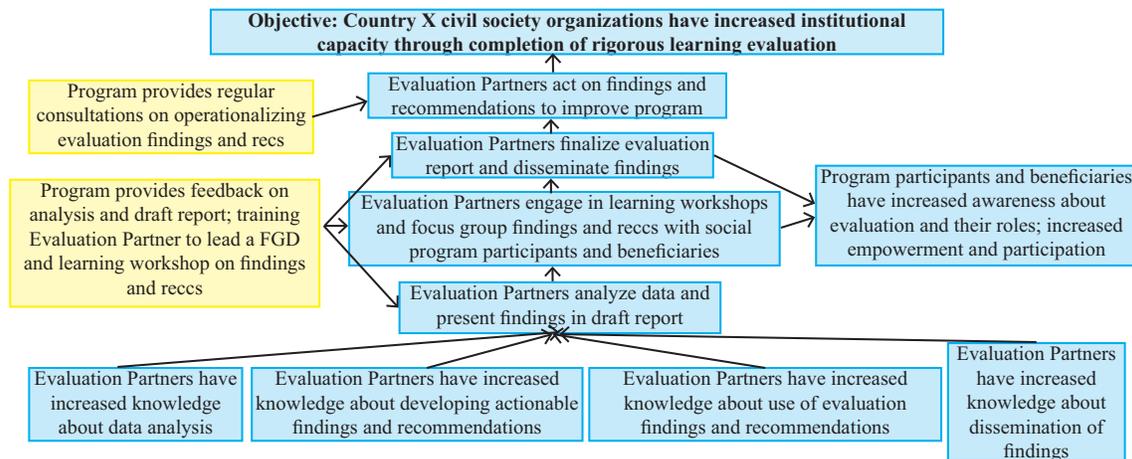
Theory of Change Statement: Parties will begin offering meaningful campaign platforms if they have access to knowledge about issues of priority to voters, and if they have the skills necessary to use these issues to gain more votes through the implementation of modern campaign strategies.

Theory of Action Statement: **[Question I]** Organization X proposes to work with political parties through trainings and individual consultations in a program focused on encouraging issue-based constituent outreach to provide voters meaningful choices. **[Question II]** Trainings and consultations will share knowledge of priority public issues of the day informed by organization X's survey results, and will focus on developing party strategies that reflect voter concerns. **[Question III]** Through trainings and consultations, program outputs will include an increased knowledge base among party members at all levels on the subject of meaningful voter outreach. Program outcomes will include the development of issued-based party outreach strategies, supported by the increase in knowledge from organization X's trainings and subsequent consultations. Impact will depend on other mitigating factors, but by helping to draft issue-based strategies and training party members on how to campaign effectively using issues rather than personalities, organization X aims to promote the development of a political environment based on a substantive dialogue during the 2011 parliamentary elections. This will be evidenced by issue-based platforms and messaging, as well as debates and town hall meetings. **[Question IV]** These achievements will depend on the policy context (including buy-in from party leadership and internal party structures), and the political environment. **[Question V]** The program also assumes that parliamentary elections will take place in 2011 as currently planned and that an electoral framework is passed and implemented in a way that is conducive to a free and open campaigning environment.

Notice that this program theory is the entire proposal summarized into a simple paragraph.

Step 4: Mapping the Results Chain

A results chain shows the sequence of results that need to occur to achieve the objective, from the lowest level result to the highest. It links the immediate results of your activities with the activities of your partners and other external factors.



Why do you want to create a results chain? Because it helps you do the following:

- Think more analytically about the cause and effect of your program.
- Identify results that need to take place for your objective to be achieved.
- Determine the activities that will influence those results.
- Identify the link between program components.
- Differentiate between external factors that influence the program and factors within your control.
- Identify key assumptions.
- Highlight potential indicators for M&E.

Plotting out the sequence of results that need to be achieved – from your activity to your objective and beyond – helps you see where the gaps in your reasoning are. Knowing this puts you in a better position to plan both your activities and M&E.

There is no rule as to the number of results you need to define per objective. Just remember that the results indicators are also an agreement between the program and the donor about: (1) what constitutes your criteria of success, and (2) what kind of information the program will provide to the donor during the program.

➤ ***Never promise activities or results that do not meet the objective! If activities don't address the objective, then they are often a waste of time and resources. Make an effort to engage with the funder to strategize a better use of funds. In the end, the funder wants what is most efficient and effective.***

Step 5: Developing Effective Indicators

Each result has a range of possible signs or indications of successful achievement. These are commonly referred to as **indicators**. Indicators tell you if you are on track or if something changed as a result of your work and how it changed.

Indicators can be quantitative or qualitative. **Quantitative** indicators provide information about things that can be expressed in numbers. **Qualitative** indicators provide information about things that are expressed through description and narrative.

Indicators

Just as results can be at the output, outcome or impact level, indicators can also be defined as being at the output, outcome or impact level. The result and its corresponding indicator share the same level of change.

Example:

To see if you're sick, you take your temperature – your temperature is an indicator of whether or not you are sick.

If you want to lose 15 pounds by August, weighing yourself regularly helps you see if you're on track – your change in weight is the indicator of whether or not you're on track.

A good indicator is:

Direct – Measures only what is intended to be measured.

Relevant – Measures a result that will lead to the achievement of the objective.

Clear – It is clear what kind of change is taking place where and with whom the change is taking place, and is unambiguous.

Definitive – You can definitively determine whether the result being measured has been achieved. The indicator consistently produces the same measure of results when used repeatedly to measure the same condition or event.

Measurable – It is actually possible to measure. Often, the change you hope to achieve is difficult to measure, especially at the impact level.

Feasible – It is realistic to conduct the measurement with the money, people and/or time at your disposal.

➤ ***It is recommended that you select a range of result levels for the simple reason that doing so provides a balanced look at how the program is going overall, and it tends to diversify the kind and complexity of data collection you do.***

Examples:

Sample Result 1: Political party leadership gains knowledge (including of polling results) about public opinion research and voter concerns.

Sample Indicator 1.1: Number of members of political party leadership that participate in public opinion polling methodology and polling results training. **(This is an output-level indicator.)**

Sample Result 2: Political parties draft issue-based campaign platforms.

Sample Indicator 2.1: Number of political parties that identify and prioritize voters' issues as part of developing issue-based outreach strategies during a training. **(This is an output-level indicator.)**

Sample Indicator 2.2: Number of political parties that draft issue-based constituent outreach strategies to guide 2011 parliamentary election campaigning. **(This is an outcome-level indicator.)**

Sample Result 3: Political party campaigns address constituent concerns.

Sample Indicator 3.1: Number of political parties that articulate messages informed by organization X's public opinion polling results in the pre-election period. **(This is an impact-level indicator.)**

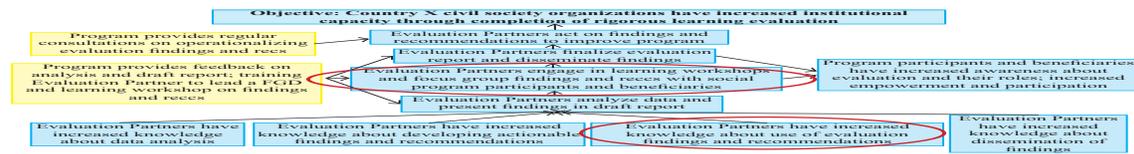
Note that while sample indicators 2.1, 2.2 and 3.1 are quantitative (they ask for the number of political parties), the data collection method will be qualitative – based on observation. For example, sample indicator 2.2 reports the number of parties that have outreach strategies (quantitative data), but the data must demonstrate that the strategies are indeed issue-based – this is qualitative data.

➤ ***Field staff input into the matrix is critical: field staff will need to be a reality check. Can you really get that data? Is the indicator off-base, too ambitious, not ambitious enough? Can you achieve these in the time frame and on budget? Does the target make sense? Is that an accurate baseline figure?***

Choosing What to Measure: Which Indicator is Best?

Now that you know what an indicator is, you can decide what you want to use them for: what you want to monitor and/or evaluate. It is often difficult to decide what to measure or evaluate. Indeed, the hardest part of M&E is not how to collect data, but choosing what to measure!

An easy way to figure out what you need to measure is first to figure out your options. Look at the results chain you have already developed. Select a range of key results, from output to outcome, and where possible, impact (if your program has that expected) to measure with indicators. For example, in the following results chain, two results are circled:



Determining How to Collect Data to Support the Indicator

M&E activities are what you do to collect data to measure and demonstrate achievement of your objective and to report to the individual indicators and evaluation points in your M&E plan.

Some sample tools for data collection are:

- Focus group discussions
- Surveys (such as public opinion polls)
- Internally-administered surveys (such as pre- and post-training questionnaires)
- Scorecards and checklists
- Observation
- Documentation review
- In-depth interviews
- Sign-in sheets from trainings

Step 6: The M&E Indicator Matrix: What It Is For, and How to Fill It Out

The indicator matrix links results and indicators with larger donor objectives, and lays out data collection methods and instruments, baseline and target information, and the data collection timeframe of your program.

Why do you need an indicator matrix? Because:

- It's often required by the donor.
- It shows you in one place how you are going to measure your success – or track your shortcomings.
- It lays out, in a simple format, what you are reporting.
- It helps keep you on track.

If you have already completed the results chain and identified indicator language, filling out the indicator matrix is only a matter of detailing the data collection method.

Filling out the Indicator Matrix

Below are details of what is entailed under each category of the indicator matrix.

Indicator: Write the indicator following the indicator guidelines above.

Indicator Type: Indicator type is the level of result it is measuring: output, outcome or impact.

Indicator Source: If you developed the indicator, write “custom” for the source. If it is drawn from another source (such as standard indicator required by the donor), note that source.

Indicator Definition: Clearly explain all aspects of the indicator that are not immediately obvious. There should be no ambiguity. For example, if your indicator is “Number of nominated civil society organization observers participating in election observation trainings,” you should define what a nominated civil society organization observer is, and what you mean by election observation trainings. For example: “Nominated civil society organization observers are individuals who are selected by their organizations as potential election observers but may not be accredited yet; election observation trainings are conducted by the program and cover observation techniques and election process regulations.” This may seem redundant, but it’s important to clarify what you mean and what you intend to do. This is especially important if you discuss such things as increased capacity. Further, the indicator definition is important internally, as in cases of high staff turnover, the indicator definition can, if properly completed, remove any chance that staff will interpret and report on the indicator in different ways.

Unit of Measurement: What are you measuring? If it’s “Number of political parties,” the unit is “political party.” If it’s “Number of civil society organizations that demonstrate improved organizational capacity” the unit is “civil society organizations.”

Disaggregation: If you are measuring individuals, as a rule of thumb, always disaggregate by gender, and if possible also by age (many programs include components aimed at encouraging youth participation, so capturing this detail is relevant), minority status, region, etc. If you say you are disaggregating, you need to follow through with that and report the disaggregated numbers in all of your reports.

➤ ***If the indicator is disaggregated, then make sure the indicator baseline and target data are also disaggregated.***

Data Source: Where is your data from? How will you get it?

Explain the answers under data source. For example, if you are collecting information on how many people attended your training, the data source may be “activity sign-in sheets.” If you are looking at increases in voter registration, the data source can be “National Election Commission voter registration list.” If you are measuring whether or not parties were able to use polling results to inform party messages, and you conducted interviews to ascertain this, this section would explain your method, which is “in-depth interviews.”

Data Collection and Analysis Method: How will you collect data and what sort of analysis method will you use? For example: “organization X will conduct in-depth interviews with 10 percent of the

participants three months after the training is completed. Organization X will use an interview protocol for all interviewees. Organization X will record each interview and transcribe the data. Then, organization X will conduct analysis of the interviews to ascertain major results of the training program and generate any program-relevant findings.”

Potential Data Source and Collection Challenges: Does your data source depend on cooperation from a political party, and how will that limit your indicator updates? Does your data collection depend on whether civil society activists are able to fill out pre/post-tests?

Use of Data: Explain how the data will be used. For example: to inform programmatic decisions, to measure progress towards the objective, and so on.

Frequency: Explain how often the data will be collected. After each training? Quarterly? Annually? Make sure the frequency is logical given the indicator definition and data source. For example, while participants can be counted quarterly, election results cannot.

Responsible Party: Explain what position will be responsible for the data collection and analysis. Staff member? Sub-grantee? Please be as specific as possible (but use the position title rather than the individual’s name).

Baseline: The baseline data tells you where you are starting from. Please spend time thinking about this and be as accurate as possible. If your indicator measures an increase in voter registration, your baseline data should come from the previous election. Please include the baseline year. If the indicator measures the number of participants at a training, the baseline should be zero, as you have not yet conducted any trainings as part of this program. The baseline year is the year the program begins. Never leave the baseline information blank.

Target: Similar to the baseline, the target is *key* to measuring your progress. Please note that it is okay if you do not achieve your target: there will likely be external factors that determine your success in achieving or surpassing targets. The important thing here is to give yourself and your donors an idea of the scale of achievement you expect. However, consistently outperforming or underperforming your targets may raise questions from donors as to whether the program is being implemented correctly. Remember also to explain how you came up with the target number.

Step 7: Including a Graphical Framework

A graphical framework illustrates all critical program components in such a way that you can clearly identify expected results. It’s a graphical representation of a complex process.

There are several different types of graphical frameworks. The results chain you discussed earlier is *one* kind, but it is usually quite messy and includes more detail than typically appear in other frameworks.

Most frameworks are used to showcase the important results and activities, and usually focus on the more intermediate-level results. Most graphical frameworks include these components:

- Your donor's objective
- Your objectives
- Your key expected results
- Your activities
- Assumptions

Note that once you simplify, you lose some of the underlying theory of the program. This is why it is critical to develop the results chain first to help you with your proposal, and only at the end to reduce it to a tidy version for submission with your proposal. Please see the sample frameworks (figures two through seven) provided at the end of this chapter for different examples of frameworks.

Step 8: What to Say in the Proposal Narrative about the M&E Plan

The program evaluation section provides a narrative description of your approach to M&E for that grant. It is the main discussion of your M&E approach; the M&E indicator matrix is an accompaniment to this section. This section should include a discussion of each of the following components. It can be organized as outlined for clarity. Each section has questions to help you develop and write the narrative:

Expected Results: What are the expected results of this grant? How do they fit with the project's program theory?

Management and Coordination of M&E: Who is responsible for M&E? Who will coordinate M&E activities? Will local partners have a role? What about contractors or consultants? How will M&E requirements be managed?

Assessments: Will the program be conducting any special assessments, such as a comprehensive baseline, midline and/or endline, or a political party, civil society or governance assessment? Describe the purpose and the initial design.

Indicators: Which indicators are qualitative? Which are quantitative? How many are output-level? How were outcome indicators selected?

Data Collection and Monitoring Tools: How will data be collected? What tools will be used? Will they be piloted? How will you try to reduce bias and ensure rigor?

Data Verification: How will the program ensure that data is accurate? What processes and checks will be put in place?

Ethics and Rigor: What sort of ethical issues are present (i.e. identity of participants)? What steps will be taken to reduce bias?

Data Security and Storage: How will data be stored? Where? Who will have access? Is there a timeframe for how long you will keep the data on hand? How will you ensure data is secure?

Analysis, Data Utilization and Learning: How will you analyze data? How will you use data? Will any data be used for program learning?

Reporting: What is your reporting schedule? To whom will you report?

Critical Assumptions and Risks: What assumptions do you have that govern your program and your M&E plan?

Sustainability: What efforts will you make to ensure sustainability of program results? For example, if results are knowledge and skills, efforts to ensure sustainability could include the dissemination of training materials and books that can be used for future reference.

Consider a More Rigorous Program Design

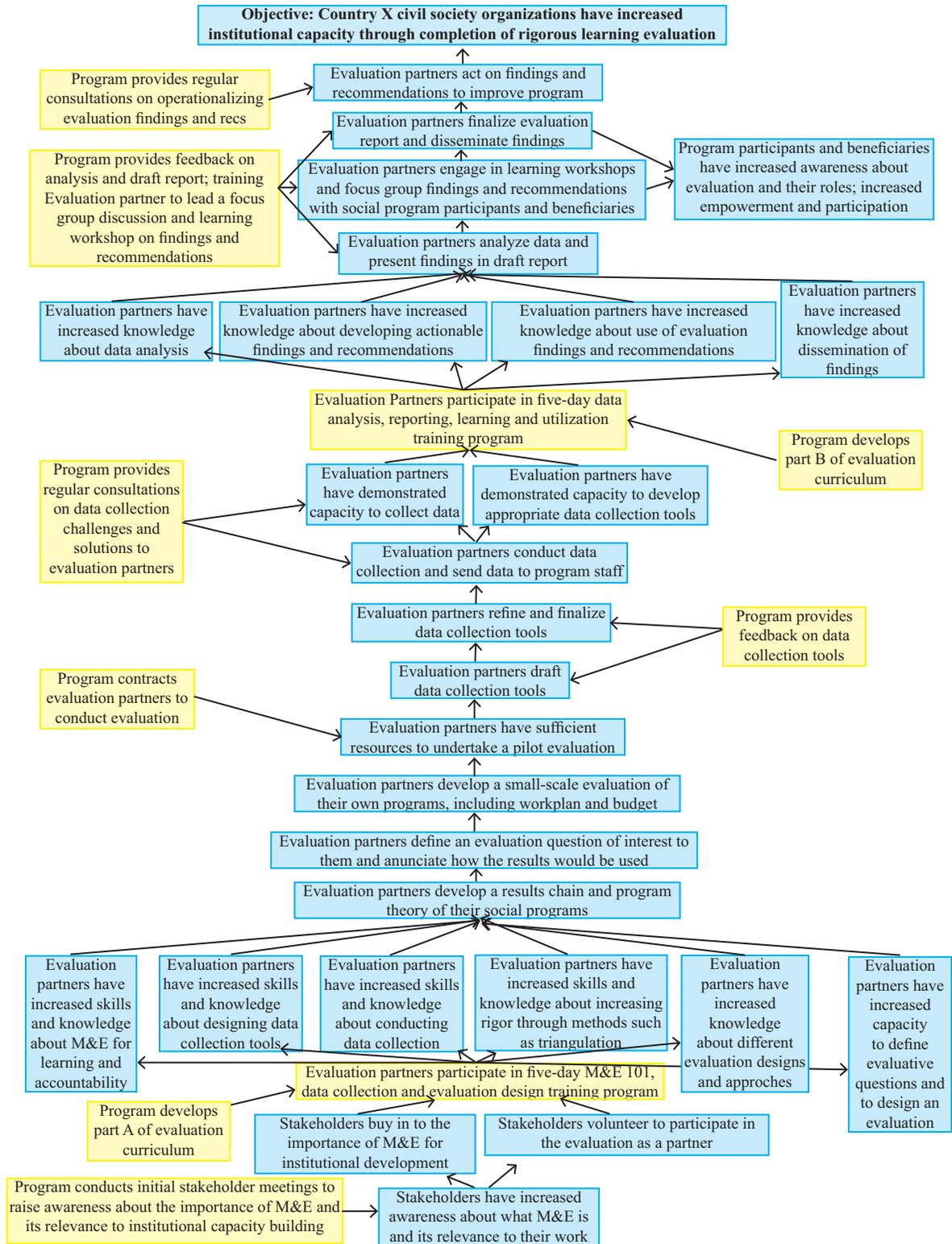
You often do not have time to take a more rigorous approach to designing your programs, simply because time allotted to proposal development is often scarce! This is regrettable, since there is a lot of learning from other programs that can go into the design of a new program, as well as many M&E methods to ensure that the program you design is as logical, efficient, effective, productive and results-based as possible. Figure two lays out some rigorous program design approaches that you should consider using if time and resources permit!

Figure Two: Evaluation Methods

Method	In What Situation Would it be Used?	What It Is For	Ideal Timeframe
Needs Assessment	If you are not sure what areas partners need assistance in, how to craft activities in order to most effectively achieve activity goals, who important stakeholders are, etc.	Determines needs of the partners, how to fill that need, and establishes standards by which the program is evaluated separately.	Proposal design stage ideally, but could also be done at any time in the program lifecycle. Also, toward the end of a grant when you know there will be an extension or a new grant. It requires several days of interviews and consultations.
Formative Country Program Assessment	If you are developing a strategy for a new country or region, or embarking on a new sector of support, and are not entirely sure what to look for or what to prioritize.	Helps prioritize country and program goals, needs, entry points, problems to address and interests. It differs from a needs assessment by taking a more macro approach, and also because, in this case, the program likely knows very little about the situation.	Ideally before the proposal process, to help develop a larger strategy.
Outcome Mapping	If you don't know what your underlying theory of change is, if you want to develop a vision and work up from activities. Developed by the International Development Research Centre.	Outcome mapping provides an excellent, detailed process to comprehensively lay out the strategic vision, mission, expected results and activities of a project, as well as indicators. It can also help a program focus on its sphere of control, influence and interest.	Proposal design stage through the end of the program. It requires one to five days of effort at the outset.
Program Theory Framework	If you want to streamline the project, clarify activities and their immediate purpose, to help get all staff thinking and brainstorming together. An effective way to get multiple offices working together on a proposal plan.	Clearly lays out each program step, what each step achieves, and what internal and external factors will influence the achievement. Helps identify criteria of success on which to base indicators.	Proposal design stage. It requires one to two days working internally with the program team and, ideally, a program partner.

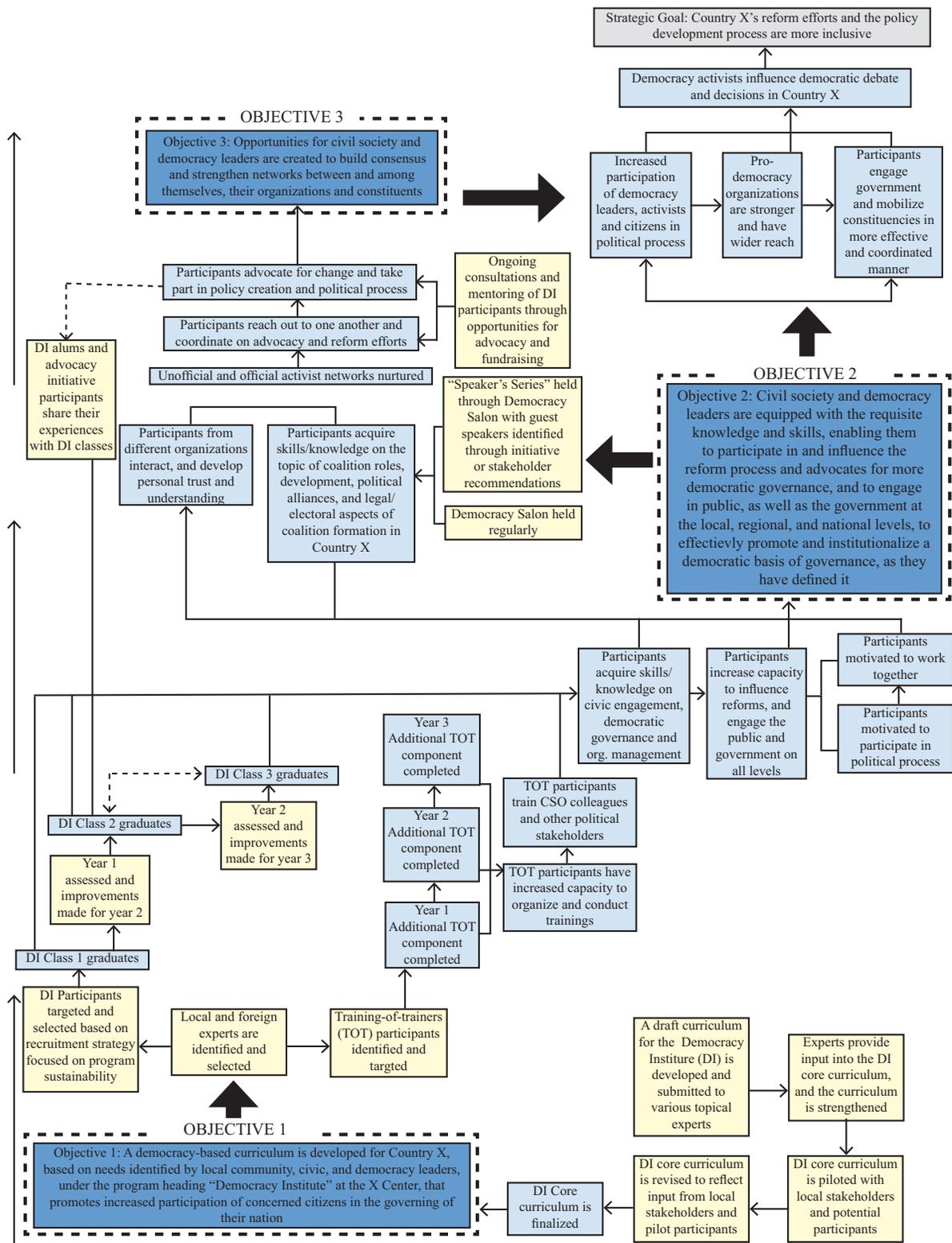
Source: IRI

Figure Three: Sample Results Chain



Source: IRI

Figure Four: Sample Results Chain



Democracy Institute (DI): Program Theory Framework
 (Accompaniment to the Program Strategy & Technical Approach, and the Monitoring and Evaluation Plan)

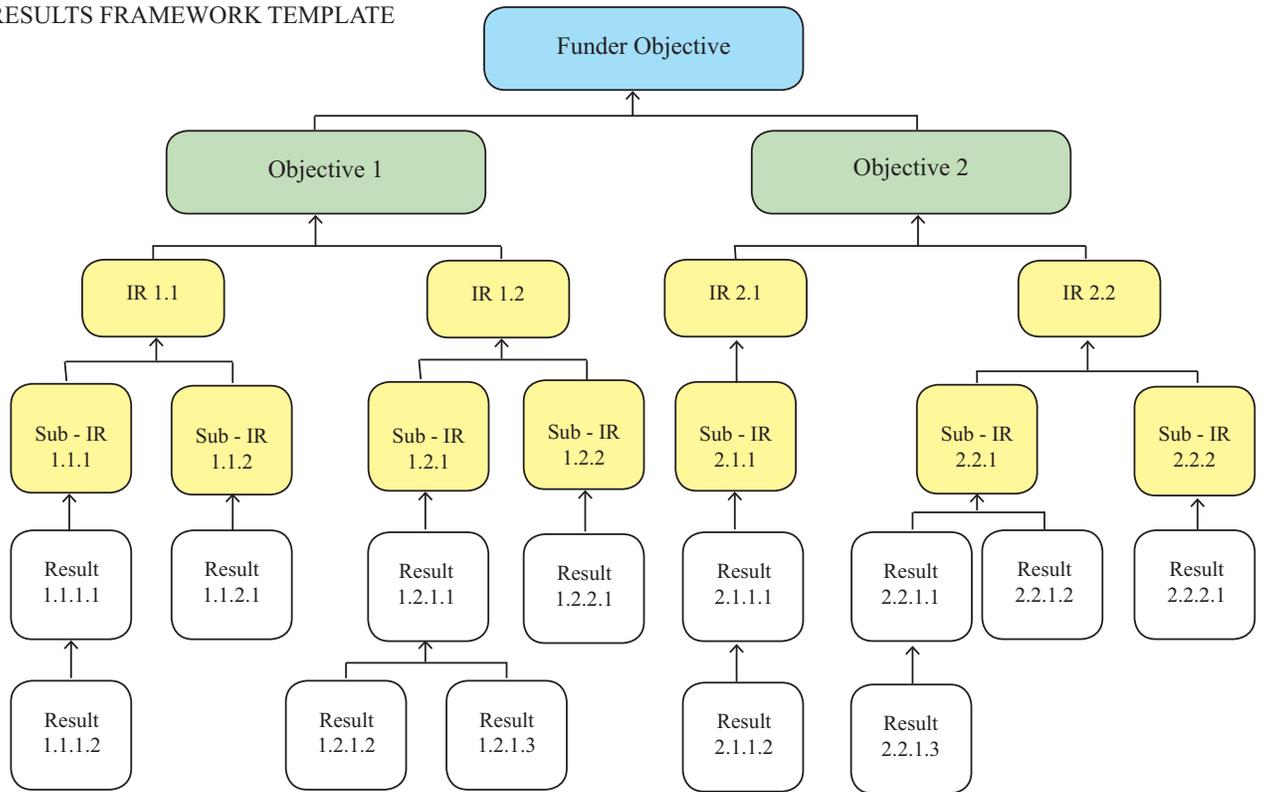
Figure Five: Sample M&E Indicator Matrix/Performance Monitoring Plan

Title of Project					*** Area shaded in blue to be completed after the grant is awarded								
Performance Management Plan													
Project Dates													
INDICATOR	INDICATOR TYPE AND SOURCE	DEFINITION, UNIT OF MEASUREMENT, DISAGGREGATION	CRITICAL ASSUMPTIONS AND RISKS	DATA SOURCE, COLLECTION, AND ANALYSIS METHOD	USE AND REPORTING		BASELINE	YEAR 1		YEAR 2		YEAR 3	
					USE OF DATA	FREQUENCY OF COLLECTION, RESPONSIBLE FOR DATA COLLECTION		TARGET	ACTUAL	TARGET	ACTUAL	TARGET	ACTUAL
OBJECTIVE 1													
Indicator 1.1	Type: Output, outcome or impact? Source: If the indicator is customized to the project, write custom. If it is an F-indicator, please use the corresponding number found in the indicator database at http://www.state.gov/f/indicators/index.htm .	Definition: Unit of measure: Describe the unit that is being measured (i.e., individual political party, civil society organization scorecard, etc). Disaggregated by: List out all variables that require disaggregation (for example: women, youth, minority status, religion, political party, province).	Briefly explain any critical assumptions and risks related to this result and objective.	Data Source: Explain where the data will be drawn from. Collection and Analysis Method: Explain what method will be used to collect the data and how the data will be analyzed. Potential Data Source and Collection Challenges: Briefly explain any challenges to data source, collection and analysis for this indicator	Use of data: Explain how the data will be used (for example: to inform programmatic decisions, to measure progress towards the objective, and so on).	Frequency: Explain how the data will be collected. Responsible party: Explain which institute and which position is responsible for collecting and using the data (e.g. field officer).	Clearly explain what the baseline is for the indicator; the source of the baseline data, and the year it was established.	Jan- March: State the indicator target for each. April- June: July- Sept: Oct- Dec:	Jan- March: State the actual figures for each. April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:			
Indicator 1.2	Type: Source:	Definition: Unit of measure: Disaggregated by:		Data Source: Collection and Analysis Method: Potential Data Source and Collection Challenges:	Use of data: Frequency:	Responsible party:		Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:
INTERMEDIATE RESULT 1.1: Intermediate results remain within the objective; think of it as an intermediate step to the objective.													
Indicator 1.1.1	Type: Source:	Definition: Unit of measure: Disaggregated by:		Data Source: Collection and Analysis Method: Potential Data Source and Collection Challenges:	Use of data:	Frequency: Responsible party:		Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:
Indicator 1.1.2	Type: Source:	Definition: Unit of measure: Disaggregated by:		Data Source: Collection and Analysis Method: Potential Data Source and Collection Challenges:	Use of data:	Frequency: Responsible party:		Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:
INTERMEDIATE RESULT 1.2:													
Indicator 1.2.1	Type: Source:	Definition: Unit of measure: Disaggregated by:		Data Source: Collection and Analysis Method: Potential Data Source and Collection Challenges:	Use of data:	Frequency: Responsible party:		Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:
Indicator 1.2.2	Type: Source:	Definition: Unit of measure: Disaggregated by:		Data Source: Collection and Analysis Method: Potential Data Source and Collection Challenges:	Use of data:	Frequency: Responsible party:		Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:	Jan- March: April- June: July- Sept: Oct- Dec:

Source: Consortium for Elections and Political Process Strengthening

Figure Six: Sample Results Framework Template

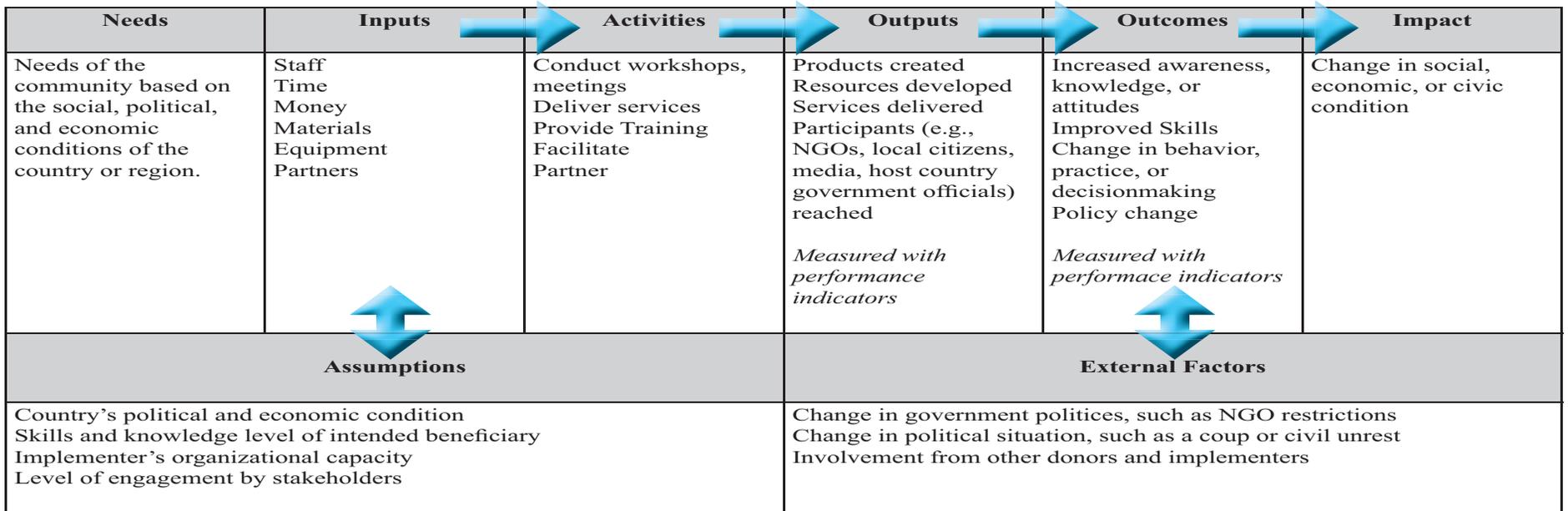
RESULTS FRAMEWORK TEMPLATE



Source: IRI; “IR” stands for “intermediate result” in the chart.

Figure Seven: Sample Graphical Framework 2

Logic Model Template



Source: U.S. Department of State's Bureau of Democracy, Human Rights and Labor

Needs Assessment Guide

A **needs assessment** is conducted to help define program objectives, strategy, partners, stakeholders and activities. Conducting a needs assessment is a key step in determining what your partners need in terms of programming, what might work to meet that need and establishing the standards by which you will evaluate whether the program made a contribution toward meeting that need.

Ideally, all programs would conduct a needs assessment. Practically speaking, consider conducting a needs assessment if any of the following situations apply to your program:

- The political context has changed significantly.
- You are starting work in a new country.
- You will be working with new stakeholders or partners.
- You are starting a new initiative or type of project that the program has not worked on before in your country.
- You are not sure who else is working on these issues or you are not sure what gaps in assistance there are.
- You are halfway through or toward the end of a program and you need to see if it should be continued or adapted.

The needs assessment ideally happens before a proposal is written, so that it informs proposal development. When you do not have time to undertake a needs assessment, it may happen directly after a grant is awarded, but it should take place before the implementation plan is developed and activities launched. The results of the needs assessments should inform the final strategy and implementation plan, and can inform how activities are designed and undertaken.

There are different ways of conducting a needs assessment, based on the objective of the assessment. Here are general needs assessment steps:

Step 1: Identify the people involved:

Evaluation expert E. Jane Davidson suggests considering needs assessment stakeholders in two groups:³

- *Immediate recipients* (your partners, such as elected officials or party members)
- *Impactees* (those people for whom something changes as a result of your program, such as the constituents of elected officials)

Step 2: Determine what method(s) to use:

There are several tools you can use to gather the data. Each tool should be used to elicit answers to all components of Step 3.

³ Davidson, E. Jane. *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: Sage Publications, 2005.

- *Direct Observation*: This is your experience and perspective gained from discussions and interaction with or witnessing program partners and other stakeholders.
- *Interviews/Focus Group Discussions*: With key stakeholders, such as staff, in-country experts, intermediary constituents, potential program recipients, partners and so on. Make sure the sampling size is slightly larger than you think you need to get a good mix of points of view.
- *Questionnaires*: Sent to stakeholders for written responses. Make sure to pilot the questionnaire first to be sure that the questions are clear and targeted.
- *Public Opinion Surveys*: Either previous surveys or existing surveys.
- *Other Existing information*: Academic literature, media reports, other funder reports, etc. can all be helpful information.

Step 3: Using the tools you selected, identify the needs of impactees⁴:

E. Jane Davidson offers the following typology to analyze needs:

- *Conscious needs*: What they know they need.
- *Unconscious needs*: What they haven't identified as needs.
- *Met needs*: Needs that have been met, to prevent program redundancy.
- *Unmet needs*: Needs that have not yet been met.
- *Performance needs*: Supporting needs to help address real needs (such as website training for an outreach website).
- *Instrumental needs*: Needs that help address performance needs (such as Internet connection or a computer for the website training).

Step 4: Once you have identified the needs of impactees, analyze the data to answer the following questions:

- What do your partners and impactees need?
- What can you do to meet those needs?

The answers to these questions will determine the program's relevance and help define the scope of your program going forward. Ideally, the individuals conducting the assessment should write a report, to highlight the key data and conclusions so that programming decisions can be made.

Formative Country Program Assessment

Sometimes you have the opportunity to travel to the field to prepare for a proposal. It is especially a good idea do this for countries where you have never worked, countries where you have been absent for some time, countries where you have had programs but not had a field office, or countries where you are taking on a new kind of program or working in a new region.

➤ ***Use a mix of these methods. At the very least it is recommended that the needs assessments include interviews and focus group discussions!***

⁴ Ibid.

These situations are characterized by any of the following:

- Little institutional knowledge about the country, region or topic.
- Little institutional knowledge about working on the ground in the country or region.
- Few stakeholder contacts or entry points.
- Uncertainty about the best programmatic approach.
- Uncertainty about what can be realistically achieved.
- Uncertainty about the degree of influence of external factors or what these would be.

A country program assessment can really take any form. One way to do it is to use a **rapid appraisal method (RAM)**.⁵ This method is simply a way of using multiple data collection methods in such a way that you get as much information as possible in a limited timeframe. It is less rigorous than standard data collection, but for a fast program assessment that aims to get a balance of views in order to inform program direction and approach, it is sufficient. If you have more time on the ground, then think about the following steps to develop your program assessment:

Prior to travel:

1. *Determine your main topical interest:* Party work? Civil society development? Democratic governance? Etc.
2. *Determine the kinds of information that is needed to know:* Examples include main actors, relationships, existing strategies, entry points, needs, political context and external factors.
3. *Determine key institutions and actors for your areas of interest:* Consult with country experts, scan media reports and journal articles. Develop a list of stakeholders.
4. *Determine what method to use to gather information from each stakeholder:* Group interviews? Individual interviews? Survey?
5. *Develop tools:* For interviews, develop open-ended questions to probe for different viewpoints, and to allow for stakeholders to prioritize their needs and interests. For questionnaires and surveys, make sure there is plenty of room for free and open response.

Once in the country:

6. *Meet with stakeholders:* Record all responses carefully, and ask for other contacts.
7. *Meet with newly identified stakeholders:* Adapt interview questions from the original interview guides.
8. *Observe:* Observe what is happening around you, make notes and record your reflections. These are important data points. If possible, take part in something, such as a town hall meeting organized by a potential partner.
9. *Monitor local news:* See what is prioritized in the local media, and observe how events or issues are being discussed.
10. *Talk to random citizens:* Hear what the average person is thinking or what the average person knows about your topical interest.
11. *Develop an initial list of findings:* As you meet with people and observe things, keep a running list of findings and a note about what sparked that finding.

⁵ Vondal, P. *Performance Monitoring and Evaluation Tips: Using Rapid Appraisal Methods*. Washington, D.C.: USAID Center for Development Information and Development Evaluation, 2010.

12. *If possible, test your nascent idea/approach with stakeholders before leaving:* After meeting with different people and being on the ground, you should have a fair idea about what you think needs to happen. Time permitting, get a group of your most informative stakeholders together and float some ideas to get their feedback and their ideas. This will prove invaluable when you return from the trip and need to put things on paper.

After travel:

13. *Finalize your list of findings:* If possible, focus group the major findings with regional experts to get a reality check.
14. *Develop your recommendations:* Similarly, float the major recommendations to outside experts to see if they have additional ideas for entry points or approaches.
15. *Develop your programmatic approach:* Now you are ready to develop your programmatic approach, which entails identifying your objectives and developing your program theory!

Outcome Mapping Guide⁶

Outcome mapping is an approach first developed by Sarah Earl, Fred Carden and Terry Smutylo at the International Development Research Centre (IDRC). They are the authors of *Outcome Mapping*, a handbook developed by the International Development Research Centre (IDRC). The handbook has step-by-step instructions on how to go through the process, and is available for free on their website: <http://www.idrc.ca>. The following section was informed by the IDRC handbook as well as other online resources on outcome mapping.

Since its creation, outcome mapping has been applied in many different ways. However, at its core, outcome mapping provides an excellent, detailed process to comprehensively lay out the strategic vision, mission, expected results and activities of a project, as well as indicators. It also can help a program focus on its sphere of control, influence and interest.

Outcome mapping acknowledges that impact is difficult, if almost impossible, to achieve, due to the complex environments in which you work. For this reason, rather than laying out explicit theories, which you may not know, outcome mapping focuses on your vision and builds up from your proposed activities. Instead of identifying indicators to measure and evaluate your work, through outcome mapping you develop graduated progress markers which essentially come from asking: what do you expect to achieve from your activity? What do you hope to see? What would you love to see? For the purposes of your donors, you can turn answers to these questions into indicators or evaluation points.

Outcome mapping should be completed before a program starts. It sets up a monitoring plan that focuses on behavioral changes and implementation strategies, with a focus on measuring the changes in behavior that correlate with each change of state. It takes a participatory approach and creates outcome ladders (dynamic indicators that have gradations of progress) that will allow the program to see how those impacted by programming are progressing towards change.

⁶ Earl, Sarah, Fred Carden, and Terry Smutylo. *Outcome Mapping: Building Learning and Reflection into Development Programs*. Ottawa: International Development Research Centre, 2001. The work is summarized here with permission from IDRC.

Below is a list of steps that are undertaken as part of this process.

Outcome Mapping Steps for Proposal Design:

You begin with a facilitated discussion that allows stakeholders to reflect on how best to measure the program. Stakeholders are determined on a case-by-case basis and could include: program staff (from headquarters and the field, if applicable), program partners (such as political parties, elected officials, etc.), donors and ultimate beneficiaries (such as constituents).

Through the discussion, which can take several days, the following are produced:

1. *A vision statement:* The ideal goal or final result toward which you're working. What would things look like if all of the problems you are working to solve were solved? The achievement may be beyond the immediate capability of the current program, but the program could make a contribution toward the vision.
2. *A mission statement:* Describe how the program intends to support the vision. This is an ideal statement on how the program will contribute. This will include how the program intends to remain effective, efficient and relevant.
3. *Identification of boundary partners:* With whom do you work? Who are the people that will bring about change? Political parties, civil society organizations, elected officials, etc.
4. *Development of outcome challenge:* Describe how the behavior of the boundary partners would change if the program was extremely successful (i.e., if the program had achieved its full potential as a facilitator of change).
5. *Development of graduated progress markers:* These are markers that will help the program determine if boundary partners are progressing towards outcomes. This allows the program to gauge what it has accomplished, while still seeing what needs to be achieved. These are categorized into "expect to see," "like to see," and "love to see" – the idea being that the program team identifies minimum progress as well as ideal progress (to allow for gradations of progress).
6. *Creation of a strategy map for each outcome challenge:* Identify strategies the program will use to contribute to the achievement of the outcome.
7. *Articulate organizational practices:* This helps the program look at how it is going to operate to fulfill its mission.

At this point, to conform to your donor requirements, it is helpful to use the graduated progress markers to develop indicators or evaluation points.

The actual outcome mapping process, though, does not use indicators, and instead focuses on monitoring outcomes and performance. The goal is to assess a program holistically, recognizing that a program needs to know not only about outcomes, but also the processes by which they are attained and the program's own internal effectiveness to achieve them. The goal is to monitor:

1. Behavior changes in boundary partners by creating an **outcome journal**: To collect data on the boundary partners' achievement of progress markers.
2. Strategies the program is employing by creating a **strategy journal**: To collect data on the program's actions in support of the boundary partners.

3. Functioning of program as an organizational unit by creating a **performance journal**: To collect data on how the program team works to remain relevant, innovative, sustainable and connected to its environment.

According to the outcome mapping approach, looking at how these three elements interrelate and the context in which change occurs is essential to program learning. By capturing information along the way, these journals become data collection guides, with which programmatic decisions are made.

*Program Theory Framework Guide*⁷

An effective way of laying out the program before writing the proposal is by developing a **program theory framework**. This is a simple table that helps you by clearly laying out what goes into a program and directly linking each programmatic element to the different levels of results you are trying to achieve.

Why would you want to do it? If you have enough time, this is an excellent way of thoroughly thinking through the program steps and getting all staff members to participate in the analysis of proposed results and factors that go into achieving them.

It is based on a pre-determined results chain, so this should be done after a problem statement, program theory and results chain have been developed and agreed upon.

Information found in this table helps streamline program design and organize a more effective and convincing proposal. It helps identify what needs to happen for the program to have an influence on each step of the results chain. It is also helpful for those who are not familiar with evaluation principles, because it lays out all possible areas where data can be collected (see Step 5: Developing Effective Indicators). See figure eight for an example.

⁷ This program theory framework approach was developed by Sue C. Funnell; for more information, please see: Funnell, Sue C. "Developing and Using a Program Theory Matrix for Program Evaluation and Performance Monitoring," Program Theory in Evaluation: Challenges and Opportunities. Spec. issue of New Directions for Evaluation. Ed. Patricia J. Rogers, Timothy A. Hacsí, Anthony Petrosina, and Tracy A. Huebner. 2000.87. (Fall 2000): 91-101.

Figure Eight: Program Theory Framework Example

RESULT CHAIN STEP	SIGNS THIS STEP HAS BEEN ACHIEVED	WHAT WILL YOU DO TO ACHIEVE THIS STEP?	WHAT OTHER FACTORS WILL AFFECT THIS STEP?		HOW TO MEASURE ACHIEVEMENT OF STEP (INDICATOR)	DATA NEEDED TO MEASURE ACHIEVEMENT?
			WITHIN PROGRAM CONTROL	OUTSIDE PROGRAM CONTROL		
(1) Party members participate in issue-based message training. 	1. Full attendance 2. Leadership or decision makers attend 3. Active participation 4. Participants walk away satisfied	1. Public opinion survey conducted 2. Survey methodology and survey results training 3. Issue-based message/campaign development trainings 4. Solicit buy-in from leadership	1. Engaging training design 2. Qualified and effective trainer 3. Environment conducive to learning (room set-up)	1. Buy-in from leadership 2. Political environment (ability to speak freely, ability to travel to a training, etc.)	Number of people trained Number of leaders in attendance Active engagement buy-in from leadership attained	Attendance records/participant lists Direct observation Written/verbal acceptance from leadership
(2) Party members gain knowledge/skills in issues and issue-based messaging techniques. 	1. Increase in knowledge about issues and issue-based campaigning 2. Ability to draft issue-based strategy 3. Ability to conduct campaign	(There is no corresponding activity.)	(There are no factors within the program's control.)	1. Level of education	Number of participants with increased knowledge of issue-based messaging techniques	Pre and post-training questionnaire Knowledge scorecard
(3) Party members are more motivated to perform issue-based campaign. 	1. Participants discuss using issues in campaigning with leadership 2. Participants build buy-in among colleagues	(There is no corresponding activity.)	(There are no factors within the program's control.)	1. Level of pro-activeness 2. Level of empowerment 3. Level of power/influence	Number of internal party trainings, consultations, or meetings regarding issue-based campaigning	Survey Focus group discussions with party members
(4) Parties/candidates develop issue-based messages and campaign strategy. 	1. Priority issues identified 2. Issue-based message developed 3. Issue-based campaign strategy developed	1. Individual consultations on polling data 2. Individual consultations on message and strategy development 3. Constituent outreach, public speaking trainings	1. Engaging consultation design 2. Qualified consultant 3. Perception of consultant expertise/legitimacy 4. Skills/knowledge based on prior trainings	1. Buy-in from leadership 2. Level of education	Number of people trained Number of consultation requests and/or meeting Number of messages tested with voters prior to campaign	Document/content analysis Participant/consulting lists Focus group with voters to test messages Interviews with leadership
(5) Parties/candidates campaign on issues of concern to voters (program objective) 	1. Consistent use of issue-based message throughout campaign 2. Issues debated between candidates or parties 3. Less negative campaigning	(There is no corresponding activity.)	1. Skills/knowledge based on prior trainings	1. Buy-in from leadership and party members 2. Level of pro-activity of candidates 3. Political environment (ability to campaign freely, electoral framework etc).	Number of parties campaigning off issues Number of debates on issues IRI polling results used in campaign messages	Staff observation of campaigns Interviews with candidates or party leadership
(6) Citizens ID party/candidates by issues. 	1. Voters can differentiate between parties based on issues	(There is no corresponding activity.)	(There are no factors within the program's control.)	1. Political environment (ability to campaign freely, etc). 2. Free and open media; relatively equal media coverage	Voters can differentiate between party messages	Focus group discussions with intermediate constituents (academia, CSOs, media) Surveys before and after the campaign period
(7) Citizens decide which party/candidate to vote for based on issues. 	1. Voters select party/candidates based off issues on which the party/candidate campaigned	(There is no corresponding activity.)	(There are no factors within the program's control.)	1. Political environment (ability to vote freely, privacy of the vote, etc). 2. Free and open media; relatively equal media coverage	Increase in votes for parties that use issue-based campaigns Increase in percent of voters who think parties stand for something	Pre- and post-election polling/survey Election results

Source: IRI, based on template developed by Sue Funnell

Chapter 3: M&E in the Field: Measuring Success with a Proper M&E System

A program’s M&E system describes how the program will use data to reflect on and communicate information to help implement the program, adapt to changing environments, report to donors and inform future programs. Remember, information is not helpful if it is not used!

A proper M&E system will include the following components:

- Established **Purpose and Scope**
- Established **Data Collection** Process
- Established **Data Analysis** Process
- Established **Data Use**, with a focus on learning, communication and reporting
- Established **Data Storage** Protocol
- Established **Ethics**

Some programs undergo assessments of their data quality, called **data quality assessments**. This section will discuss these as well.

Established Purpose and Scope

After the grant is awarded, ask yourself the following questions:

- What do I really want to know about my program?
- What do I need to know to be sure I am on track?
- What do I need to know to be sure I am achieving my results?
- How can I collect information efficiently to know that?
- What kind of staffing and resources do I need to gather and process that information? Do I have enough of these resources?
- Do I need any formal assessments or evaluations, such as a needs assessment, stakeholder assessment or anything else?

Note: M&E “system” versus “plan”

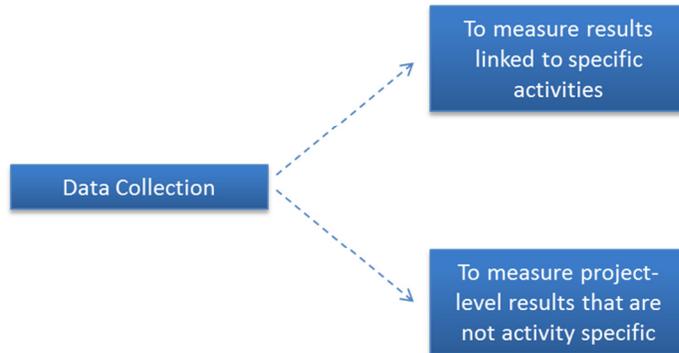
Your M&E system is whatever system you have in place—formal or informal—to manage the collection, utilization and dissemination of information about your program.

The system should be formalized into a written M&E plan specific for each grant/project. Many funders require a formal M&E plan as part of a grant/contractual agreement.

Once you have answers to these questions, review the indicator matrix, and refine your indicators and related information. Then, address your M&E needs and activities in your program workplan, or develop an M&E workplan.

Data Collection

You gather data to understand two things: (1) where you are in implementing your program (i.e., how your activities are going, which is basic monitoring) and (2) changes in behavior/state due to your work (i.e., project level results that are not specific to a particular activity, which is more evaluative but also important for monitoring).



Immediate results linked to specific activities tend to be outputs and sometimes outcomes. Project-level results tend to be at the outcome and impact level.

Data helps you make program decisions, learn from program results and inform others of your progress and results.

The M&E plan, including the indicator matrix, constitutes the minimum requirements for monitoring your performance and measuring your results.

What Do I Need to Consider?

Data collection takes place throughout the life of a grant. It often begins before a program starts (e.g., establishing a baseline or conducting a needs assessment), and continues after program activities have concluded, to collect information on program results and to inform the final report or future programs.

There is no single best way to collect data. In fact, it's generally advisable to use a mixture of techniques to ensure objectivity and **triangulate** data (using different researchers, data types and data sources to corroborate findings). Below are some questions to consider with regard to data collection as you prepare activities and look through your work plan and M&E system:

- Do I need a baseline assessment? An endline assessment?
- Do I need to pilot test my data collection method?
- What are my available resources and time?
- When do I need to start collecting data for an indicator, and how often must I collect it?

➤ *There has been a trend in recent years toward quantitative data: it's often easier to manipulate and understand. For lower-level results like outputs, it can be straightforward: number of people trained. But this number does not tell you anything about the larger system you are trying to change. In addition, collecting quantitative data for higher level results can be difficult, and sometimes misleading; quantitative figures often depend on qualitative data and analysis. For monitoring purposes, quantitative data is usually used to gather information on lower-level results like outputs, while qualitative data is often used to describe higher-level results.*

➤ *Each time you do a program activity, ask yourself: can I use this opportunity to collect data?*

- How will I use the data I collect?
- Who will analyze the data?

Data can be qualitative, quantitative or a mixture of both. Qualitative data uses words, pictures, video and narrative to describe something. Quantitative data uses numbers to describe something. Neither is better than the other: it depends on what you need!

Here is a breakdown and comparison of each method:

Comparison of Data Collection/Analysis Methods				
Method	What it is	Example Data Collection	Benefits	Drawbacks
Quantitative	Use of numbers to tell a story	Surveys, multiple choice questionnaire, counting, etc.	<ul style="list-style-type: none"> • Responses are fixed so in-depth analysis is not required (unless inferential statistics are required for analysis of surveys), though it requires extensive preparation at the outset to define variables. • Able to aggregate vast sums of data into manageable information. 	<ul style="list-style-type: none"> • Extensive training required for anything outside of basic multiple choice questionnaires. • Often based on closed-ended data collection tools which cannot capture unintended consequences. • Often difficult to understand variables. • Not flexible.
Qualitative	Use of words and narrative to tell a story	Interviews, focus groups, observation, content review, pictures	<ul style="list-style-type: none"> • Captures unanticipated stories. • Can answer “how” and “why” questions. • Very flexible, can integrate complex stories together. 	<ul style="list-style-type: none"> • Significant work. • Can be difficult to find trends in data. • Difficult to boil down in presentations.
Mixed Methods	Use of quantitative and qualitative data to tell the story	Mixing a post training questionnaire with a group interview	<ul style="list-style-type: none"> • More robust than a single method. 	<ul style="list-style-type: none"> • Takes more time and requires more data collection, often from the same participants.

Making Sure the Data is Actionable

Collecting data is meaningless if the findings are not used. When planning for data collection, make sure you ask yourself: how will I use the data? If you can’t give yourself a specific answer, then you probably won’t use the information.

Examples of data use:

- Report to indicators for accountability.
- Tells you if you are on target to meet your objectives.

- Tells you if you are on the right track: e.g., if questionnaire results show that participants do not fully understand one of the training topics, perhaps you should rethink the way you present the content in future trainings.
- Helps you learn from what you are doing in order to inform future programs.

Examples and Guides to Common Data Collection Methods

Now that you have thought about the bigger picture of your data collection, you can turn to data collection methods. These methods can be used for any level of result: output, outcome or impact. The level it measures depends on how you use the method.

There are many different places where you will get data to report to your indicators.

Often a program will have access to existing program data (**primary data**), such as: participant lists, web statistics, survey results, etc. Or a program will have access to data from other groups (**secondary data**), such as: reports by other organizations, news clippings, etc. The kind of data you have access to will depend on your project type and the environment in which you work.

To collect new primary data there are common M&E data collection methods that most projects can use. Below is a table showing the differences in methods.⁸ These methods can be used for any level of result: output, outcome or impact. The level it measures depends on how you use the method.

⁸ IRI would like to acknowledge the efforts of Social Impact, Inc. in helping to develop this table.

Common Data Collection Methods			
<i>Refer to the Data Collection Section for guidelines and tips for each of these methods.</i>			
Method	Uses	Strengths	Challenges
Public Opinion Survey	<p>Polls are an ideal tool to measure the progress of a program, including its reach and impact. Polls are best with programs that reach a large part of the population or have been working for years on a single issue and seen more systemic change.</p>	<ul style="list-style-type: none"> Numerical responses enable statistical analysis. Question design can be controlled. Can reach many people quickly and efficiently. Large sample size means results typically representative of bigger population. Anonymity of instrument might improve honesty of respondents. 	<ul style="list-style-type: none"> Can be expensive to administer. Data analysis can be technical. If not designed correctly, questions may not adequately capture actual impact. Can only capture results of programs that have broad reach or scope.
Self-administered surveys (pre-/post-training questionnaires)	<p>Questionnaires can be an ideal means to measure various things, such as the effectiveness of trainings, to collect data on participants, or to test changes in knowledge.</p>	<ul style="list-style-type: none"> Inexpensive Anonymous 	<ul style="list-style-type: none"> Depends completely on proper/unbiased question design and structure of responses. Cannot capture unanticipated responses.
In Depth Interviews (Individual or Group)	<p>Interviews help you to gain a deeper understanding of program outcomes, gauge why and how a program might (or might not) have had an impact, and learn more about the experiences and perceptions of stakeholders.</p>	<ul style="list-style-type: none"> Able to enter into another person's worldview. Can explore complex issues in depth and capture unanticipated results. Flexible - questions can be altered as interview progresses. 	<ul style="list-style-type: none"> Difficult to conduct effectively. Time consuming. Subject to interviewer bias. Difficult/time-consuming to analyze interview data. Sample size typically not representative of population. Can be expensive.
Focus Group Discussions (FGD)	<p>FGDs are useful for efficiently and relatively inexpensively capturing the perspective of a select group of people. Useful in capturing the interaction between people. Sampling is an important consideration for FGDs.</p>	<ul style="list-style-type: none"> Efficient - can capture the perspective of several people in one session. Group discussion helps to stimulate memories and past experiences. 	<ul style="list-style-type: none"> Difficult to schedule people together. Planning of session takes considerable effort. Focus can be lost. Candor may be more difficult during group sessions. Requires a skilled moderator.
Participant Observation and Visual Evidence	<p>Useful to directly observe changes as a result of your programs and assess whether partners use techniques learned during training. Goal is to systematically view an activity until patterns become apparent.</p>	<ul style="list-style-type: none"> Can directly experience an activity. Can capture aspects of an activity, such as intensity and perception. Do not have to rely on others' account of an event. Visual evidence provides quick descriptions of a physical environment. 	<ul style="list-style-type: none"> Requires disciplined and prolific note taking. The observer can often influence behavior. It is often hard to interpret or categorize behavior correctly. Observer will inevitably have own views/perceptions of activity and setting (difficult to be objective). If possible, have two simultaneous observers to reduce bias.
Documentation Review (Content Analysis)	<p>Review documents to discern patterns.</p>	<ul style="list-style-type: none"> Find patterns in large volumes of information. 	<ul style="list-style-type: none"> Can be difficult and time consuming to find and review large volumes of documents. Difficult to analyze and categorize information. Subject to researcher bias.
Scorecard/ Checklists	<p>To measure the progress of a partner.</p>	<ul style="list-style-type: none"> Effective at tracking progress of a partner over time. Can allow for systemized comparison of multiple partners. Provides a degree of objectivity by reducing bias in data collection and analysis. Can facilitate the quantification of qualitative information. 	<ul style="list-style-type: none"> Depends on measures set out in scorecard (not able to capture changes outside of scorecard list).

Data Collection Alternatives for Common Democracy and Governance Activities

Whenever you do a program activity, an M&E component should generally be included. To help you include evaluative components into your program activity, this table lists different data collection possibilities for common activities. This list will also help you fill out the indicator matrix by providing lists of possible M&E data collection method and sources.

Major Categories of Processes (i.e. Activities) and Options for Measuring Associated Results				
Program Activity/ Process	M&E Activity <i>Types of M&E activities that may be useful to accompany the program activity</i>	M&E Tool <i>Tools you will need to undertake this M&E activity</i>	Result Level <i>This will depend on the program and tool used</i>	Use / Purpose of This Activity <i>Why you would undertake this M&E activity</i>
1) Trainings, Workshops & Presentations (formal learning-related activities)				
	<ul style="list-style-type: none"> Pre-/post-test of knowledge 	Pre- /post-test	Output	To assess actual knowledge uptake and perceived knowledge uptake
	<ul style="list-style-type: none"> Track participation 	Participant lists	Output	To track participation
	<ul style="list-style-type: none"> Post-test only of knowledge 	Post-test	Output	To assess perceived knowledge uptake
	<ul style="list-style-type: none"> Post-event questionnaire 	Questionnaire	Output	To obtain feedback on training and other follow up information
	<ul style="list-style-type: none"> Training observation 	Observation checklist	Output	To assess participation, observed reactions, group dynamics, etc. during the training program. Particularly useful for events that bring together different groups (example: women and men for a gender training) or potentially acrimonious groups (example: civil society and parliamentarians)
	<ul style="list-style-type: none"> Group discussion at end of session 	Group discussion facilitation guide	Output	To gather immediate feedback and perceptions of the training/workshop/presentation; to gauge how much participants think they learned
	<ul style="list-style-type: none"> Photographs, videos and other visual evidence 	Camera, video, audio equipment	Input	To provide evidence of training/workshop
	<ul style="list-style-type: none"> Collecting training products 	None	Input	To provide evidence of training/workshop outputs
	<ul style="list-style-type: none"> Document/content review 	Document review checklist	Output/Outcome	To assess the quality of what is produced by participants during a training/workshop (such as the quality of a draft mission statement), or to assess quality of what participants produce after leaving the training/workshop

Program Activity/ Process	M&E Activity <i>Types of M&E activities that may be useful to accompany the program activity</i>	M&E Tool <i>Tools you will need to undertake this M&E activity</i>	Result Level <i>This will depend on the program and tool used</i>	Use / Purpose of This Activity <i>Why you would undertake this M&E activity</i>
	<ul style="list-style-type: none"> Pre-/ post- (or post only) focus group discussion 	Focus group discussion guide	Output/Outcome	To assess perceived changes in knowledge, to obtain feedback on training, to brainstorm next action items
	<ul style="list-style-type: none"> In-depth interview immediately after the training: phone or in-person 	Interview protocol	Output	To assess whether participants changed perceptions based on the training, built relationships, learned new information or plan to implement any part of the training
	<ul style="list-style-type: none"> In-depth interview several weeks after the training: phone or in-person 	Interview protocol	Outcome	To assess whether participants changed behavior and/or used information from the training (example: follow on activities)
	<ul style="list-style-type: none"> Survey immediately after the training: via email, online or SMS 	Survey or questions	Output	To assess whether participants changed knowledge or plan to act using information from the training
	<ul style="list-style-type: none"> Survey several weeks after the training: via email, online or SMS 	Survey or questions	Outcome	To assess whether participants changed behavior and/or used information from the training (example: follow on activities)
	<ul style="list-style-type: none"> SMS-blast questionnaire 	Short list of questions	Output/ Outcome	To gather immediate feedback on any topic; can be an outcome if it asks about changes in behavior

Data Collection Tools

There are any number of methods and tools to collect useful data. The following is a brief introduction to the more common data collection tools.

Public Opinion Research

Polls can also be a tool for M&E. They are an excellent way to gather data to measure progress in the implementation of your activities and any change that has occurred. They are often a missed opportunity when it comes to M&E! Remember, however, that polls are often expensive and require technical expertise. Here are some examples of how polls can be useful for M&E:

- ***Using Polls to Measure Progress in the Implementation of Activities***

Imagine that your program is helping a CSO watchdog by paying for advertising space in major newspapers to include information on parliamentary affairs. Your M&E plan could simply count the estimated readership of each paper to assess the reach of the information on parliamentary affairs. This may depend on estimates provided by the newspapers themselves, which may or may not be accurate.

If your program is already conducting a poll, it could be used to get a more accurate picture of how many people actually received the newspaper, and *how many actually read those specific program-sponsored advertisements*. This is a better measure than readership estimates.

- ***Using Polls to Measure Change***

Imagine that your program is going to train and mentor government officials on constituent outreach – specifically, conducting regular public forums in each city. Let's say you have two polls planned for your program for message development training. Your M&E plan could simply count the number of public forums that government officials hold, or the number of attendees. While useful data, it does not get to the real purpose of the public forums: encouraging public participation in the political process and strengthening confidence in government and the democratic process.

If your program is going to train a significant number of government officials and is expected to have an extended reach, you could insert extra questions into the poll to measure, for example, the public's perception of how engaged its officials are in each province; whether or not the public received information about outreach events; whether or not they acted on that information; and the level of public participation in town hall meetings.

When conducting a poll at the beginning and the end of the program, the poll can show evidence of change. For example, the poll can show change in the public's opinion of its government officials from the beginning of the program to the end. This would give you an indication of change.

Self-Administered Surveys (such as Training Questionnaires)

Self-administered surveys – such as post-training questionnaires – are often used to measure how well the trainer did, how well the training was organized, etc. Such questionnaires, properly crafted, can do much more. For example:

- ***Pre-/post-training or event questionnaires can provide a more rigorous assessment of change.***
It's common to administer post-training questionnaires to gather feedback on the event. But a pre-training questionnaire can serve as a baseline, and a post-training questionnaire can serve as an endline. Perceptions, awareness, knowledge and expectations can be compared between baseline and endline. While you can measure some of this change through post-training questionnaires only (no pre-test), it is not as rigorous.
- ***Self-administered surveys in the form of post-training questionnaires can be used to gather data about program participants and provide important program data points.***
Perhaps a program wishes to know whether participants learned new information by attending a party witness training, to find out whether party members were getting information on election regulations from other sources (and if so, from where), and to gather sufficient information to further track participants. The program team therefore distributes post-training questionnaires to all 9,000 participants of its referendum program. In another example, a program designs a basic questionnaire template to be used in all trainings and events to gather sufficient information to track network members and gather data necessary for M&E purposes. The questionnaire template includes a space for questions from specific trainings, with sample questions. Data from the questionnaire will be included in the network's global member database.
- ***Guided surveys can be used to guide program staff in conducting proper follow-up with program participants.***
A program may want to find out whether election training program participants applied the skills learned in the training, and whether these skills contributed to their success on Election Day. Program staff design a questionnaire to measure these aspects. Then, a few months later, they pick a random sampling of training program participants and call each participant to administer a survey. The questionnaire response guides their telephone questions, allowing for objective and comprehensive data for more rigorous analysis.

Questionnaires can be filled out by participants themselves, or if participants have limited reading or writing abilities, with the help of staff. Below are some different permutations of self-administered questionnaires, and some tips for which you would use in which situation:

	Why You Would Use It	Advantages	Disadvantages
Pre-Test* Only	To gather demographic information and to inform event content	<ul style="list-style-type: none"> • Participants are more likely to fill the test out comprehensively at the beginning of an event. • Can prepare participants for event topics and thereby contribute to learning. 	<ul style="list-style-type: none"> • Does not provide any assessment of event results.
Post-Test* Only	To gain a snapshot self-assessment by an individual after an event	<ul style="list-style-type: none"> • Does not need a baseline. • Only requires a test at the end of the event. • Possible to measure perceived changes. 	<ul style="list-style-type: none"> • Less rigorous measurement of change. • Not possible to measure actual changes, only perceived.
Pre-/Post-Test*	To compare a baseline to an endline to gain a real sense of change before and after an event	<ul style="list-style-type: none"> • Can compare change more rigorously. • The pre-test may prepare participants for the event topics and thereby contribute to learning. 	<ul style="list-style-type: none"> • Takes more time to administer. • Challenging to maintain confidentiality while comparing the pre and post-test from the same individual.

* Test can be synonymous with self-administered questionnaire.

Here are some tips on how to develop the best possible questionnaire⁹:

- **Clearly define the objective of the questionnaire and develop a list of things you would like to know.**
This will make the questionnaire flow logically and help you gain appropriate data and insightful information. If you are having trouble writing questions, it's probably because your objective is not clear.
- **Keep the questionnaire as short as possible.**
Define precisely what information you need, and write as few questions as possible to obtain that information. Ideally, respondents shouldn't need more than 15 minutes to complete the entire questionnaire.
- **Only ask one question at a time.**
For example, write: "to what extent is the material clear?" rather than: "to what extent is the material clear *and helpful*." Material can be clear but not helpful, and vice versa. The multiple variables will confuse the respondent or they may only respond to one part of the question.

⁹ For further information, please refer to: Fowler, Floyd J. *Improving Survey Questions*. Thousand Oaks, CA: Sage Publications, 1995.

- ***Avoid open-ended questions, wherever possible.***
People won't answer in depth. There will probably be a lot of bias (i.e., less literate participants will not contribute). Also, with larger samples, responses to open-ended questions take much more time to analyze. Rather than open-ended questions, categorize all possible responses to multiple-choice, and include a write-in spot for obscure categories relevant to only a few people ["other (please specify)"].
- ***Allow for write-in answers for any numeric responses.***
For questions that ask for a numeric answer, allow these to be write-in responses (open-ended). Why? Because doing so prevents you from accidentally fixing the answer. If your question asks how many hours per day does the participant watch television, and the range is zero, one, three or five+ hours per day, the respondent may be too embarrassed to pick the biggest option even though (s)he may watch seven hours per day.¹⁰ Also, you may pick the wrong range, which will distort the responses; this is often true for age. If you have a youth training, you may not pick a range that accommodates older participants, when in fact many are not "youth" but are forced to select an age range that is considered "youth."
- ***Be as clear as possible.***
Rather than say "(1) very often (2) often (3) sometimes (4) rarely (5) never," write "(1) every day or more (2) 2-6 times per week (3) about once a week (3) about once a month (5) never." People can have different interpretations of "very often."
- ***Make sure your answer options make sense for the question.***
When asking respondents to rate their feelings, knowledge, etc. against a scale, be sure to choose a rating system that makes the most sense according to the questions you are asking. For example, it is not appropriate to ask respondents to rate their feelings on a scale of 1 to 5 when they questions being asked elicit a binary (yes/no) response.
- ***Limit rating scale points to four or five options.***
A rating scale (never, rarely, occasionally, fairly often, etc.) is often confusing the more points you have.
- ***When necessary, provide "Do not know" and "Choose not to answer" choices among possible responses.***
This prevents the respondent from answering falsely just because the real answer is not available, or because of confusion or uncertainty, reading difficulty, reluctance to answer or question inapplicability. However, make sure to not include this option for questions that do not need it; sometimes respondents choose it simply to finish the survey more quickly, which means your results may be skewed!

¹⁰ Example drawn from: Doucette, Anne. "Applied Measurement for Evaluation." Lecture at The Evaluator's Institute, George Washington University. Washington, D.C.: July 2012.

- ***Be cognizant of why you are offering a neutral response option (for example: “stays the same”).***
 Sometimes you cannot be sure the participant really has a neutral opinion. If that is the case, the data will not be accurate. However, questions that measure change usually require a neutral response option.
- ***Clarify whether or not you required extreme stances in response choices.***
 Sometimes you want answers at the extreme end of the spectrum (“I strongly agree”). At other times, you want to ensure that people with some uncertainty can still make a choice (“Somewhat agree”). Decide at the outset what answer choice is appropriate for your survey.
- ***Watch out for ingrained assumptions.***
 For example, watch out for conventional interpretations of words (i.e., “emotional” is often assumed to be a negative response when it comes to the workplace, where it could be a positive response when it comes to family).
- ***Write questions so that all possible answers seem equally acceptable.***
 This will prevent the question from seeming to lead the respondent to an answer.
- ***If you need respondents to rank a list of items, try to keep the list as short as possible, and ask for the single most important item and the second most important item.***
 People cannot be expected to rank more than six things, and it is an easy question on which to make mistakes. Consider whether or not you really need to know more than the top three items. Data beyond that is rarely useful in analysis.
- ***If you need to include sensitive questions, put them at the end of the questionnaire.***
 First, try to rewrite it to be less sensitive! Also, those who are upset by the question may not complete the remainder of the questionnaire if it is at the top.
- ***Decide whether anonymity is important or not.***
 If you also use the questionnaire to track a lot of participants for future follow-up, then the questionnaire shouldn't be anonymous. If you only really need contact information from a few participants, consider getting that information separately, especially if the questionnaire has sensitive questions. Or, institute a code number for each questionnaire and have the code linked to a participant list.
- ***Be consistent in what demographic information you ask for, particularly in a training program.***
 Establish what information you really need for analytical purposes and for following-up in the future. Some of the data may not seem applicable now, but aggregated over time can provide interesting information, especially at the end of a grant.
- ***Pilot test the questionnaire.***
 This will show any question or formatting issues that may prevent you from getting the right response, as well as any translation issues. It will also weed out unclear questions.

- ***Take the time to make the questionnaire attractive, clearly laid out and clearly printed.***
Make sure that questions do not spill on to another page; start a new page for the question if necessary.
- ***Prepare the method by which you will analyze the data at the same time as you draft the questionnaire.***
This will help you format the questionnaire in a way that is most conducive to inputting the data. This is essential to using the data later. If you have gathered 500 questionnaires and inputting the data is arduous because of the way the questions are designed, you are unlikely to complete the project or you will delay the data management, which in turn delays the data analysis.

Focus Group Discussions¹¹

A focus group discussion (FGD) is a structured discussion among a group of people about a specific subject. You can use FGDs to help design polls, analyze polling results or to test messages, speeches and campaign materials to see if they resonate with voters. FGDs can also be used to gather data on your programs from an M&E standpoint, especially when it is important that you capture the interaction between individuals, or if you are looking for new responses. FGDs are different from group interviews. A group interview is a way of getting a lot of responses from different people in a short amount of time. In a focus group, the *interaction* between participants is important.

Step 1: Select your topic and design your questions.

The discussion can last anywhere from twenty minutes to two hours, depending on the scope of what you would like to find out and how long you think you can maintain the attention of your participants. A one-and-a-half-hour session would generally need five to six questions.

Step 2: Select your group of individuals.

FGDs are most effective when there are between six and ten participants. Participants should be carefully selected depending on the kind of information you need. There is no strict rule, but consider: gender, level of education, age, ethnic/religious/tribal affiliation, socio-economic status, membership in political party or civil society organization, and experience in relation to the topic.

Step 3: Select your moderator.

The success of the information garnered from a FGD is highly dependent on the quality of the moderator. The moderator must have strong communication skills, be adaptable, be able to listen and control nonverbal reactions, be friendly and not be intimidating.

Step 4: Set up the space and decide how to record the information.

Participants should sit at a round table or in a circle so that all members can see each other.

Step 5: Decide whether the discussion will have an audio recording or will be videotaped, and ensure focus group participants are comfortable with the recording method.

Practice with the recording devices and check them periodically throughout the session to ensure they are working. Determine in advance who is responsible for transcribing the recording.

¹¹ For additional reading, refer to: Stewart, David, Prem Shamdasani, Dennis Rook. *Focus Groups: Theory and Practice*. Cambridge, MA: Harvard University Press, 2006.

Step 6: Guide the discussion.

- Start with an introduction to the topic to establish contact, explain the need for the focus group, and walk through the agenda.
- Explain the ground rules for the discussion, and explain how the session will be recorded.
- Start with the first question. Make it simple. Allow participants to think about their answers before facilitating discussion. Where necessary, summarize the result of the discussion out loud to participants for the benefit of the person taking notes.
- Move from simple to complex questions, or from least sensitive to most sensitive questions.
- Work to ensure even participation. Call on those who are more reticent. If necessary, allow each person one minute to answer the question. Discourage interruption.
- Close the session, summarizing key themes and inviting participants to give any final comments.

Step 7: After the discussion, write down immediate reactions and observations.

If hand notes are taken, read through them to clarify anything that does not make sense before you forget what was said.

Step 8: Analyze results.

Plan to transcribe, and where necessary, translate the discussion. Go through the discussion carefully and draw out general attitudes, specific opinions, and other inferences.

Step 9: Prepare the presentation of results.

Make sure the FGD results are not wasted! Prioritize the completing of a report draft, presentation, or method by which you will convey the results of the FGDs.

In-Depth Interviews¹²

Interviews can be an important means of gaining a deep understanding of changes in perceptions, attitudes, and to gather facts and anecdotes. Here are some examples of when interviews might be a useful data gathering method:

- You have conducted extensive trainings with elected officials on the importance of responsiveness to constituents, yet the trained officials have not improved in this area. You could conduct interviews with these officials to understand why.
- You have worked with CSOs on various capacity building initiatives. While undertaking this work, the political environment for civic activism worsens. You could interview CSO members to understand how the political environment is impacting their progress, which can then help you re-tool your program accordingly.

An interview is more than just a conversation with a program participant or beneficiary. Interviewing requires skill and proper preparation. The right questions must be asked in the right way to capture high

¹² All citations in this section refer to the following: Patton, Michael Q. *Qualitative Research & Evaluation Methods*. 3rd Edition. Thousand Oaks, CA: Sage Publications, 2002.

IRI staff have benefited from Patton's book and lectures on this subject, and encourage those seeking more information on interview techniques to refer to this book, specifically Chapter 7, "Qualitative Interviewing."

quality data. The quality of information obtained during an interview is largely dependent on the interviewer.

There are different types of interviews, but consider following a semi-structured interview approach. In this approach, questions are formulated in advance. This approach has several benefits, including: 1) it helps to make sure that the interview focuses on the important topics; 2) it serves as a guide for a less experienced interviewer; 3) it can be provided in advance to an interpreter or facilitator for translation and piloting; and 4) the interviewer is also free to go off script and further explore topics that come up during the interview.

Here are some things to keep in mind when conducting interviews for M&E purposes:

1. *The interviewer is responsible for ensuring the interview goes well.*

- *Be prepared and engaged:* Be aware of your demeanor, body language and energy level throughout the interview. Make sure that the interviewee feels that you value his/her opinions and are truly interested in what s/he has to say. Try not to schedule too many interviews in a day so that you can give each interviewee your full attention.
- *Be neutral:* In order not to influence the responses of the interviewee, the interviewer must establish and maintain neutrality throughout the interview. Explain to the interviewee at the beginning that they should feel free to answer (or not answer) questions as they see fit. Then, be mindful of any conscious or unconscious reactions you have to what they say – remember that the tone of your voice, your facial expressions and your body language can convey subtle messages and potentially introduce bias into the interview!
- *Guide the interview:* It is your responsibility to make sure that the interview stays on track and on time. Veering off-topic can sometimes yield unexpected, yet important, information, but remember that you probably have limited time to get the information you need. Therefore, try to keep the interview on track and on schedule. Use cues and transitions to help the interviewee know where you are in the interview, particularly if you are covering a lot of topics. Finally, feel free to gently course-correct if the interview is going off track. Evaluation expert Michael Quinn Patton suggests using either verbal or non-verbal cues: stop nodding, stop taking notes, and sit back (all cues that the interviewee should stop talking) or simply interrupt the interviewee. As Patton points out, while this might seem impolite, it is disrespectful to the interviewee if the interviewer is not making good use of the time being spent.

2. *How and when you ask questions is important.*

- *Start off easy:* It is better to start with a question that is easy and comfortable for the interviewee to answer, but still relevant to the research topic(s). A good example is asking an interviewee to describe his/her experience with a program or personal background in democracy/advocacy work. This helps the interviewee understand the scope and tone of the interview and prepare to answer subsequent questions. Patton suggests avoiding potentially controversial questions, which can put the interviewee on the defensive from the outset, as well as demographic questions, which may condition the interviewee to respond with brief, purely factual answers.
- *Ask one question at a time, and ask it clearly and directly:* While interview questions can (and should) be asked in a relaxed, conversational manner, don't try to fill space by asking

several questions at once or offering potential answers. Formulate your questions in a way that encourages open-ended responses, but make sure it's clear what is being asked. Rather than "Would you say the training was good, like maybe on scale of one to five, how helpful was it...would you come to another one?" try "What were your impressions of the training?" If the interviewee appears confused, feel free to offer clarification – but only after allowing him/her time to process the question and think about his/her response.

- *Try to not ask "why" questions:* Patton recommends against asking "why" questions, because they imply that the interviewee should understand the reasons behind the occurrence of an event, and that the reasons are even knowable. They can also imply that a person's response has been inappropriate. So instead of asking, "Why did you undertake this program activity?" you might ask, "What is it about this activity that you thought suited your organization's needs?"
- *Tread carefully when asking sensitive questions:* Sometimes you will need to ask questions, such as those about the political context or their political activities, that may be sensitive or uncomfortable for the interviewee. As mentioned earlier, don't start off with these questions! Instead, ask them later on, once the interviewee has had a chance to feel at ease with the process overall. Patton's "presupposition" format can be a useful way to introduce these questions. "Presupposition" questions assume that a potentially controversial or difficult answer is quite normal and acceptable. So instead of asking "Has this program had any challenges?" you assume the answer is yes, and ask instead, "What challenges has this program had?" Note that the interviewee can still say "none" with this question – so the question is not biased. Finally, when asking sensitive questions, be cognizant of the interviewee's emotional state. If appropriate, remind them that their responses are voluntary but will be kept confidential should they choose to respond.
- *Follow up and ask for clarification as needed:* With a semi-structured interview, it is perfectly alright to ask additional questions that are not on the protocol for the purposes of following up or clarifying responses. Follow-up questions should be conversational and casual, allowing the interview to continue to flow freely. It may be advisable to prepare a few sample follow-up questions in advance, particularly for less experienced interviewers.
- *The closing question is important:* At the end of the interview, be sure to allow the interviewee a chance to offer any thoughts that weren't addressed in the interview. Even the most carefully constructed interview protocol can't anticipate everything that might be important to an interviewee! Patton frames this as allowing the interviewee to have the "final say": ask, for example, if there is anything that the interviewee would like to add, or if there is anything that should have been asked, but wasn't. Be sure to ask the question while the interviewee can still offer a complete and thoughtful response – don't wait until after you've packed up your notebook and walked him/her to the door.

3. *Keep a careful record of the interview.*

- If at all possible, record the interview. This allows you to focus more completely on the interviewee and not be too concerned with taking extensive notes (which can actually distract you from what the interviewee is saying). Recordings also are more accurate and unbiased, since they are a verbatim record of what was said, and are thus an important way of ensuring transparency and objectivity in your data collection and analysis efforts. In group interviews, you may even consider video recording so that you can keep better track of who is saying what. Of course, be sure to obtain participant consent before any recording begins.

- Even with a recording, you should still take notes to remind you of important or interesting concepts, either for follow up or later analysis. As Patton observes, taking notes also sends a message to the interviewee that what they are saying is important. Taking notes can also help you mentally “place” the interview and specific comments that were made to help you remember details when you come back to the notes/transcripts. For example, make notes of the location, the interviewee’s demeanor, anything unusual or interesting that happened, etc. When the interview is over, briefly return to your notes and add any initial impressions and ideas that you think are worth further exploration.

Scorecards and Checklists

Scorecards and checklists can be used in a variety of ways to inform program implementation and measure program results. Properly used, they can allow for a systemized comparison of multiple items or one item over time. They can also provide a degree of objectivity by reducing bias in data collection, which benefits program choices and/or analysis of people, organizations, events and situations.

Here are some ideas for how to use scorecards and checklists:

- You are working to build the organizational capacity of a political party. You could conduct a baseline capacity assessment at the program outset using a scorecard to form a baseline, and another assessment at the close of the program using the same scorecard to form the endline.
- You would like to work with three CSOs, but you have a large number from which to choose. You want to have an objective decision-making process. You could use a checklist or scorecard to outline your priorities in your choice of CSO. Then fill one out per CSO, and you will have a basis for a decision.
- You will be tracking the quality of town hall meetings planned and implemented by a partner CSO, who you have trained to conduct town halls. You would like to observe the town halls over the course of the year, and track the CSO’s change in demonstrated capacity to undertake town halls. You could do this through an observation checklist or scorecard.

Scorecards and checklists are similar but have clear distinctions. With a scorecard, values are weighted. So, something that is not as important as something else is weighted less. Items in a scorecard can be valued on a scale, such as zero to five, or they can be binary with a value of zero or one. With a scorecard, a final score is produced, taking weights into account. With a scorecard you can say: “Organization X improved from 50 percent to 65 percent.”

A checklist may also provide a numeric value, but it is generally not weighted and there is no definitive score. Checklist items are binary: you achieve the item or you do not, so for each item you receive either a zero or a one in value. With a checklist you can only say: “Organization X improved on or achieved 13 of the 20 checklist items.” Note that in a checklist you can’t give a percentage score, because items are not weighted, and there is no score. You can, however, say what percentage of checklist items was achieved: “Organization X demonstrates competence in 60 percent of checklist items.”

Scorecards and checklists can be broken down into two types: goals-free and goals-based¹³, as follows:

- ***Goals-free scorecards and checklists (universal criteria):***
A goals-free scorecard or checklist uses universal criteria. For example, a goals-free scorecard for organizational capacity will consider what the universal hallmarks of strong organizational capacity are, regardless of where different organizations being assessed are in their development. So, for example, with a goals-free scorecard, you could compare a brand new organization with an organization that is 20 years old, and their scores would reflect their different levels of maturity. Unsurprisingly, it is challenging to develop this scorecard, because you need a huge range of criteria. You would use a goals-free scorecard or checklist if you needed to compare the absolute states of multiple organizations. You could also use it for a single organization over time or as a spot-check assessment, but given the amount of work it would take to develop a universal scorecard, it is not recommended; rather, you are better off using a goals-based scorecard or checklist.
- ***Goals-based scorecards and checklists (individual criteria):***
A goals-based scorecard or checklist uses individual goals to determine scorecard/checklist items and criteria. An individual organization can be measured over time using a goals-based approach; the baseline would be zero (no goals achieved), and the endline would look at how many goals were achieved. In this way, it does not matter where the organization is in its development, because your baseline is where it is now, and the criteria is where you want it to be at the end of the program (the goal). If measuring multiple organizations, note that you can only compare the rate of their individual progress; you can't assess the absolute level of development of each organization against each other. So, you could say that some organizations are progressing faster in their goal accomplishment than others, but you could not say that one organization is absolutely more advanced than another.

¹³ The terms “goals-based” and “goals-free” were first defined by Michael Scriven with regard to the use of criteria to define evaluation approaches; see: Scriven, Michael. *Evaluation Thesaurus, 4th Edition*. Newbury Park, CA: Sage Publications, 1991.

Here is a comparison breakdown of the different types of checklists and scorecards, and in which situation you would mostly likely use which type:

Scorecard/ Checklist Objective:	Goal-Free (Universal Criteria)		Goal-Based (Individual Criteria)		Recommendation
	Scorecard	Checklist	Scorecard	Checklist	
Snapshot / One-Time Assessment					
Single organization; snapshot assessment	✓	✓	✓	✓	It is recommended to use a goal-based scorecard or checklist for a single organization, snapshot assessment. A universal criteria checklist or scorecard is possible but not recommended unless you need to compare the single organization against an existing measurement.
Multiple organizations; comparing absolute levels of each as compared to the other	✓	✓	-	-	It is recommended to use a checklist rather than a scorecard, as a snapshot comparison need not provide a score, but could compare the number of checklist items addressed. A scorecard is more difficult to construct.
Change Over Time					
Single organization; comparing baseline to endline (or intermittently)	-	-	✓	✓	It is recommended to use a goal-based scorecard or checklist for a single organization, assessment over time. A universal criteria checklist or scorecard is possible but not recommended unless you need to compare the single organization against an existing measurement, because the criteria may not be specific enough to the organization.
Multiple organizations; comparing individual changes or progress, not absolute levels	-	-	-	✓	If you only need to compare the rate of change/progress, it is recommended to use a checklist with individual goals. A scorecard is possible, as is a universal criteria scorecard or checklist, but probably not necessary.
Multiple organizations; comparing changes in absolute levels of each as compared to the other	✓	-	-	-	It is recommended to use a universal criteria scorecard, as the weighting system and the score will provide a more accurate measurement of change across multiple organizations.
Note that a scorecard or checklist is only as accurate and useful as the effort put into establishing appropriate criteria. The most difficult to construct is a goals-free (universal criteria) scorecard, followed by a goals-free checklist. A goals-based (individual criteria) checklist is the easiest to construct, with the goals-based scorecard slightly harder.					

Tips for Designing the Scorecard/Checklist:

- *Clearly define your scorecard/checklist objectives.*
Is your objective to measure progress at the beginning and end of the program? To compare across units? Clarifying your objective will help you keep the scorecard/checklist focused. Note that in the table above five general categories of objectives have been identified.

- *Take care to differentiate between identifying potential for improvement and evaluating performance.*
Potential for improvement and performance are very different things. An organization may be poorly structured and managed, but produce great results, and vice versa. Thus, a program that works to build CSO organizational capacity, but that writes a scorecard that will measure CSO performance, will not get the data it needs to inform program implementation or measure program results regarding organizational capacity.
- *Whenever feasible, involve your stakeholders in the design.*
Sometimes involving the people/organizations you are using your scorecard/checklist to measure actually helps in the design and data collection itself. If you are writing a scorecard to measure the performance of a government ministry, you may not include a key component of the ministry's work or organization that would have been flagged had the design involved ministry staff.
- *Make the scorecard/checklist variables as simple as possible.*
If using the list requires a judgment be made by the individual completing it, having simple variables will reduce bias. Also, yes/no questions are preferable to questions that ask you to rate on a scale. Why? Because it's easy to make a mistake or introduce bias with a scale, unless each scale parameter is spelled out comprehensively.
- *Decide whether you are measuring over time, or a snapshot in time.*
If you are measuring an organization over time, you need a much deeper set of unambiguous criteria, to allow for a wide range of change, as well as the potential that the person(s) conducting the endline assessment will not be the same person who constructed the scorecard/checklist or conducted the baseline.

Tips for Fielding the Scorecard/Checklist:

- *Train the individuals responsible for fielding the scorecard/checklist.*
This will ensure consistency in data collection.
- *Whenever feasible, involve several people in the scoring decisions.*
This will prevent haphazard guessing or quick judgments. For example, in the illustrative political party scorecard, there is a question regarding the existence of local party offices. One program had difficulty measuring this variable across different parties: how do you compare an empty room with a party sign out front to a fully staffed and functioning party office? The final, tailored scorecard/checklist either should have broken this item down to several other variables (is the party office staffed regularly, etc.), or defined very clear parameters used by several people together to define the rating for that variable for all parties subject to the scorecard/checklist.

Observation and Visual Evidence

In many instances, your observations can be legitimate and important sources of data for your programs. For example, let's say you have trained government officials on the importance of holding town hall meetings, and how to do so. Your indicator could simply measure how many meetings have taken place, or how many follow-up activities were initiated by the officials as a result of the meeting.

However, these indicators would not answer the following questions: Did officials provide important information to the public? Was there active participation by constituents? Did citizens leave appearing satisfied with the meeting?

For some programs these are important indications of program progress, which are not captured in the first indicators mentioned above. Observation, correctly performed and documented, could provide data to answer these sufficiently.

Observation lets you:

1. Gain direct information.
2. Understand ongoing behavior, processes or events.
3. Examine physical evidence, products or outputs.
4. Supplement data drawn from other sources that may not be enough.

Here are other examples where observation may be a useful data source for your programs:

- Assessing whether candidates used techniques emphasized in trainings in their campaigns (speeches, posters, radio addresses, television advertisements, etc.).
- Monitoring trainings by CSO trainers who participated in train-the-trainer programs.

Tips for making the observation as objective as possible:

- If you will be observing and/or comparing multiple incidences (for example, the campaigns of several different parties), create a list of items you will be looking for to help track data and observations. See “scorecards and checklists” for more ideas.
- Be disciplined in documenting your observations. Patton suggests writing them down during the observation or immediately after.¹⁴
- Triangulate your observations as much as possible. Triangulation involves using multiple sources or kinds of data. In the case of the campaign example, having several different people observe and compare their notes, or observing several different events, is preferable for data legitimacy.
- When reporting results, take care to consider, and if necessary report on, the strengths and limitations of the observers.

The data from your observation can be either quantitative or qualitative. If you observe multiple events and use checklists to manage the data, then your results could report quantitative information, such as: “Four of the five trainings observed had very active audience participation. Active audience participation was defined as...” If you are reporting on singular events, your information may be recorded in a descriptive narrative, which is qualitative data.

Content Analysis (Document Review)

Documents and other items, like video, can be good sources of data. A proper content review will help you find trends or patterns in the data from which to draw conclusions or ascertain program results. Here are some examples where a content review may be a useful data source for your programs:

¹⁴ Patton, Michael Q. *Qualitative Research & Evaluation Methods*, 3rd Edition. Thousand Oaks, CA: Sage Publications, 2002.

- Parties were trained on how to draft election manifestos. Your organization would like an objective analysis of whether or not your advice on content appeared in the manifestos themselves. A document review would allow you to analyze the manifestos against pre-determined criteria.
- Your organization trained journalists to be more balanced in their reporting. A content review of media reports could guide you in a review of media reports to assess whether the reports improve in quality.

Step 1: Decide what you want to analyze, why you want to analyze it, and an appropriate sample size. There is no fixed rule to the sample size, but be prepared to explain your reasoning behind your selection. If you are looking at party manifestos, you will probably need to see the manifestos from all parties with which you work. If you are looking at congressional research reports, you probably do not need all reports, but a smaller sampling and a mix of authors and subjects. Consider a random sampling, which would eliminate some bias.

Step 2: Collect the documents.

If you cannot find enough of the right documents, the analysis may need to be abandoned. Each document must be checked to make sure it is appropriate (e.g., not missing sections; on the right topic).

Step 3: Decide what variables (criteria) you will use to analyze the documents.

For example, if you trained parties on how to write proper manifestos, ask yourself: what makes a good or bad manifesto? What should I be looking for in the document to ascertain this? These will constitute your variables. Limit the number of variables as much as possible, and organize these variables into a numbered checklist or scorecard.

Step 4: Scan the documents and code for each variable in the checklist.

Coding is discussed in detail in a subsequent section (*Qualitative Data Analysis*).

Step 5: Organize document sections by variable, and look for trends and patterns within each variable.

Step 6: Analyze the data and draw out conclusions.

Be careful to track how you came to these conclusions. Keep the list of variables and the coded documents on file in case you need further analysis or to prove certain conclusions.

Data Collection Approaches for Special Circumstances

There are some situations when traditional approaches to data collection may not be appropriate. Here are some approaches that can be taken instead:

	What is it?	In what situation would you want to use it?	What does it involve?
Most Significant Change (MSC)¹⁵	MSC is an approach that focuses on the collection of significant change stories to inform program results and activities, as opposed to vast sums of data. It is often associated with what is called an “indicators free” approach to M&E. For your purposes, it can be used in conjunction with your M&E plan to provide deeper evidence of program outcomes. It is more participatory, in that it empowers participants to select and tell their stories. These stories are then vetted by a panel using pre-determined criteria, and the final stories are used as part of program M&E. The vetting process is intended to make the stories more rigorous than anecdotes.	If program activities are more open-ended in nature; if the program is operating in a very complex environment in which outcomes are not clear; or in contrast, if the program is focused on customized technical assistance such as long-term mentoring.	Set up a schedule of regular change story collection; identify priority topics; identify key informants/participants; collect stories through interviews, written communication, telephone, etc.; write up the story in one to two pages; panel selects the top stories (those demonstrating the most change according to the pre-determined criteria) to inform program M&E.
Rapid Appraisal Method (RAM)¹⁶	RAM is a way of packaging different evaluation methods to quickly collect data.	If the program has limited time on the ground to collect data. This is NOT a substitute for a proper M&E plan. This is for programs that do not have field representation, that only have limited opportunity to collect data, or, if a program needs to make a quick decision and does not have time for a lengthy assessment.	Any data collection method can be refined for RAM. For example: <ul style="list-style-type: none"> • interviews conducted with critical stakeholders (only) known for possessing important information • community discussions (such as through public forums) • mini-surveys that are not representative

¹⁵ Davies, Rick, and Jess Dart. *The Most Significant Change (MSC) Technique: A Guide to Its Use*. Cambridge, UK: Davies and Dart, 2005.

¹⁶ Vondal, P. *Performance Monitoring and Evaluation Tips: Using Rapid Appraisal Methods*. Washington, D.C.: USAID Center for Development Information and Development Evaluation, 2010.

In addition, some programs operate outside of the country in which the program is taking place, whether from headquarters or other field offices. This can be a challenge, because you depend on other people collecting data for you. Here are some common challenges and a list of ideas to mitigate them:

Challenge	Tips
There is limited manpower to implement any activities, including M&E activities; you are juggling activities and M&E activities remotely so it is difficult to complete everything.	<ul style="list-style-type: none"> ● Include M&E activities in contracts. ● Include M&E activities in the workplan as program activities so that they are integrated. ● Reduce and refine the quantity and type of data to collect, focusing on key components; communicate these to your donor.
Program staff/facilitators/implementers do not feel it is part of their job description.	<ul style="list-style-type: none"> ● Include M&E activities in the proposal in the activity section itself as program activities. ● Include M&E activities in the workplan as program activities. ● Headquarters-based staff can train implementing partners and facilitators on their roles/responsibilities in M&E. ● Include M&E activities in contracts.
You lose sight of activities that took place six months ago that must have results now.	<ul style="list-style-type: none"> ● In the workplan, set aside two weeks every few months dedicated to follow-up on past activities. ● After a major activity is completed, set yourself a reminder for three to six months down the road to follow-up with the participants of that activity.
Contractors are not as concerned about results and M&E as the program is.	<ul style="list-style-type: none"> ● Include M&E activities as separate, paid deliverables in the contract. ● Provide contractors with activity report templates as well as follow-up data collection templates and instructions. ● Separate the contract into two smaller contracts: the first contract provides the bulk of the budgeted funds to complete the activities and provide immediate data on those activities; the second contract is for six months later for the remaining amount specifically for follow-up data collection on the results of the activities. Or simply extend the life of the contract and make the final deliverable be follow-up M&E information/data.

Program and M&E Activity Workplan

The indicator matrix can be difficult to use on an everyday basis because it is very detailed. It is suggested that you integrate data collection requirements from the indicator matrix into your program activity workplan.

Alternatively, you can develop a **program and M&E activity workplan table**. This simplifies the workplan, which is often in a narrative form, and integrates activities with the agreed-upon M&E activities. The M&E activities can help you collect information for the indicator matrix as well as general program information.

- *Many print off a large, poster-size version of this workplan and post it on a common office wall so that all staff can work off of it and it can be referenced on a regular basis.*

The workplan can be in any format you and your team find most helpful, however, it should include the following components:

Program and M&E Workplan								
Activity/Result	Associated Indicator (if relevant)	Data Collection Activity (incl. method, frequency)	Responsible Person	Timeline				Notes
				Q1	Q2	Q3	Q4	

The benefit of this format is that it provides a basic timeline of major activities and milestones, without all the accompanying narrative.

Increasing Data Collection Rigor

Data is only as good as the quality of the data collection activity, method and tools used. For example, there could be *bias* in the data. Bias occurs when the researcher's interests and/or world view impacts both the data collection and/or process of analysis; bias can also come from the data source or the method used. There are many solutions to reduce bias, including randomization, involving two or more researchers in the evaluation and triangulating data collection using multiple methods. The following are ways to ensure that your collected data is as good as possible.

- *For larger evaluations or data collection efforts, it's important to pilot test tools and have a record explaining any changes. This is particularly important for programs that will undergo a data quality assessment.*

Piloting Tools

Pilot-testing tools is important to ensure that they not only get you the information you need, but are also politically/culturally sensitive, not too long or confusing, etc. It's important to pilot-test tools, particularly if you plan to gather data from a lot of different people.

For example, if you intend to use a single questionnaire for a training program that will reach 1,000 people, it's critical to test the questionnaire well beforehand so you don't waste time and effort processing 1,000 questionnaires that are not helpful. Without pilot-testing, participants often provide the wrong kind

of information, are confused by the format, circle more than one response, have translation problems, and so on.

Conducting Baselines and Endlines

Ideally, the M&E system will include data collection at the beginning and end of a program. This allows you to assess changes over time. For example, how will you know if you improved if you did not know where you started from? It is sometimes possible to collect baseline data retroactively, but not always. For programs that are working on capacity development of a group of people or of an organization, a baseline and endline are important.

Triangulating

When you triangulate, you use different data collection methods (interviews, surveys, etc.), data sources (training recipients, impartial observers) and data collectors (for example, two different field staff members) to corroborate findings. Triangulating will increase the rigor of the findings. For example, if you want to measure the level of expertise after a training, you would collect information from recipients of the training as well as impartial observers, such as journalists or party leaders, and/or mix the methods you use, such as by using both interviews and pre- and post-training tests.

➤ ***Often, the program has a natural comparison group, like when you have to stagger training events because of time or resource limitations. If it is important to your program to track changes in participants compared with non-participants.***

Using Comparison/Control Groups

An advanced way to increase rigor is to compare program participants (**treatment group**) with individuals that did not have access to the program (**comparison/control group**). Comparison group participants are selected to closely resemble the treatment group in many demographic variables (age, socio-economic status, minority status, education, etc.). A control group is randomly selected. Comparison and control groups can provide a counterfactual, which is a glimpse into what the treatment groups might have still looked like if the program had not taken place. Without a comparison/control group, you are less sure that any observed changes are the result of programs.

Using Random Selection

Random selection can be used as part of a formal evaluation, or to help lend rigor to everyday data collection. For example, for everyday data collection, random selection of follow-up interviewees is a good way to reduce selection bias. For formal evaluations using random selection, a group of people is randomly selected to participate in the program, or to serve as a control group. The randomization must take place well in advance of the program activity. Random selection is a technique to ensure that the only difference between control and treatment groups is their participation in the program. Without randomization, differences in the two groups could be attributed to external factors, and it would therefore be difficult to isolate the cause for the change.

Using a Participatory Approach

Often, you depend on data from your participants. This can be difficult to gather, particularly since participants often do not know what you really want to know, and/or are not invested in the process enough to commit to getting accurate or regular data. Engaging participants from the offset can fix many of these problems by building program buy-in and engendering interest among participants themselves about the result of the project.

Data Collection Tips

Here are some tips to increase the efficiency and usefulness of your data collection:

- Think about data collection as you would a regular program activity. Data collection is often the last thing you think about, which makes it an added burden. But remember that data collection is a requirement, just like a training program is a requirement.
- Integrate data collection into your existing system for activity planning rather than creating a new system.
- Assign responsibility for data collection to specific individuals.
- Use staff meetings to prepare for future program activities as well as future data collection efforts. This is especially helpful if your office has a database or M&E officer, so that they know what is coming up and how to help.
- For data collection that is linked to a specific activity, require data collection and analysis at the same time as the general report. Some programs have linked reimbursements to the submission of collected data and its analysis.

Data Analysis

How you analyze data is as important as what data you collect. Through analysis, data becomes information that can be used:



➤ ***Remember that it is how you deal with the finding, rather than the finding itself, that is essential. The finding itself is just information. If you don't even see the right information, then you will miss the bigger picture.***

The process of analysis is finding themes through sometimes disparate data. Through analysis, you transform data into information, and thereby into findings.

Data analysis sounds fancy but is not necessarily difficult. The important part of data analysis is being as objective as possible. Different people can analyze the same data and draw different conclusions if they hold on to their preconceived ideas and assumptions. While it is natural to look for certain things in the data, it is critical to be rigorous and allow yourself to see an opposing or unattractive finding.

The analysis method you select will depend on your data: quantitative, qualitative or mixed methods.

Qualitative Data Analysis

Step 1: Data Reduction

Data reduction is the process of selecting, focusing and simplifying your data to a manageable quantity. It can be done by finding patterns and themes and associating pieces of data with those themes.

Determining the themes or patterns

You can predetermine your themes or patterns and then look for them in the data (*a priori codes*) or you can look at the data and generate themes or patterns that emerge (**grounded codes**). There is no right or wrong way; it depends on your needs. If you are evaluating press releases after training an organization on how to improve their outreach, you will probably have a set of things you are looking for; in this case you would probably have predetermined themes. If you are analyzing transcripts from an interview, you probably have major categories of information you are looking for, but not the theme itself; in this case you will probably look for patterns. Often, you will do both: start with some predetermined themes, and refine them as you go through the data.

Coding the data

Coding data sounds complicated but is a simple process. Work through the data and highlight or otherwise indicate in the content what piece belongs to what code. You can also code for things like the placement of an article on a page, length of an article, tone of an article, etc. If you already know what you are looking for, it can be helpful to develop a coding review form that helps guide your coding. Here are some tips for coding your data:

- When coding, it's important to write memos to yourself at the same time. That way you can record your ideas, reactions, etc., and track draft findings along the way.
- Sometimes after you code data you may find that you have to go back and collect more data if there is not enough already available. The more you prepare in advance and design strong data collection tools, the less likely this will be.
- Be careful to not make more of the data than what is really there! It is tempting to hang on to an interesting theme, even if it becomes clear that it does not really bear out in the data.
- Sometimes if you see a piece of data that seems important, but you are not sure why, you can set it aside. Go back and look at your sidelined data and see if there are other themes to explore.
- Be aware of the biases you bring to analysis. As Patton observes, you are putting your worldview onto someone else's world. You are putting meanings onto someone else's perspective.¹⁷ Patton suggests that one way to test the robustness of your worldview is to

¹⁷ Patton, Michael Q. *Qualitative Research & Evaluation Methods*. 3rd Edition. Thousand Oaks, CA: Sage Publications, 2002.

present the codes to people whose world is being analyzed, such as the program participants themselves.¹⁸

If you already know what you are looking for, it can be helpful to develop a coding review form that helps guide you through the coding.

Step 2: Turning Data into Findings

Once you have the data coded and organized, findings will begin emerge. You need to look through the coded data to interpret what you are seeing.

There is no secret recipe to establishing a finding. You can quantify (enumerate) some data from a code, for example: “6 out of 10 respondents indicated that they could not attend the training due to lack of affordable transportation.”

The critical point is that a finding must be substantiated by sufficient evidence. For each finding, explain how often it was reinforced in the data, the kinds of individuals that attested to it, etc. Here is a (fictional) example table drawing out findings to see the evidence for each question:

Evaluation Question	Interview Question	Finding/Response
Have citizens increased their participation in local government decision making in effective ways, and if so, how and why?	1) What is your opinion about the relationship between local government and citizen advisory groups?	Respondent 1: “... During the meeting, it was clear to me that the relationship is bad, because the representatives of the citizen advisory group were too afraid to say anything...”
		Respondent 2: “... After the training, I realized that part of my responsibility is to try to make my people's voice heard, even if it is difficult... but it was so difficult, during meetings.”
		Respondent 3: “... By the third roundtable discussion I felt like we were making improvements, making progress. I felt like the representatives from the advisory group actually had helpful feedback, rather than just complaints. Also, I am more comfortable talking to many of them. Before, I was afraid to talk to them because I know they are angry, because they don't understand.”

Quantitative Data Analysis

Generally, quantitative data analysis serves two purposes: 1) to determine characteristics of a specific population, and 2) to generalize characteristics from a sample group to a larger population.

Generalizing characteristics to a larger population

Let’s say you want to extrapolate findings from a sample population to a larger population. This can be done through public opinion polling – you administer surveys to a relatively small group of people in order to draw conclusions about trends within an entire country’s population. This type of analysis requires highly specialized statistical and methodological skills and should not be attempted without

¹⁸ Ibid.

assistance from an expert! As such, this handbook will focus on the second type of quantitative data analysis, which focuses on a specific population.

Determining characteristics of a specific population

Let's say you have gathered pre- and post-test data from participants in a training. Your analysis will be limited to that group of participants: did they increase their knowledge? What topics did they already know a lot about? What questions did they still get wrong in the post-test (i.e., what topics could be better explained at future trainings?). This type of data analysis is fairly straightforward because you are drawing direct conclusions about a small group of people based on data taken directly from them.

If you use an online platform such as Survey Monkey, analysis is very easy because results are set up in a user-friendly way. In contrast, if you have hard-copies of the questionnaires, you will need to do most of the data entry and analysis work yourself.

You can do data entry in different ways, with Microsoft Access or Excel being most commonly used. Microsoft Access requires more work on the front-end and is difficult to tweak, but it is easier for other people to upload data without making mistakes. Microsoft Excel tends to be the default because it is straightforward to use and highly adaptable. This section provides advice using Excel as the platform.

Step 1: Assign an identification number to each participant

Creating identification (ID) numbers is important for confidentiality: to protect the identity of the participant, but also to reassure the participant that his/her answers will not be used against him/her. There are different ways of creating IDs. You only need them for multiple use questionnaires, such as if you are tracking responses over time in the form of a baseline and endline. Single use questionnaires do not need IDs, *unless* you are collecting demographic information at the same time as other confidential information.

Assigning IDs can be relatively straightforward. At the top of each form, ask the participant the following information:

- What is the last letter of your first name? ---- *E.g., O*
- What is the last letter of your last name? ---- *E.g., E*
- What is the number of the month were you born in? ---- *E.g., 8*
- What is the first letter of the city you were born in? ---- *E.g., A*

Their unique ID is: OE8A

This information should not change, and it should be not be sensitive; of course, pick options based on the specific circumstance! You will need to ask them this information each time they fill out a form.

Step 2: Enter the data into an excel spreadsheet

There is no strict rule, but for ease of sorting and counting, formulate your spreadsheet in this way, with participant/cases listed in different rows, and the variables listed in different columns:

Training Database

Participant ID	Gender (0 = Male; 1 = Female)	Party	City	Age	2. To what extent do you think you learned about the following topics (1 = Already Knew; 2 = Little; 3 = Some; 4 = Quite a Bit; 5 = A Lot)				3. During the training, we talked about a variety of campaign methods. Which of the following do you expect to do? (Check those you intend to do) (1 = intend to do; blank/0 = did not check)		
					2a - Topic A	2b - Topic B	2c - Topic C	2d - Topic D	3a - Door-to-door	3b - Cold calls	3c - Fundraising
101	0	Party A	City X	29	1	3	4	4	0	0	1
132	1	Party B	City X	33	3	5	4	3	1	0	1

Your questions may extend very far out to the side – that’s ok, it’s what Excel is made to do! If you are doing a baseline and endline questionnaire (such as a pre/post-test, or a post-test after six months), then it is recommended that your original database include the post-test questions. It should be like the following:

Baseline/Endline Training Database

Participant ID	Gender (0 = Male; 1 = Female)	Party	City	Age	Question 1: Have you ever used the following campaign methods? Please select all that apply					
					Baseline			Endline		
					3a - Door-to-door	3b - Cold calls	3c - Fundraising	3a - Door-to-door	3b - Cold calls	3c - Fundraising
101	0	Party A	City X	29	1	0	0	1	0	1
132	1	Party B	City X	33	1	0	1	1	1	1

Open-ended questions require a little more work. First, after reading through them, develop some natural categories. Then, add in a variable column for each of those categories in your database, and indicate with a “1” whether that respondent gave an answer to that category. Like this:

Participant ID	Gender (0 = Male; 1 = Female)	Party	City	Age	Please describe one or two reasons why you think your radio program is not reaching its target audience.				
					Technical Challenges	Show Timing	Topic Selection	Other	Comment
101	0	Party A	City X	29	1				“We do not have the proper equipment.”
132	1	Party B	City X	33		1			“The station won’t let us air our show during peak times””

This way you can count how many respondents mentioned each category, and also get a sense of other responses.

Step 3: Entering invalid or non-responses

Sometimes the questionnaire is not clear and the participant answers a question incorrectly (such as by providing multiple answers for a “select only one option” question). If the invalid or non-response rate is high, then your data may not be useful to aggregate.

Usually you can work with what you have. For invalid or non-responses, enter a random number that is not used in any of the database, like -9. For example:

Participant ID	Gender (0 = Male; 1 = Female)	Party	City	Age	2. To what extent do you think you learned about the following topics (1 = Already Knew; 2 = Little; 3 = Some; 4 = Quite a Bit; 5 = A Lot)			
					2a - Topic A	2b - Topic B	2c - Topic C	2d - Topic D
					101	0	Party A	City X
132	1	Party B	City X	33	3	5	-9	3

You can count the instances of this number through Excel. *Do not leave the cell blank or with a zero!* Zeroes are counted and factored into averages you automatically calculate and blanks are usually ignored. Also, you may in the future need to know how many invalid or non-responses appeared; if you use a blank or zero, you probably won't recall whether a blank or zero in the data set is there (a) because that was a legitimate option and the participant chose it, (b) because you incorrectly entered the data, or (c) because it was actually a non-response or invalid response.

➤ *Excel has handy tools and online tutorials for analyzing data according to all of these variables.*

Step 4: Analyzing the data

There are several ways of working with the data in the database to produce findings. Here are some ideas:

Counting: As simple as it sounds, enumerating the number of responses in a descriptive form can be very informative. For example: “20 of the 30 respondents responded that they ‘learned a lot’ about topic D,” or “85 percent of respondents requested more advanced training on fundraising.” It is easy to count in Excel.

Cross-Tabulation: Cross-tabulating is the process of comparing several variables across all or some of the respondents. For example, you may be interested to know whether the women that came tended to be younger or older women:

ID Number	Age				Gender (0 = Male; 1 = Female)
	18-30	30-45	45-60	60+	
001			1		0
002				1	0
004				1	0
008			1		0
009			1		0

Here, you can see that the women who came were between 18 and 45, while the men were all 45 or older.

Using cross-tabulation is quite easy in Excel! Just access the sorting/filtering function.

Mean: The mean is what is normally understood as the average. With the mean you simply add up all the values of a variable and divide it by the number of respondents.

Mode: The mode is the value that is repeated the most. For example, with the mode you could say: “Category B was the most-selected response.”

Median: The median is the value that is in the middle of the data set. So, out of nine values, the fifth value is the median. This is not very useful when analyzing questionnaire responses.

Note: Be careful creating an average (mean) out of any data that is presented as words. For example, if a respondent selects a worded response from five choices (1 = strongly disagree; 2 = disagree, 3 = neutral; 4 = agree; 5 = strongly agree), you can give the percentage of people that reported a particular response (50 percent responded “strongly agree”) or the absolute number of people who selected that response (10 out of 20 answered “strongly agree”), but it might be misleading to say that the average respondent gave an answer of 4.3 out of 5 because a score of 4.3 was not assigned a value as a response option.

Step 5: Assessing the findings¹⁹

After you develop an initial list of findings, it is important to discuss these findings either with the program team, or with the participants themselves.

With the program team, ask yourselves these kinds of questions:

- Does this finding ring true?
- Why do you think you have this finding?
- What does this finding mean for us?
- How can you act on it to improve your program?
- What actions should you take, and who is responsible for carrying them out?

With the participants, potential discussion topics include:

- Does this finding surprise you?
- Does this finding fit your experience?
- What actions should be taken to respond to the findings?

Data Use

Data that is not used is a waste of time, energy and resources. Data can be used for more than just reporting to your indicators. Data can:

- Determine program progress
- Improve activity implementation
- Inform strategy (should you stay on course or adapt?)
- Communicate to funders
- Communicate to other stakeholders
- Inform future programs

Here are some examples of ways to ensure that the data is being used:

¹⁹ Discussion questions based on IRI’s experience and informed by Patton, Michael Q. *Utilization-Focused Evaluation: The New Century Text*, 4th Edition. Thousand Oaks, CA: Sage Publications, 2008.

- When planning to collect data, determine how it will be used at the outset, and assign someone the responsibility of using the data (whether in reports, to inform program plans, etc.).
- Make data reports a part of weekly staff meetings; have program staff report on the data that has been collected and discuss what the data mean for the program.
- Organize regular learning hours for program teams to reflect on the program thus far, using collected data.
- Have program staff draft regular data reports (such as analysis of questionnaire responses) and have them include recommendations based on the data.
- Share the data with your donors in meetings with them.

Data Storage

Data storage is important to consider for three main reasons: (1) institutional memory, (2) security of sensitive data (names, opinions, etc.) and (3) transparency of findings/reporting.

When deciding on how to store your data, consider the following:

- What data is important to retain?
 - Do you need to make back-up copies of the data? (Usually advisable!)
 - How long do you need to retain the data? Can there be an expiration date?
- Who needs access to the data?
 - Do you need to protect any information?

Ethics of Data Collection and Analysis

Ethics plays an important role in evaluation in two key ways:

1. Protecting your sources (i.e. research participants).
2. Ensuring the integrity of the evaluative process.

You can subscribe to the American Evaluation Association's *Guiding Principles for Evaluators* which can be found on the AEA site: <http://www.eval.org>.

Ethics Fundamentals

Here are some fundamental ethical considerations when conducting any form of data collection:

Gain informed consent from any interviewee or data collection participant.

Participation should always be voluntary and not subject to any pressure. Informed consent means that the researcher provides enough information about the evaluation to research participants so that they are able to make an informed decision about whether or not they want to participate in the research. Here are some things to consider including in an informed consent statement:²⁰

- Briefly discuss the purpose of the research

²⁰ Please note that this is not a definitive list of items to be included in an informed consent statement. We recommend you refer to guidance developed by your organization/institutional review board, funder, or other oversight body to ensure that your informed consent statement meets established criteria.

- Explain what the research will involve (time requirements, etc.)
- Tell the participant about any risks they might incur or benefits they could gain from participating
- Explain how the data will be used and/or distributed. Will data be confidential? Anonymous? Who will have access to the data during and after the research? For example, it is good practice to let the interviewee know that s/he will never be referred to directly in a report by name and that evaluations report on findings that resonate across multiple interviews (i.e., interview responses are aggregated).
- Explain how data will be captured (note taking, recording, etc.). If recording, the participant needs to give permission to be recorded.
- Finally, make sure the participant that their participation is voluntary, that their decision to participate or not will impact any relationships s/he has with the implementer, and that they may choose to stop participation at any time.

Throughout the data collection process, monitor the participant to ensure (s)he is comfortable and still consents to be part of the effort.

Sometimes people don't really understand what is being asked of them at the outset. So it's your responsibility to ensure that throughout the process, s/he is comfortable and still wants to participate.

Ensure security of people and data.

Security is very important to consider at the outset of any evaluation effort. Think through what you need, including the following:

- Security of data: Use passwords, do not keep hard copies, encrypt any audio recordings, etc.
- If you promise anonymity, deliver it! Anonymity means that nobody, including you, knows who the person is.
- If you promise confidentiality, deliver it! Confidentiality means you know who the person is but won't reveal that information.
- If using a direct quote, attain that person's permission to cite them in that context before doing so; otherwise, do not include the person's name or any other identifying information.
- If you receive requests (from funders, other evaluators, etc.) to turn over your raw data, make sure you are not violating any promises of confidentiality/anonymity before agreeing to such requests.

And respect cultural sensitivities!

Chapter 4: Evaluations – Taking Stock of Your Programs

Although evaluation is the “E” in M&E, people are often confused about what evaluation really means. Too often, evaluation is lumped in with program monitoring, the assumption being that simply collecting data about the program means that the program is being properly evaluated. For example, the indicator matrix is focused on the “M” in M&E, meaning monitoring, since it predominantly monitors program performance and progress toward results. But simply collecting and recording data is not evaluation, although this work can, and should, support evaluation.

➤ ***The results of an evaluation can be used long after the program has ended to help inform future work. For these reasons, you should include evaluation results as part of the program’s permanent record and disseminate the results throughout your organization.***

So what exactly is evaluation? Consider the E in M&E to be the process by which data is collected and used in a systematic way to answer questions focused on the how, why, and so what. Whereas with proper monitoring you can tweak the program along the way, with evaluation you can make the bigger decisions: Should you change program direction? Are you achieving your desired results? How are you achieving them? Should the training program in year two adjust its training content and delivery mechanism?

While many monitoring activities from your indicator matrix are evaluative in nature (they tell you whether you are on track and help you change course), they do not constitute a formal evaluation. These types of evaluations pose a question that provides answers that you can use to assess program results and program implementation, or to promote learning and accountability. Data is collected and analyzed to answer that question. In contrast to an indicator matrix, an evaluation brings together multiple activities with the express purpose of answering a particular question about the program.

Why a Formal Evaluation?

A formal evaluation is important because it takes you out of the daily focus of program implementation and indicator measurement, and helps you think more broadly about the program and how you are approaching it.

There are a myriad of situations where you would want to conduct an evaluation. Here are just some examples:

- You know that one part of your program is seeing a lot of unexpected results, but your indicators have not been capturing them. You would like to evaluate these results so that your reports can include these successes and contribute to organizational learning.
- You are halfway through a long-term

Quick Note: Evaluation versus Assessment

Bottom line, the terms **assessment** and **evaluation** are often used interchangeably to mean the same thing. However, they are sometimes differentiated in the following way:

- In an **assessment** you analyze information *to make decisions about a program*.
- In an **evaluation** you *analyze achievement, often against a set of predetermined standards*. Most evaluations need to begin at the outset of the program to collect baseline information. They are often completed at the end of a program to evaluate its overall success, but can also be done during the program to inform decisions.

training program. You want to know which parts of the training program curriculum and if the teaching method are working, and where you can improve for the next round of trainings.

- You are coming to the culmination of a five-year program. Over five years, the program is bound to have achieved impact, but you are not sure what to look for since the program has changed so much over the course of the grant period. You would like to evaluate the entire program at the end to help explain the program in the final report and contribute to organizational learning.
- You are worried that perhaps the program is not as relevant to the current political environment as it should be.

Formal Evaluations: Who Conducts? Who Commissions?

Depending on who commissions and/or conducts an evaluation, it is described as either internal or external. You usually refer to **internal** evaluations when they are undertaken by your organization's staff, whereas **external** evaluations are conducted by third parties (like an independent research firm).

However, even if the evaluation is conducted by a third party, if your organization determines the evaluation questions and selects the evaluation firm, it is technically still an "internal" evaluation because the program controls the process – the evaluation firm ultimately answers to us. Thus, whether an evaluation is considered internal or external actually depends on who controls the evaluation, which has repercussions for the perceived objectivity of the evaluator.

Internal Evaluations Conducted by Program Staff

Organizations can internally implement most evaluation designs, though it is generally advisable to outsource larger, more quantitative or summative evaluations to a third party to increase objectivity and credibility.

When is an internal evaluation appropriate?

- Your program operates in a politically sensitive environment, and having someone external come in and interview your stakeholders is simply unfeasible.
- Your program does not have the funds to cover an external contract.
- You need someone who really knows the history of your program.
- You need an evaluation done quickly.

When deciding whether an internal evaluation is appropriate, consider the following pros and cons.

Pros:

- The evaluation is conducted by someone who understands the programs and is sensitive to the realities on the ground.
- The evaluation is conducted by someone using your program's contacts and relationships.
- The evaluation is often more feasible as program staff are involved from the start in the evaluation conceptualization.
- Regular communication is easier as a result of immediate program buy-in.
- Internal evaluators are often better placed to craft recommendations to which internal stakeholders can and will commit.
- Internal evaluations can often be accomplished more quickly and cost less.

Cons:

- Staff do not always have the level of expertise to conduct specialized evaluations or to manage quantitative data.
- Staff have vested interests in the outcomes or have ingrained assumptions which may affect evaluation design and data analysis. These issues can introduce bias into the evaluation, which means the evaluation might be less rigorous and helpful.
- Staff may simply not have enough time to conduct an evaluation.
- Staff conducting the evaluation may have working relationships with program staff managing the program, affecting objectivity either because they do not want to jeopardize future relationships or are prejudiced in favor of the program.

External Evaluations Commissioned by Your Organization

Evaluations commissioned by your team but undertaken by a third-party may include members (or teams of members) of full-fledged evaluation firms, independent academics or practitioners. Many donors now consider an evaluation to be a basic component of a program, with preference given to the use of third-party evaluators for reasons of perceived objectivity and rigor. For this reason, when appropriate, please consider including an external evaluation as part of your M&E plan.

External Evaluations Commissioned by the Donor

Funders are increasingly commissioning evaluations of DG work through outside evaluators, academics and evaluation firms. This is an opportunity to showcase your work! While this can sometimes be perceived as a scary experience, it does not need to be. Donor-driven evaluations can be an opportunity to be recognized for your achievements. In addition, if you are able to provide constructive input into the process, you may be able to add in sub-questions to get information that would be of use to your program.

Tips for Working on a Donor-Driven Evaluation:

- *Provide as much information on program theory as possible at the outset*
Evaluation criteria for goals-based evaluation – the most prevalent type of evaluation – depend on a clear explanation of what the program intends to achieve. Because your proposals are often approved far in advance of program implementation, and because of the complex environments in which you work, your programs often do not look exactly like what was originally proposed. However, original proposals are often the only thing that evaluators have to work off of when developing the evaluation design. For this reason, provide the evaluator with the most updated workplan, results chain or LogFrame, at the outset. If they don't ask for it, volunteer it!
- *Carefully review the evaluation design and methodology*
Often, evaluation designs are developed in a vacuum, with little awareness of the program context. Evaluators rarely have as much knowledge about the country and the program as you do. For this reason, it is important that you help the evaluation by looking at the design and seeing if you think it is realistic. Look at the timeline: are elections or other events coming up that could prevent good data collection? Look at the methods proposed: do you happen to know that some organizations are particularly biased, and can you suggest other organizations to include in order

to balance the sources? Look at the evaluation questions and criteria: are these questions of use to you? Are the criteria appropriate, given your program design? Has the evaluation appropriately considered the program's goals and expectations?

- *Make sure to collect data for your indicators and analyze the data rigorously (throughout the implementation of the program)*

An evaluation is only as good as the data off which it is based. An evaluation team will almost always look at the data collected for your indicators; often an evaluation team will depend on that data! This means that the data in your indicator matrix needs to be high-quality data that is rigorously collected and analyzed. It also means that if you haven't been collecting data for your indicators, not only will the donor now be acutely aware of it, but the evaluation will suffer, and the evaluator may not have enough data to discuss your results!

- *Know your rights!*

You have a right to read the evaluation design and methodology, and to know the criteria against which you will be judged. You have a right to know the competence of the evaluators. You have the right to know upfront what the expectations are insofar as program responsibilities in the evaluation and its timeline (such as providing data).

- *Ensure that the evaluation safeguards the integrity and safety of program partners*

Your program will be in the country long after the evaluation team has left. Thus, the program has a much greater incentive to maintain relationships. It is important to engage in a discussion with the evaluation team to know how it plans to ensure these relationships, or the program, are not harmed.

- *When in doubt, ask!*

You should never, at any point in the evaluation, be confused or in the dark as to what is going on. Ask the evaluation team or ask your donor.

Tips for Reviewing Draft Evaluation Reports:

When an evaluation is complete, most likely you will be given an opportunity to review a draft report and to provide comments.

- ***Transparency***

There must be transparency of purpose, design, methods, data, findings and recommendations, including the inclusion of any tools and templates used.

- ***Accountability***

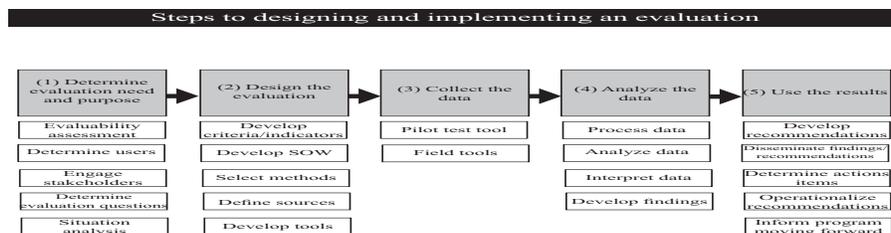
The report should be accountable to principles of ethics, such as participant confidentiality and security.

There are two items in a report that are dependent on your careful read of the draft report: factual inaccuracies, and omissions and requests for more information.

- ***Factual inaccuracies***
Because the evaluation team is not as knowledgeable about the program as you are, it is bound to make factual errors about dates, names, locations, etc. These mistakes are natural and should not discount the validity of the evaluation: simply provide the correct information.
- ***Omissions and request for more information***
Sometimes an evaluation finding will seem strange or counter to what you have observed. In these situations, it is important to request more information about how the finding was derived. All findings should be substantiated with evidence from the data itself.

Designing and Implementing an Evaluation: What to Consider

Evaluations come in all shapes in sizes. They also can take place over different time periods, from a few days to several years. There is no right way to do an evaluation; the only gold standard is if the evaluation serves its purpose in the most rigorous way possible given available resources. However, there are some basic steps to undertake when designing and implementing an evaluation.



Step 1: Determine the Evaluation Need and Purpose

The need for and purpose of the evaluation will drive all decisions about its design, methods, analysis, etc. It is important to think through whether an evaluation is appropriate at this time: this is called an evaluability assessment. Not all programs are ready for certain types of evaluations.

The evaluation question should address the purpose of the evaluation and inform the design.

It is also important to think through what resources – funds, expertise, time, etc. – are available to do the evaluation. This is called a situation analysis.

Evaluability Assessment

An evaluability assessment determines whether an evaluation is possible and worthwhile. It asks these kinds of questions:

- Is the program designed in such a way that allows for evaluation?
 - Are objectives clear?
 - Is the program logical with the underlying theory justified?
 - Are expected program results clear?
 - Does the program keep sufficient records?
 - Is there sufficient budget?

- Is it feasible to collect data for the evaluation?
 - Are there sufficient data sources?
 - Would program managers participate in the evaluation by providing records and facilitating data collection as necessary?

- Would the evaluation be useful?
 - Is the program at a stage where an evaluation would be used?
 - Would an evaluation be credible to stakeholders?
 - Are intended users interested in an evaluation at this time?
 - Is there sufficient buy-in?

At the end of the assessment, you should be able to decide whether the evaluation should take place, or whether the program needs to be tweaked or thought through more to prepare for an evaluation.

Determine Evaluation Questions

Your evaluation question will depend entirely on the need for and purpose of your evaluation. Here are some general purposes for an evaluation commissioned or conducted by internally, along with a corresponding example:

- To assess **results**. Example: As a result of your training program, have participants more effectively advocated on gender issues to local government officials?
- To assess **implementation**. Example: Was the timeline for the intervention appropriate? Were the right regions for the intervention selected?
- To promote **learning**. Example: Do elected officials respond to constituents differently based on whether the program or civil society organized the town hall meeting?
- To ensure **accountability**. Example: To what extent has your program delivered on objectives as set out in original program design?

When a funder is evaluating a program, it will often look at the following criteria²¹:

- **Relevance**: Was the program suited to the priorities or policies of the beneficiary and donor?
- **Effectiveness**: Did the program achieve its objectives?
- **Efficiency**: Was the approach cost-effective in relation to its achievements?
- **Impact**: What were the positive and negative effects of the program?
- **Sustainability**: To what extent will program results continue after the program has ended?

²¹ *Principles for Evaluation of Development Assistance*. Paris: OECD-DAC, 1991.

From these main purposes and criteria, you then develop evaluation questions. An evaluation question is generally comprised of major question(s) along with their sub-questions that the evaluation will seek to answer. Here is an example:

Main question: To what extent did the intervention contribute to more effective candidates for elections?

Sub-Question: To what extent did the door-to-door campaign training contribute to the implementation of the door-to-door campaign technique by party members?

There are different types of evaluation questions, but in general they fall under the following categories:

- **Descriptive:** Descriptive questions ask “what is.” They describe a program, measure change, observe a process, describe results or provide a snapshot of the state of a program component.
- **Normative:** Normative questions ask “what should be.” They compare the program against benchmarks, expectations or other values.
- **Cause-effect:** Cause-effect questions try to determine attribution or contribution. They look at causal relations.

Your evaluation questions can be a mixture of these types. A good evaluation question is one whose answer will be used! Ask yourself:

- Would the answer to my question be of interest to key audiences?
- Would the answer to my question reduce uncertainty?
- Would the answer to my question yield important information?
- Would I be able to act on any answer to the question?

Situation Analysis

Once you have your question selected, it is important to think through how best to answer the question. This will be affected by your resources, both internal and external. A situation analysis considers the following:

Internal:

- Key people and their expertise
- Time constraints (grant end date, reporting requirements and staff time)
- Budget constraints
- Logistical constraints (transportation)
- Buy-in

External:

- Security
- Buy-in constraints from stakeholders
- Other environmental constraints

Step 2: Design the Evaluation

The evaluation design depends on the evaluation question and your situation analysis. Every evaluation is different! Here are some components, frameworks and approaches that you will most likely consider in designing the evaluation.

Developing Evaluation Criteria/Indicators

Most evaluations and their questions need a set of defined criteria against which to measure what is being evaluated. These can be specific indicators or expectations (such as expected results). The evaluation will look at the current state of affairs and compare them to these indicators or expectations.

Indicators or expectations can come from a number of sources. Here are some examples:

- The objectives and expected results as defined by the program through the proposal and workplan.
- A baseline or a prior period of performance (such as a previous grant).
- Academic research or other analysis relevant to the program, including expert opinions.

Evaluation criteria should be appropriate, relevant to the program and sufficient to answer the evaluation questions and overall evaluation purpose. It is important to ensure that the evaluation end-users buy into these criteria before the evaluation begins. This is generally done through the inception report or scope of work that details the design, which is developed by the evaluator.

Major Categories of Evaluation Designs

The evaluation design depends entirely on the purpose of the evaluation and the evaluation questions and sub-questions. At times, a funder or the evaluation commissioner will prefer a specific design or method. Note that some designs or methods are not appropriate for some types of questions. Evaluations can happen at any time during the life of a program, depending on your need and purpose. Here are types of designs focused on different periods of the project lifecycle:

- ***Formative evaluation*** is used to make decisions that inform and improve program design and implementation. It is generally conducted at the beginning of a program, or part-way through, to inform direction or tweak approach. Examples of formative evaluation include: needs assessment, stakeholder assessment, baseline assessment, systems mapping, community mapping, etc.²²
- ***Process evaluation*** is used to assess the effectiveness of the process by which the program is being implemented, and whether the program is reaching its milestones. A process evaluation can look at just about anything: whether the M&E system is providing and disseminating information properly, whether a training program is targeting the right audience and is appropriate to

➤ ***Evaluations with an experimental design are often referred to as randomized control trials (RCT). In the field of DG, they are often referred to as impact evaluations. For the latter, note that this does not refer to “impact” as it is commonly defined in a results chain.***

²² “Formative” and “summative” as first defined by Michael Scriven. For more information, please refer to: Scriven, Michael. “The methodology of evaluation”. Stake, R.E. *Curriculum Evaluation*. No. 1. Chicago: Rand McNally, 1967, pp. 39-89.

participant needs, whether the program met or diverged from the intended strategy, or whether the milestones or objectives have been achieved.

- **Summative evaluation** is used to look at what the program has resulted in, often at the outcome or impact level. It often compares the results of the program to its original objectives, but it can also be **goals-free**. Summative evaluation is what people normally think of when they think of an evaluation.

An evaluation can focus on one or several of these time periods simultaneously. For example, a summative evaluation should inform the next program, and so it is also a formative evaluation. A process evaluation can also be done at the end of the program to understand program milestone achievements.

Evaluation designs can also be defined by the criteria (or lack thereof) against which an evaluation will be evaluated²³:

- **Goals-Based**
The vast majority of evaluations are goals-based, in that they evaluate the program based on the explicit program goals (objectives). Most evaluations that use criteria, indicators or expectations are goals-based.
- **Goals-Free**
Goals-free evaluations are often used in situations where it's not clear what the goals are/were, or where the situation is changing so quickly that the stated goals may no longer be relevant. Developmental evaluation is an example of a goals-free evaluation.

	What it is:	Situation you would use it:
Goals Based	Goals-based evaluation uses specific criteria against which the program is compared. Evaluation focused on accountability usually requires goals-based evaluation.	If you want to know whether the program has achieved, or where it is relative to achieving, its objectives or expected results. If you want to know whether the theory of change is accurate or appropriate.
Goals Free	Goals-free evaluation ignores any predefined program goals, objectives or expectations, and just looks at what has been accomplished, both positively and negatively. It often assigns value to that accomplishment. Evaluation focused on learning often has goals-free components.	If your program is in formative stages or is being conducted in a highly complex environment where it is not possible to establish clear criteria at the program outset. If you want to know what has been achieved and do not need to know whether that was in reference to something specific; this is often the case for learning purposes.

An evaluation can include components of both a goals-based and goals-free evaluation, depending the evaluation questions and sub-questions.

²³ The terms “goals-based” and “goals-free” as first defined by Michael Scriven; see: Scriven, Michael. *Evaluation Thesaurus, 4th Edition*. Newbury Park, CA: Sage Publications, 1991.

Finally, evaluations can be categorized according to whether and how they define a **counter-factual**, or what would have happened had the project not taken place. Evaluations do this by creating comparisons between groups that have received the program and those that did not. Evaluation methodologies that use comparison methods are grouped into three categories of experimentation based on how the comparison is achieved. The degree of experimentation that is used depends on your program and situation. Generally, the more experimental the evaluation, the more rigorous the results. However, this does not mean that one degree is necessarily *better* than another. The best methodology is the one that most directly addresses the purpose of the evaluation, the needs of the program, and the context in which the research must take place.

	What it is	Benefits	Drawbacks
Experimental	Total randomization of treatment and control group.	Very high internal validity.	Randomization may not work; expensive; complicated; not always possible or practical; some ethical considerations.
Quasi-Experimental	Randomization is not total; certain groups are selected, and units are randomly assigned to a treatment or control/comparison group, or there is no randomization at all.	Closer to true experimental design. Best when randomization is not possible or not practical.	Since units are not randomly assigned, there could be alternative explanations for the results.
Non-experimental	No randomization, no separate comparison/control group. You can compare the same group of participants before and after the program. Subjects can act as their own pseudo-control group if there is a baseline.	Easier to implement, less expensive, can serve as pilots and help to identify variables that are important to the program's success.	Participants can improve over time without participating in the program and this would not capture that; often mistaken attribution; context can change; difficult to interpret findings.

The majority of evaluations conducted within the DG field are non-experimental. At times you are able to achieve a quasi-experimental design. There have been few truly experimental designs carried out, and these have tended to focus on outputs and lower level outcomes, as opposed to larger outcomes and impact. However, elements of experimental design, such as randomization and control/comparison groups, can be incorporated into facets of the various evaluations. For example, simply collecting baseline information can help constitute a form of comparison group, since it helps show what the state of the system was before the program began.

Evaluation Approaches

In addition, there are numerous approaches to evaluation. Consider the following:

- ***Participatory***
A participatory approach includes all stakeholders in all aspects of the evaluation, from the design of the evaluation questions themselves to the data collection and analysis. You may hear some forms of this approach called democratic evaluation, stakeholder-based evaluation, or participatory action research. This can increase the relevance of, ownership over, and utilization of the evaluation.
- ***Empowerment (or transformative)***²⁴
An empowerment approach uses evaluation concepts and methods to increase the capacity of stakeholders to improve their own program or service. By teaching them about evaluation throughout the evaluation process, you will increase their capacity to monitor and evaluate; by increasing their capacity to undertake monitoring and evaluation, you are also increasing their capacity to achieve the goals of their own programs.
- ***Appreciative Inquiry***
An appreciative inquiry approach takes the view that focusing on the positives through evaluation can be more effective, especially in situations where there is fear or skepticism surrounding evaluation, when stakeholders are unfamiliar with each other, when relationships have soured, or when you want to build appreciation for evaluation.
- ***Utilization-Focused***²⁵
A utilization-focused evaluation approach judges the merit of an evaluation on whether it is used. Thus, the entire evaluation is built around its use. All decisions are made in light of whether it will increase the usability and credibility of the evaluation with the end users.

These approaches influence how you undertake the evaluation design. An evaluation can incorporate one or several of these approaches.

Step 3: Collect the Data

Data collection for evaluation purposes is very similar to data collection for monitoring purposes. The only difference is that data for a formal evaluation is streamlined to answer specific evaluation questions.

Step 4: Analyze the Data

²⁴ The empowerment evaluation approach was developed by David Fetterman. For more information, see Fetterman, David and Abraham Wandersman. *Empowerment Evaluation Principles in Practice*. New York, NY: The Guilford Press, 2005.

²⁵ The utilization-focused evaluation approach was developed by Michael Quinn Patton; for more information, see Patton, Michael Q. *Utilization-Focused Evaluation: The New Century Text*. 4th Edition. Thousand Oaks, CA: Sage Publications, 2008.

Data analysis for evaluation purposes is very similar to data analysis for monitoring purposes. The only difference is that data for a formal evaluation is streamlined to answer specific evaluation questions.

Step 5: Use the Results of Evaluations

As this handbook has noted on numerous occasions, M&E findings and recommendations are useless unless they are *used!* To help ensure use, here are some ideas:

- **Engage intended users from the start²⁶**
Make sure that the people who will be using the evaluation results (most often, implementing staff and/or senior leadership) are part of the design and implementation decisions.
- **Focus group the findings to inform recommendations**
When findings have been developed, sit down with knowledgeable stakeholders and present findings to them. Work with them, decide what these findings mean, and turn them into actionable recommendations.
- **Lead a learning workshop to turn recommendations into action items**
After recommendations have been developed, lead a workshop with all stakeholders to develop a workplan to operationalize the recommendations. Use a response matrix to track progress.

Evaluation Response Matrix				
Recommendation 1:				
Program Response:				
Key Action	Timeframe	Responsible Party	Status	Comments
1.1				
1.2				
1.3				

- **Build in utilization steps into the evaluation itself**
In the evaluation design, build in key moments when the evaluator and evaluation stakeholders ensure the evaluation is useable. For example, build in a deliverable focused on testing question design relevance with staff.
- **If contracting the evaluation, add in a learning workshop as a contract deliverable**
This will ensure that it happens, and it will convey to the evaluator that you are serious about using evaluation results – and will encourage the evaluator to ensure findings and recommendations are actionable.

²⁶ These recommendations are informed by Michael Quinn Patton’s book as well as IRI experience. For more information, see Patton, Michael Q. *Utilization-Focused Evaluation: The New Century Text*. 4th Edition. Thousand Oaks, CA: Sage Publications, 2008.

Step 6: Disseminate Evaluation Results

Not only is it important to use the evaluation results to improve the program and inform future programs, but it is important to maximize the evaluation by sharing lessons learned. Here are some ideas for disseminating evaluation results:

- Develop an evaluation two-pager that lays out the main lessons and results with tips for future programs. Share this inside your organization, with your donor and at any DG events.
- Cite the evaluation results in new proposals as supporting evidence for the validity of your program approach.
- Include a discussion of the evaluation, its results and how your program has addressed the recommendations in reports, especially final reports.
- Use it in your organization's public relations materials and online platforms, including your website, Facebook pages, Twitter feeds and other social media platforms.