RESEARCH DESIGN TO EVALUATE THE IMPACT OF USAID DEMOCRACY AND GOVERNANCE PROGRAMS

Kenneth Bollen
Pamela Paxton
Rumi Morishima

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The main goal of this report is to develop a methodological and analytical strategy for evaluating USAID Democracy and Governance (DG) programs, along with suggestions about administering and managing the effort. USAID requested this report as part of the SORA project.

The structure of USAID places constraints on evaluations that must be accounted for in an evaluation research design. Three of the most important are: (1) a lack of information on mission activities in the Democracy and Governance area, (2) the influence of missions and embassy officials on USAID evaluations, and (3) low incentives for undertaking reasonably rigorous data collection and evaluation. Overall, these constraints point to the need to elevate the understanding, importance, and rewards of good evaluation in the agency. The constraints also point to the need for enhanced collection of data on activities, programs, and sectors in the DG central office. Finally, they indicate that at least some evaluation methods should be less dependent on missions than has been true in the past.

There are a number of trade-offs involved in doing evaluation. Two of the most important are (1) the trade-off between breadth of coverage and depth of coverage and (2) the trade-off between retrospective and prospective evaluation. USAID will be unable to evaluate *all* of its activities and programs, and it may have difficulty unambiguously attributing impacts to USAID funding, if only retrospective evaluation methods are used.

This document makes a number of sequential recommendations:

- First and foremost, USAID must improve central office information on the inputs of activities, programs, and sectors, as well as related outputs, outcomes, and impacts. Collecting basic information on USAID programming must be accomplished before activities, programs, or sectors can be chosen for evaluation.

- Once basic data collection is complete, we recommend that USAID convene a task force, made up of academics and USAID officials, in a series of three meetings. In its first meeting, the task force would choose a set of activities and programs to be investigated in more detail. During its second meeting, the task force would decide what mix of methods to use in evaluating each chosen activity or program. The task force would meet a third time to recommend activities and programs to investigate in the future with a randomized experiment. Although not officially part of the SORA project, here and elsewhere in the document we strongly recommend this prospective focus. The task force would use the data collected as part of the initial evaluation phase, as well as information gleaned from focus groups of USAID officials on evaluation priorities.

- The table below identifies the methods that we recommend the task force consider, and their strengths in terms of evaluating outputs, outcomes, or impacts.

Table 1: Methods of Evaluation			
	Output	Outcome	Impact
Randomized Experiments		√	√
Quasi-experiments		√	√
Surveys of Individuals & Groups	√	√	√
Interviewing and Site Visits	√	√	
Country Expert Sector Overviews			√
Cross-national (quantitative)			√

Each of these methods is illustrated with examples taken from the DG area and the costs of using each method are estimated. We recommend that each activity or program that is chosen by the task force be matched to the optimal combination of methods for evaluation.

- While activities and programs are being designated for further evaluation, cross-national analyses at the sector level could begin. There is data that already exists on spending by sector (i.e., civil society, rule of law, etc.) and this can be used without need for task force input.

The evaluation approach that we recommend differs from the current USAID DG approach to evaluation in several ways. First, multiple methodologies are recommended rather than a single one. Second, more attention is given to evaluations of *impact*, rather than only immediate outputs. Third, we recommend that measures of inputs and impacts be explicitly operationalized and measured consistently over time and across countries. Fourth, we recommend that, wherever possible, control groups (individuals, organizations, or countries that did not receive USAID programming) be investigated as well as the groups that *did* receive USAID programming aid. Finally, we recommend that baseline data (that can be used to compare groups before and after USAID programming) be used.

Activity 1: Research Design. Develop: (1) a methodological and analytical strategy for a multiyear research effort aimed at evaluating the impact of USAID Democratic Governance (DG) programs, the project known as the Sector Operational Research Agenda (SORA); and (2) make concrete, practical recommendations regarding the administrative, management, personnel, budget and level of effort needed to sustain SORA for the life of the program.

INTRODUCTION

In May 2004, the Office of Democracy and Governance at USAID will celebrate its tenth anniversary. During that time period, support for democracy programming by USAID has increased almost eightfold, reflecting the general worldwide increase in democracy creation (Carrothers 1999). In the early years, the focus of programming was on getting started and developing legitimacy in relation to other USAID programming. As the program matures, however, there is increasing interest in evaluating whether USAID's democracy programs are meeting their objectives. This internal trend mirrors external pressures toward more accountability from the GAO and other sources. This document outlines a general strategy for evaluating USAID's democracy and governance programs. Our task is to develop a methodological and analytical strategy for evaluating USAID Democracy and Governance (DG) programs, along with suggestions about administering and managing the effort.

At the most general level, evaluation has been defined by the Joint Committee on Standards for Educational Evaluation (1981) as the "systematic investigation of the worth or merit of an object. . ." This general definition masks important differences in types of evaluation, from organizational reviews to cost-benefit studies. As USAID desires to know which of its activities work and which do not, their focus is on impact evaluation.

An *impact evaluation* assesses changes in the well-being of individuals or groups that can be attributed to a particular program or activity. It is aimed at providing feedback to help improve the effectiveness of programs and activities. Information generated by impact evaluations can be used to inform decisions on whether to expand, modify, or eliminate activities.

In this document, we start by describing aspects of the USAID structure that affect evaluations and then raise some of the issues that affected past USAID evaluations. This is followed by a research design perspective on previous USAID evaluations that sets the stage for the new methodology we propose. Next, we outline our general suggestions on the steps to improve the evaluation process. Fundamental to all evaluation efforts that we suggest is an improvement in the collection of basic data, discussed in the next section. Next, we give an overview of the multiple methods that we recommend. Following this, we have separate sections describing each of the methods that we recommend: randomized experiments, quasi-experiments, surveys, focus groups and site visits, country sector expert overviews, and cross-national longitudinal analyses. The last two sections of the text give examples of implementing the research design and provide the next steps for USAID, along with time lines. Finally, we include a series of appendices that give additional details about topics discussed in the main text.

One issue that we discuss throughout the document is the contrast between the current approach to evaluation at USAID DG and the evaluation approach that we recommend. One indication of the difference is that typical past evaluations relied on only one methodology whereas we recommend five or six methods. In addition, we see the collection of basic data on USAID programming as an essential first step to evaluation. We will say more about this in later sections, but suffice it to say here that we recommend changes in the DG approach to evaluation.[1]

ISSUES RELATED TO USAID STRUCTURE

It is important to outline the USAID bureaucratic structure in which an evaluation of USAID activities would take place. The structure of USAID places certain constraints on any attempt at evaluation that must be accounted for in a research design. We describe three of the most important below.[2]

Structure Problem #1: **There is a lack of consistent, centralized information on mission activities in the Democracy and Governance area.** Activities are designed, beginning to end, by missions officials. Ideas for activities, the allocation of resources, design, choice of implementation partner (where possible), and most monitoring occur within the missions. More importantly, missions are not required to report their ideas or resource allocations to the central office. Thus, the central office, in many cases, does not know specific information about democracy and governance programming across the countries in which USAID is working.

Compounding the problem is the fact that mission officers move from mission to mission every two to four years. Upon arrival at a mission, a new mission director is under no obligation to continue the activities begun by his or her predecessor. Thus, activities may switch substantially from year to year with changes in mission officers. And, there is often a loss of institutional memory on activities after a few years, as most mission staff have moved to new and varied posts. Finally, for a variety of organizational reasons, implementers may in some cases be the most comprehensive source of information about programming, and they have less incentive to retain such information over the long-run.[3]

---

[1] Another issue that we should mention is that USAID staff asked that we concentrate on evaluation methods that are applicable to past activities rather than ones that require the collection of new information. In other words the emphasis should be on retrospective rather than prospective methods. All but one of the approaches that we propose can serve either retrospective or prospective purposes. Although we do make recommendations for prospective evaluations, our emphasis is on retrospective evaluations.

[2] Information for this section came from interviews with USAID staff, internal reports, and the USAID webpage. USAID staff was quite forthcoming and honest about the difficulties the USAID bureaucratic structure poses for evaluation. For an evaluation to succeed one must have a realistic picture of how things work and we are grateful to USAID for providing such candid information.

[3] This is particularly true of the early years of the Europe and Eurasia region.

Structure Problem #2: Agents outside the democracy and governance central office affect USAID evaluations. Mission officials (as well as embassy officials) approve all trips by contractors, as well as USAID/Washington staff, into the country. This approval is commonly known as country clearance. Although the primary purpose of country clearance is to ensure the safety of visitors, it is also used to prevent overload on mission and embassy staff by outside demands. Occasionally, therefore, missions may deny country clearance to evaluators. For example, in a recent six-country study undertaken by USAID, although Peru was part of the original sample of countries, the sample had to be revised when mission officials refused access to the country.

Mission officials may also be influential once evaluation teams have arrived in the country. Mission officials may have their own agenda for the evaluation, naturally wanting to highlight certain areas for positive evaluation. Furthermore, political forces within the country may similarly impact the activities that can be evaluated by a research team. Either may shift the team away from the intended objectives of the evaluation. For example, in a recent study of South Africa, the evaluators were unable to study civil society due to pressure from the ANC.

The incentives of missions and political forces in the country will have to be taken into account in designing and implementing the evaluation. For example, precise country samples may be hard to predetermine. On the other hand, a cooperative mission increases the odds of a successful evaluation effort.

Structure Problem #3: Incentives for undertaking reasonably rigorous data collection and evaluation do not exist. Many missions have very limited funding for democracy and governance programming, and rigorous evaluation can take up a large percentage of a small program. In addition, some of the best evaluation methods require allocating funds or programming based on methodological concerns, rather than maximizing the "likelihood of success," which can make mission officers eschew the most rigorous evaluations, especially in small programs. Finally, mission officers may feel content with the data collection required as a part of "re-engineering," even though, in reality, re-engineering has allowed missions to choose different measurement yardsticks each year. Overall, recent years have seen a decline in the accountability of missions to central offices as many of the checks and balances needed for accurately and objectively measuring progress have been removed.

ISSUES IN PAST EVALUATIONS

A primary reason for this research design project was USAID's lack of satisfaction with most, though not all, past evaluations of USAID activities. In a separate report, we reviewed a sample of evaluation reports that were selected with the help of USAID staff. A detailed account is in Bollen, Paxton, and Morishima (2003). Here we characterize the work as a whole, recognizing that there will be exceptions to each generalization that we make. We focus on problems with previous evaluations since these represent areas that we hope to address in the research design for new evaluations.

Evaluation Problem #1: Over-reliance on a single approach to evaluation. A typical evaluation involved a team of evaluators who used a combination of "desk study" and on site interviews that occurred during a two to three week period. Each report relies heavily on the evaluators' interpretation of these interviews and the desk study information collected. Quantitative information is very limited. When quantitative data are provided, their usefulness is undermined by their lack of correspondence to SORA goals or the data being in too aggregated a form. Overall, these evaluations have predominately relied on qualitative methods based on the interpretation of interviews and reading material. The shortage of time, the limited types of people interviewed, the number of activities or countries, and the heavy reliance on interpretation of this material make it difficult to do a rigorous evaluation.

Evaluation Problem #2: Lack of information on what was done. In a significant number of evaluations, it was not always clear what was done. For instance, the exact timing of the activity, the money spent, personnel invested, and other resources that went into an activity were not provided. Sometimes, the dollar figures for a group of activities might be reported, but it was not clear how these were allocated across activities.

Evaluation Problem #3: Lack of information on what resulted. Typically, the results were not presented for the immediate output, proximate outcome, and general impact. Results tended to emphasize immediate outputs rather than examining outcomes or impacts at the institutional or country level. Some evaluations reported results, but there was so much detail that it was hard to see the forest from the trees. Another problem is that the conclusions are too general such as stating that there is a lack of "political will."

Evaluation Problem #4: Failure to consider alternative explanations. A primary reason for doing an evaluation is to determine which activities worked. To know this requires that we rule out alternative explanations for any effects that are observed. In most situations, there are domestic and international factors that could influence the outcome or the short and long term impacts. Most USAID evaluation reports do not systematically consider alternative explanations. If they are mentioned, they are rarely carefully pursued to rule out as the primary determinants.

These four problems are ones that should be addressed in a new research plan for evaluation. For instance, excessive reliance on a single methodology is not a good idea since each methodology has weaknesses as well as strengths. The evaluation should provide basic facts on what was done in each activity and what the results were. Finally, the evaluations must consider alternative explanations even if it is not possible to fully rule them out.

RESEARCH DESIGN OF PREVIOUS EVALUATIONS

Further insight into previous USAID evaluations comes from describing them from a research design perspective. A classic and still influential perspective on design comes from the work of Campbell and Stanley (1963) and Cook and Campbell (1979). In their notational system, an X stands for a treatment and an O stands for an observation of the cases under study. In our work, the X represents an USAID activity intervention (or cluster of activities). The typical USAID evaluation report that we reviewed could be characterized as a "One Group Posttest Only Design." In the Campbell-Stanley-Cook notation, this is:

One-Group Posttest-Only Design		
Intervention Group:	X	O

This table shows that the USAID activity or "intervention" occurs first. Then either at the end, or midway through the intervention, observations are made on the presumed consequences of the activity.

Two major limitations of this design are: (1) there is not baseline information collected before the USAID activity and (2) there is no control group of nonrecipients. The former would enable evaluators to determine if change has occurred. Without a baseline observation, we must rely on impressions of changes from those involved in the intervention. As we explain in our full review of these evaluations, all parties involved in the evaluation have a stake in a positive assessment (Bollen, Paxton, and Morishima 2003). Based on our conversations with staff and on our readings of the evaluations, we believe that this has affected some reports.

Turning to the second limitation, a control group would help evaluators to see if similar changes occurred in the group *not* receiving the USAID intervention. Widespread changes might already be on the way in a country such that those receiving or not receiving the intervention would have changed *regardless of the USAID activity*. Observing a control group provides a check on whether the changes observed are uniquely tied to the USAID intervention. In addition, a control group could be helpful in determining the role of other alternative explanations in changes observed in the intervention group.

We are not arguing that having a baseline measure and a control group would solve all problems. If the design remains quasi-experimental or nonexperimental, there will still be threats to unambiguously attributing effects to a USAID activity. However, we are arguing that the typical research design of past USAID evaluations is one of the weakest. Further, it is possible to come up with alternative designs that will improve the ability to attribute effects to USAID activities.

One further problem of previous evaluations bears mention. As we have stated, in many previous evaluations, information is missing on the inputs as well as the outputs, outcomes, and impacts. This means that we do not fully know "X" or "O." That is, we do not know basic characteristics of the intervention or output (e.g., money spent or the start date of the USAID activity). Thus, one part of our suggested research design must be the collection of better information, a topic that we discuss in detail below.

To preview how our recommendations in this document diverge from the typical USAID evaluation of DG programs, we highlight six differences: First, we will recommend that multiple methodologies be used to evaluate any chosen sector, program, or activity. Second, we recommend more attention to evaluations of *impact* wherever possible, rather than immediate outputs. Third, we recommend that measures of inputs and impacts be explicitly operationalized and measured consistently over time and across countries. Fourth, we recommend that, wherever possible, control groups (individuals, organizations, or countries that *did not* receive USAID programming) be investigated as well as the groups that *did* receive USAID programming aid. Finally, we recommend that when evaluating any chosen sector, program, or activity, a thorough search for baseline data (that can be used to compare groups before and after USAID programming) be conducted.

OVERVIEW OF RECOMMENDED STEPS FOR EVALUATION

In this document, we present a number of steps to the successful evaluation of USAID DG programming. The rest of the document is ordered along these lines:

- Data collection on USAID sectors, programs, and activities
- Selection of activities, programs, and sectors for initial study
- Selection of multiple methods to evaluate the chosen sectors, programs, or activities
- Midstream assessment of initial evaluation approach
- Revisions/Additions to approach

The first step in the process of evaluation is one that can start immediately: improving central office information on the inputs of activities, programs, and sectors, as well as the outputs, outcomes, and impacts that result, in all the countries where USAID is active. Currently, very little information about USAID programming is available in the central office. Collecting basic information on USAID programming must be taken before activities, sectors, or programs can be chosen for evaluation.

We next suggest that USAID, in conjunction with a task force of outside experts, undertake the difficult task of choosing the activities, programs, or sectors for evaluation. As will be discussed below, USAID cannot evaluate every activity undertaken in every country. Therefore, decisions about what activities, sectors, or programs to investigate in more depth must be made. There are a series of trade-offs in this process that we outline below. The knowledge of USAID programming (e.g., cost, length of programming, number of missions undertaking) collected in the previous step will be essential to this process.

The third step is to decide on a mix of methods to evaluate the chosen activities, programs, or sectors. In this document, we describe six methods, provide examples of how they can be used to evaluate USAID activities, programs, or sectors, and describe their strengths and weaknesses. The specific mix of methods chosen will depend on the activities, programs, or sectors chosen for further investigation.

A midstream assessment is needed to correct any problems with the initial approach to the evaluation, and decide whether important activities, programs, or sectors were left out of the evaluation. This assessment will lead to revisions to the evaluation approach and either new methods or a different mix of methods to use in the transition to the final phase of the evaluation project. Also, there may be additions or deletions to the group of activities, programs, and sectors USAID is choosing to evaluate.

The specific sequencing of these steps that we recommend is noted at the end of the research design. Before the specifics, we discuss the major steps of the research design in general and in detail.

Finally, it should be noted that while USAID, through its SORA project, is particularly interested in a retrospective look at its programming, we include some suggestions for prospective evaluations as well. Prospective evaluations are the best way for USAID to gather concrete, rigorous information about the effectiveness of its programming. Taking a few steps in the present to collect better data on programming, and set up randomized experiments, will yield benefits for USAID far into the future.

COLLECTION OF BASIC DATA

The collection of basic data on USAID programming is a fundamental requirement of a high quality evaluation of USAID activities, programs, and sectors. The required data largely concern the inputs and outputs of USAID activities, so we focus on these in this section. Further discussion of the data for outcomes and impacts occur in the sections pertaining to each method.

A necessary first step to the implementation of this evaluation design is for each sector in the central DG office to conduct a census of activities in its area. In our discussions with USAID staff we have found that most view the activity as the basic unit. Within each sector (e.g., civil society) there is a collection of activities that fall under it. Yet there is not a list of *all* activities in which USAID engages for each sector. One exception is the Rule of Law area, where Lynn Carter, Mark Dietrich, and Molly Hageboeck created a census of activities for the E&E region that appears in appendix B. Having an analogous listing of activities for each sector would provide USAID a census of the types of activities, classified by sector. In some cases, an activity spills over into more than one sector, but this could be indicated in any such list of activities with cross-referencing.

Forming an activity census could be accomplished in a number of ways. Sector staff could have a meeting to brainstorm all activities, based on their expertise in the area. Alternatively, a sector could survey missions operating in their sector for the activities they are undertaking. Some sectors already have information that will be useful in conducting this census. For example, the governance sector has a database of activities in legislative strengthening covering 1991 to 2001.

Once the census is complete, the sector should organize the activities, again much as has already been done in the Rule of Law sector. The census and organization is an essential first step, because this design requires that a task force choose specific activities to be evaluated in more depth than others. Thus, it is essential for USAID and the task force to know *what activities missions are doing*. Organizing similar types of activities together will further aid USAID and the task force in *concretizing in a consistent manner the activities being undertaken in their* *sector*. Two existing organizational schemes that may aid in the organization are the USAID DG Results Framework appearing in Appendix C, and the DG Office Taxonomy of activities, appearing in Appendix D.

In addition to knowing what activities and programs are done, we need to know what goes into them. That is, what are the inputs to an activity? What are the inputs to a program? In an ideal situation, there should be information on the money, personnel, and material resources that were put into each activity or program. Categorizing the expenses by type would help; having the number of people and hours devoted would also aid the effort as would a description of the material resources and training or technical advice that was offered as part of the activity

We realize that this information may not be available in all cases. However, we see this as an ideal to strive for. *USAID should make every effort to collect data on the most disaggregated level that it can*. Part of these data must be quantitative, but a qualitative description would be valuable in providing information not tapped by the numbers, or simply not available.

Ideally, USAID would collect information on the following variables:

1. Money spent
2. Personnel (e.g., person hours)
3. # years
4. Other donors (#)
5. Other donors ($)
6. Implementing partner (US vs. local)
7. Geographic location (local (e.g., municipalities) vs. national)

To reiterate, for each of these variables, USAID should collect data for the most disaggregated unit possible. If information by sector is available, good, if it is available by program, better, and by activity – even better still.

For each of these variables, USAID should also collect data for as many years as possible. That is, USAID should collect information as far back in time as possible, while also collecting information for the most recent years.

Toward this end, the central office may want to consider allocating funds to each mission to hire additional support for a period of 3-6 months, so that the administrative personnel who know where the information is located could spend their time collecting it. For a limited initial investment, USAID could gain valuable retrospective information about its programming.

More aggregated information, sometimes on the sector, sometimes on a program, is more readily available than data on activities. The most frequent use of data on the sectoral level will likely be in the cross-national analyses we propose below. We defer discussion until that section.

Although SORA focused on outcomes and impacts, we think it would be worthwhile for USAID to collect data on immediate outputs as well. This would be most efficiently accomplished with the collection of inputs since the two are so closely related. For example, in the case of training, this would include how many and how long individuals were trained should be collected at the time that information on the activity is collected. We can think of outputs as corresponding to the major types of activities that USAID undertakes: training, technical assistance, the provision of material resources, and institutional development. In the case of training activities, USAID would want to collect information on how many instructors were used, how many people were trained, for how long they were trained, and how dispersed the training was (where was it done in the country). Technical assistance outputs include how many organizations or people were consulted with, and what their positions were. Institution building outputs include documenting the development of any new centers or institutions. In addition, information should be collected on how many, where they are located, how many employees they have, as well as early indication on whether they are being used. Finally, when material resources were donated, outputs include what was bought, and whether it was sufficient to meet demand.

Bruce Kay (USAID) prepared an initial report on the availability of data in the central office. We include this report as appendix J, along with some comments by Margaret Sarles (USAID).

Future USAID Data Collection

In addition to collecting basic information on the inputs and outputs of activities, we have several other recommendations for future USAID data collection. Although this section takes a prospective view, we **strongly urge** USAID to consider these recommendations. If USAID begins to collect better data on its programming now, it will bear fruit for years to come.

First, we recommend that evaluation be built *into* project proposals, rather than having a decision made about evaluation *after* the project is in the field or completed. An obvious reason for this suggestion is that having evaluation as part of the activity would enable the collection of important information prior to the implementation. In other words, baseline measurement would be possible. For virtually all of the research designs and methods that we discuss below, there would be great benefit from the collection of baseline data. Baseline data for the group receiving the activity and for a control group would be the most ideal situation. So a first step in any USAID activity would be to collect baseline information before the USAID activity is begun.

Another advantage of planning for an evaluation is that it might permit the use of more powerful research designs than currently used. A randomized experiment, for example, requires considerable planning for randomization of the intervention. If the evaluation is only considered after implementation, randomization is not possible.

Second, we recommend that all future evaluations contain a standardized core of information on inputs and outputs. Having a core, common template for future evaluations would ensure that (1) mission officers and the DG central office were receiving appropriate information about the successes and failures of various activities, and (2) the evaluations could be used for more systematic and comprehensive assessments of DG programming in the future. In making this recommendation, we recognize that there are specific issues related to each country and activity, and that these idiosyncrasies must certainly be captured in an evaluation. Therefore, it is important that the common template we suggest represent only the “core” of the evaluation, and that space be available to evaluators for other impressions and issues. Regardless, there are certain issues that are common across evaluations and USAID should make sure that these issues are addressed in each and every evaluation performed as part of the scope of work. In appendix E, we have an example of the standardized core of information that we recommend.

Third, we recommend that USAID develop a set of measures to be annually collected from the person in charge of DG activities in each mission. This should include an inventory of current DG activities and their inputs. It also could provide more subjective impressions of the status and trends in each of the DG sectors. Open ended questions on areas in the country that require the most work or where the greatest progress has occurred would also be helpful. We have included, as appendix F, a draft of such a questionnaire being developed by Eric Kite, Margaret Sarles, and other DG staff.

POTENTIAL METHODS FOR EVALUATION

Improving evaluation of democracy and governance programs requires a set of methods that address both the problems of previous evaluations and issues of USAID structure.

We suggest six methods, to be used alone, and in combination to evaluate the set of activities, programs, and sectors chosen by USAID and the task force. Table 1 lists the six methods and indicates the result (output, outcome, or impact) for which they are best designed. For example, cross-national longitudinal studies will be best for assessing very general levels of impact, while interviews and site visits are better designed to capture outputs and outcomes.

Table 1: Initial Methods of Evaluation			
	Output	Outcome	Impact
Randomized Experiments		√	√
Quasi-experiments		√	√
Surveys of Individuals & Groups	√	√	√
Interviewing and Site Visits	√	√	
Country Expert Sector Overviews			√
Cross-national (quantitative)			√

The set of methods we propose has a number of important features.

1. **Rigor**. As outlined above, traditional evaluations performed by USAID suffer from a number of problems. We attempt to strengthen the methodology used in USAID evaluations by adding more rigorous methods (such as randomized- and quasi-experimental designs) while strengthening the methods traditionally used by USAID (such as demanding systematic data collection from site visits).

2. **Flexibility**. The design is flexible. If one method does not seem appropriate for a particular research question or country, another is possible (It is important to note that we are not recommending that missions choose a method of their liking for evaluating one of their projects. Rather we see that a mix of methods will be selected by an agent outside of the mission. The method mix will be determined by the optimal, practical combination of methods for evaluation of a particular activity, program, or sector).
3. **Attention to multiple results (output, outcome, and impact**). The methods are differentially better suited to output, outcome, or impact analysis. We have weighted the designs more heavily toward impact analysis, to reflect SORA interests, but all types of results can be considered with this design.
4. **A combination of quantitative and qualitative approaches**. Combined, these two approaches provide a fuller picture of USAID programming.
5. **Triangulation**. It is essential to employ multiple methods, measures, researchers, and perspectives in an evaluation. Each method has its own strengths and weaknesses. By the proper choice of combinations of methods in investigating any given activity, program, or sector, we can compensate for the weaknesses of a single method.

In the next section, we describe what can and cannot be accomplished and the trade-offs involved in selecting different methods for evaluation, as well as the trade-offs in choosing activities, programs, or sectors for evaluation. Following that, we stress what can and *can not* be accomplished in an evaluation. Both sections lead to our recommendation that a task force be formed to use the data collected in the earlier stage to make recommendations about the activities, programs, or sectors to be evaluated in more depth. The task force would also recommend the optimal set of methods to evaluate each chosen activity, program or sector. We have six sections to describe each method in greater detail. Each section describes the method, provides a detailed example related to USAID programming, explains the advantages and disadvantages of the approach, and finally provides an estimate of its cost to implement.

TRADE-OFFS IN SELECTING DIFFERENT METHODS AND ACTIVITIES

In making a decision about the activities, programs, or sectors to be chosen for evaluation, USAID and the task force will face two major trade-offs. Explicit recognition of these trade-offs is necessary for USAID to choose where it wants to invest its evaluation resources, for what ultimate return.

Breadth vs. Depth

On the one hand, USAID could choose to evaluate all of its programming at a very abstract level – achieving *breadth*. That is, by focusing at the sectoral level (e.g., civil society, rule of law) USAID could evaluate its programming at a very general level, making statements such as “across countries more spending in elections and political processes is associated with higher levels of legislative effectiveness.” Several problems characterize this approach: at such a high level of abstraction it may be difficult to show an impact of USAID presence or funding, and, if found, USAID will not know *which* of its programs or activities led to the association. Furthermore, to assure that outcomes were due to USAID activities, it will be necessary to control for other factors that might explain the relationship.

On the other hand, USAID could choose to evaluate specific activities or programs in more *depth*. In this case, USAID could make statements such as "in Bolivia, the training of judges activity increased judges' knowledge and understanding of the law" or "in Indonesia, funding legal resource centers increased citizen access to justice." In this case, USAID would have very specific information on the relation between an activity and outputs and outcomes. However, only a small portion of USAID's total portfolio of activities could be evaluated at this level of detail. In addition, this approach would not necessarily rule out other explanations for the observed associations.

Retrospective vs. Prospective Evaluation

USAID, through its SORA project has two conflicting goals in the context of evaluation. On the one hand, USAID wants to attribute outcomes or impacts to USAID programming. In order to accomplish this, the most rigorous methodological design (namely randomized experiments) is needed. However, randomized experiments must be planned in advance, and therefore are *prospective* in their orientation. On the other hand, USAID wants a *retrospective* focus in its evaluation. Retrospective evaluation, using existing data to look for possible quasi-experiments, will be less able to make attribution to USAID programming. There is a wide range of threats to the conclusion that USAID produced a particular outcome or impact if only retrospective data are used.

WHAT CAN AND CANNOT BE DONE?

Ideally, we would like to be able to evaluate *all* USAID activities in the DG area and to have an approach that would *unambiguously attribute* the unique contributions of USAID to any outcomes that result. However, it is important to recognize what is and is not feasible to do.

What is feasible? We believe that it is quite feasible to improve on the existing typical DG approach to evaluations. Though more details are provided below, these improvements will come from using multiple methods customized to an activity, program, or sector, using baseline data and control groups where possible, and considering other explanations for observed effects. Considering future evaluation efforts, we believe that it will be possible to find activities that are suitable for randomized experiments. Randomized experiments will provide some of the clearest evidence of whether an USAID activity works or not. We also believe that the collection of new data in surveys or in other ways will enable the measurement of additional characteristics that could be part of an evaluation of USAID efforts.

What cannot be done? USAID cannot evaluate **all** of its DG activities. The first problem is simply the sheer volume of activities undertaken by USAID and its missions. There are literally hundreds of different activities occurring in 70 or more countries. The second problem is that each activity that is evaluated would require clear information on inputs (money, people, other resources) and outputs. These data are not readily available, and, in some cases will never be available.

It also is clear that for many USAID activities, programs, and sectors, there is a high degree of uncertainty in attributing effects to USAID. Each outcome or impact in which we are interested is influenced by other factors over which we have no direct control. Ethnic rivalries might play a dominant role in the peacefulness and fairness of an election. This factor might trump anything that USAID does. Or other donors might be working in the same sector as USAID and make it hard to distinguish the separate influences. When USAID activities were implemented in the past, research designs were not used that could take account of these confounding influences. The result is that for activities or programs undertaken in the past, we cannot unambiguously attribute effects to USAID activities. Some steps can be taken in this direction by using multiple methods and approaches that explicitly consider alternative explanations, but some uncertainty will always remain. Reducing this uncertainty involves investing more resources in prospective evaluations.

A TASK FORCE TO HELP USAID MAKE THE HARD DECISIONS

As indicated in the sections above, USAID has a number of difficult choices to make before it can evaluate any of its activities or programs in depth.[4] Further, these choices can not be made until basic data on activities and programs are collected. That is, without first knowing what activities programs are being undertaken by missions, USAID will not know the full range of activities and programs that *could* be evaluated. And without knowing the resources being invested in these activities (in terms of dollars, time, or number of missions), USAID will not be able to prioritize programs or activities along any reasonable set of criteria (e.g., “revealed preferences” in the form of dollars spent). For these reasons, we have recommended that the evaluation begin with a census of USAID activities, the collection of data on those activities, and an attempt on the part of USAID officials to group activities into similar types.

The next step of the evaluation would therefore involve the convocation of a task force in a series of three meetings. In the first meeting, the task force would take the information collected by USAID and make decisions on which activities and programs to be investigated in more detail. The goal of this meeting would be the selection of fifteen priority activities and programs to consider for evaluation. Once these activities are decided upon, the task force would convene a second time for the decision about what mix of methods to use in evaluating each chosen activity or program. The goal of this meeting would be to pick five to six activities that could be evaluated with a reasonable mix of methods. The winnowing of the 15 to 5-6 would be based primarily on what methods could conceivably be used for their evaluation, with primacy given to activities where it is possible to form control groups and have baseline measurement. The task force would meet a third time to recommend activities and programs to investigate with a randomized experiment. Again, although not officially part of the SORA project, we strongly recommend this prospective focus. We recommend that this task force be made up of the original technical advisory board as well as an equal number of USAID officials.

To further aid the task force in making its difficult decisions, we suggest that USAID undertake a series of focus groups, both of central office officials, and of mission officials, to discuss USAID evaluation priorities. Hard data on money spent and mission programming provide only one

---

[4] Evaluating sectors is more straightforward, as we discuss below.

source of information on USAID priorities – these focus groups should provide the harder-to-measure "fuzzy" priorities that have been important to USAID over time. The focus groups would be undertaken concurrently with USAID data collection, and would have a number of additional benefits. For example, focus groups could increase internal cohesion and agreement among the officials of the DG central office. In addition, they would help create learning for USAID officials.

To summarize: taking the census of activities, data collected on activities, programs, and sectors, as well as the results of the focus groups, the task force would have a first meeting and be charged with prioritizing a set of fifteen activities and programs to evaluate in more depth. In a second meeting, the task force would be charged with choosing the optimal mix of methods to evaluate a sub-set of six activities and programs, based on the availability of quasi-experimental designs, existing survey data, etc.. In a third meeting, sites for prospective randomized experiments would be identified.

In the sections that follow, we describe each method that could be used to evaluate USAID activities in greater detail.

Estimated Cost of Task Force Meetings

1. *Focus groups*, including several D&G personnel (Option A), w/ mission directors (Option B)
    a. One focus group, off-site in a dedicated facility, with one moderator and an assistant, taped and vidographed, $8,000-10,000
    b. Travel for 10 mission directors, $20,000
    c. Senior project analyst, 3 days per focus group session (3 x 2 x $500), $3,000
    d. Telephone interviews with 5 mission directors, 2 RD days, $1,000
    Option A: Two focus groups without mission directors attending, $20,000-24,000
    Option B: Two focus groups, with mission directors attending one, $39,000-43,000

2. *Task force meetings*, three one-day meetings of technical advisory board, RD, and AID staff
    a. Travel, honoraria, and per diems for advisory board, total, $36,700
    b. Senior analyst (3 x $500/day), $1,500

    Total, $58,200-62,200 (Option A)
    $77,200-81,200 (Option B)

RANDOMIZED EXPERIMENTS

Randomized experiments are the "gold" standard of research designs. The goal of an experimental design is to assess the change that has occurred in a target population while isolating the effects of other relevant factors. This is typically accomplished by *randomly selecting* the individuals or groups to receive the intervention. Applying the intervention at random is important because, on average, this will eliminate the association between the intervention and other variables that might influence the outcome or impact. Put another way, randomization provides a means to account for other factors that might influence the outcome or impact by making the intervention uncorrelated with these other factors.

There are many possible randomized experimental designs. A classic simple one is the Pretest-Posttest Control Group Design. Using the Campbell-Stanley-Cook notation introduced previously, we briefly present the essence of the design. Consider a USAID activity intervention, denoted in the following diagram by X. We observe (denoted by O) a group that received the intervention before and after the intervention, where the subscript indicates the time period. We observe the control group (who did not receive the intervention) at the same time points.

A Basic Randomized Experimental Design:

Intervention Group:	R	O_1	X	O_2
Control Group:	R	O_1		O_2

The R represents randomization of the intervention. In this particular design both the experimental and control group are measured before and after the application of the treatment. In this design, the control group captures changes that would have occurred without any intervention on USAID's part. Differences observed between the intervention group and the control group can therefore be attributed to the intervention.

Such a design has several advantages. One is that by having both before and after measures we can assess whether the activity intervention leads to a different amount of change than occurs in the control group. This means that if there is a general trend in the O variable then this will not be attributed to the intervention, since the same trend should occur in the control group. Or, if there are omitted variables or historical events that are influencing the trend in the intervention group, the same omitted variables are likely to influence the trend in the control group. By comparing the intervention and control groups we can rule these omitted factors out as a source of the difference between them. This leaves an effect of USAID programming as the explanation.

Implementation of a randomized experimental design requires four steps: First, a contractor/mission must randomly assign, *before* an intervention takes place, who will be exposed to it and who will not. Second, the contractor must design indicators to monitor both the intervention and control group before and after the intervention. This could be a survey in the case of training judges, an observation of minority voters at polling stations in the case of a voter registration drive, or a review of the local government tax base in the case of local government reform activities. Third, using the indicators, a contractor must monitor both the intervention and control group before and after the intervention. Finally, once the data is collected, the contractor must analyze it to determine whether a desired outcome or impact has occurred.

An Example: Legal Resource Centers

There are a variety of USAID activity areas where it might be possible to try randomized experiments. For instance, training programs are a common activity of USAID for legislators or judges. It would be possible to randomize who did and did not receive the training. However, applying the randomization to communities rather than to individuals within the same community is likely to be preferable because it lessens some problems. If we randomized across individuals within a community, it might be hard to apply the activity intervention to the experimental group and not the control group. For instance, if the training is perceived as valuable, members of the intervention group might share knowledge with members of the control group. This would be less likely to occur if the units being analyzed were communities.

For our example, we focus on the creation of legal resource centers. Falling in the rule of law area under access to justice programming, the development of legal resource centers can help traditionally disadvantaged groups, such as women and minorities, understand their legal rights and how to pursue those rights in the court system. Typically USAID might start a pilot project to introduce twenty legal resource centers across three hundred municipalities in some country.[5]

A randomized experiment measuring the efficacy of legal resource centers would begin by randomly selecting the communities or municipalities to receive the legal resource centers.[6] If a mission had the resources to develop twenty such centers in three hundred municipalities, then the twenty centers would be randomly assigned throughout the three hundred municipalities.

Before the centers began operating, the contractor would observe *both* the municipalities that were to receive the legal center, and those that were not (or a randomly selected group of those not to receive a center). Observation could take the form of a survey of the population (or just the disadvantaged population) about their rights and how to pursue them. Or, it could entail counting the number of women filing complaints with the court or contacting lawyers.

The contractor would observe the same municipalities again after the legal center had been operating for a suitable length of time (e.g., one year). There may have been changes in the society over the year that led to increases in the number of women filing complaints across the whole country. Such an increase would be evident in the municipalities that did not receive the USAID legal centers. However, if the legal resource centers were having their own effect, then the increase in the intervention municipalities should be even higher than those observed in the control municipalities. The contractor could use a variety of statistical methods to attribute an effect to USAID.[7]

---

[5] We thank Jose Garzon for providing information on typical USAID pilot programming in this area.

[6] Note that randomly selecting the municipalities to receive the resource centers could be viewed as the most fair way to distribute a valuable good.

[7] Other examples are possible. If voter education through a media campaign was the activity of interest, USAID could randomize which communities received the media campaign and which did not. Indicators of voter knowledge and voting behavior could be measured in each community before the activity commenced. After the voter education was completed, the two sets of communities could be measured again to see whether the education

Advantages and Disadvantages to Randomized Experiments

The main advantage to randomized experiments is that they are widely viewed as the "gold" standard of research designs. If done properly, USAID would have strong evidence of an effect of an activity, evidence that would be difficult for others to critique.

This design, like all designs, has its disadvantages. If cases drop out of either group, then this raises the possibility that observed differences are due to the nature of the cases that remain. It is possible that the intended intervention (X) might lead to changes in other variables that are the real causes of the observed differences. In addition, the degree to which the results of the experiment generalize beyond these two groups is always an open question. See Cook and Campbell (1979, pages 341-386) for further discussion of the challenges and opportunities of randomized experiments.

Other disadvantages include the need to plan ahead for a randomized design. Since USAID interventions are not typically assigned randomly, actions need to be taken before the commencement of an activity, including observation. Thus, a randomized experimental design could only be used to evaluate USAID activities in the future, not those that have already been undertaken.

Another disadvantage is that the contractor would need strong cooperation from a mission to undertake this research method. The mission would need to agree that randomization could occur, and allow the contractor to observe the intervention group and the control group both before and after the activity intervention.

Emphasis of method (sector, program, or activity level)

Randomized Experiments are best designed for activities and programs. Corresponding to this, the method is likely to provide the greatest insights into outputs and outcomes. It is possible that USAID might be able to observe sector-level effects with a randomized experiment, but most experiments would not target effects at that level.

Estimated Cost of Randomized Experimental Designs

Since USAID requested only retrospective methods, we do not include cost figures for randomized experiments, which are inherently prospective. We do, however, make the **very strong** suggestion that USAID consider setting up an incentive system, whereby missions choosing to undertake a randomized experiment to evaluate one of their activities or programs, would receive not only the money to pay for the experiment (including technical help on its design) but additional funds to spend on that activity. The additional money would help mitigate mission concerns that randomizing would jeopardize the possible success of their programming.

---

campaign had an effect such that the intervention group had different knowledge and behavior than the control group.

QUASI-EXPERIMENTAL DESIGNS

Sometimes randomization is not possible. In the case of USAID activities, the intervention may already have occurred by the time it must be evaluated. Or, there may be political concerns that prevent the random assignment of the activity intervention. Or, ethical considerations may mandate nonrandom assignment.

The key difference between randomized- and quasi-experimental designs is that quasi-experiments *do not use random assignment* to create the comparison groups. Instead, the control group and intervention group, while different due to the presence of the USAID activity, may also differ from one another along a number of other characteristics. Thus, the task of interpreting the result of a quasi-experimental design is more difficult – the effect of the intervention must be separated from any effects due to the initial noncomparability of the two groups.

Although there are a variety of quasi-experimental designs (see Stanley and Campbell 1963 or Cook and Campbell 1979), we again focus on the most straightforward as an illustration.

The Basic Quasi-Experimental Design

Intervention Group:	O_1	X	O_2
Control Group:	O_1		O_2

The key difference between this design and the randomized design discussed in the previous section is the lack of randomized assignment (R) to the intervention and control groups. In this design, the control group again captures changes that would have occurred without any intervention on USAID's part.[8] However, because of differences that may exist between the control and intervention groups, interpretation of the results is more difficult. See Campbell and Stanley (1963:40) for a concise representation of various quasi-experimental designs and the threats to validity they can or can not address.

Because random assignment is not a part of quasi-experimental designs, one can either design the evaluation ahead of time, or look for natural quasi-experiments that have already occurred.

[8] There are any number of other factors that may influence positive change in a control group over time. For example, all judges in a country may be internally motivated to reduce their backlog of cases. Thus, all judges may show some improvement in that outcome over time. Without a control group, it would be difficult to differentiate this internal motivation from any training the judges had received from USAID. See Cook and Campbell (1979:51-55) for a list of possible confounding factors.

An Example: Civic and Voter Education

Civic and voter education is an activity undertaken as part of the civil society sector. In addition, it is used in the elections and political parties sector as a part of election assistance.[9] Typically USAID would contract with an NGO or coalition of NGOs to provide civic or voter education to the population. Activities undertaken by the coalition might include broadcast and print media campaigns (e.g., radio spots, public service newspaper ads) to educate citizens about elections or participating in a democratic process. The NGOs might create community forums, such as town hall meetings, to get citizens excited about influencing politics. Members of the NGOs might physically "hit the street" with posters, flyers, and pamphlets informing citizens about democratic participation. In the case of voter education, large concerts (with popular rock bands) might be organized to "get out the vote."

In most cases, funding, or the distribution of NGO membership, would not allow for a nation-wide effort, so it is naturally targeted to particular areas of the country. Thus, it is possible that USAID and the task force could identify a country in which USAID operated in civic and voter education in only some localities.

In this case, the contractor would collect information before the USAID intervention on the outcome or impact variables of interest for areas of the country that experienced the education drive, and for those that did not. For example, if available, the contractor could collect information on local voter turnout, minority participation, or other outcome variables of interest to a civic/voter education campaign. In addition, the contractor could collect the same information after the USAID intervention.

Quasi-experimental designs are not free from threats to inference. In this situation, for example, there might be differences in levels of urbanization between the areas that received the education drive, and those that did not. If urban areas are more likely to increase their voter turnout over the time period of the study,[10] the quasi-experimental design described here could not separate that effect from a USAID intervention effect. Overall, this means that that analysis of the data can only make *plausible* attribution to USAID activities.

This example covers only one possible type of quasi-experimental design. Other types exist. One that may be particularly promising for USAID evaluation is the interrupted time-series design.

Advantages and Disadvantages to Quasi-Experimental Designs

The main advantage to a Quasi-Experimental Design is that it is methodologically more rigorous than the types of evaluations typically undertaken by USAID. Having baseline measures and a control group permits more careful evaluations than not having these. Also, it can take

[9] We thank Sundaa Bridgett for providing detailed information on civic education campaigns in the elections area to use in this example.
[10] An interaction of selection and maturation.

advantage of naturally existing situations where USAID has implemented an activity in only some regions, or with some individuals.

The disadvantage to this method is that we do not have randomization of the intervention. As we stated earlier, randomization provides greater insurance against other factors that may influence results.

Emphasis of method (sector, program, or activity level)

Quasi-Experimental designs are best designed to evaluate programs or activities (and possibly sectors). Like experiments, the level depends on the intervention and its hypothesized consequences. If it is narrow, then the activity level is most appropriate whereas broader intended consequences could permit program or even sector assessments.

Estimated Cost of a Quasi-Experimental Design

a. one senior analyst, half-time, 1.5 years (360 days x 0.5 x $500/day), $90,000
b. one research assistant (180 days x $150/day), $27,000
c. data, software, computer, etc., $20,000

Total, $137,000

SURVEYS

Surveys serve two purposes in our evaluation plan. First, existing surveys provide a source of information on individuals and groups who are relevant to USAID activities. Second, survey methods permit the collection of new information customized to the needs of the evaluation. As a source of data, surveys will be used in conjunction with some of the other methods that we discuss. For instance, in randomized experiments targeted at changing individual attitudes, a survey could provide information of the baseline and post-intervention attitudes in the experiment and control groups. Or in the case of multi-country surveys, country aggregated responses could be part of the cross-national analyses that we describe below. Thus the surveys are more a method of data collection than a set of analysis techniques.

Though surveys have broad applicability, we believe that they will be better at collecting some types of information than others. For example, we would prefer to have an administrative system to collect routine information on the inputs and possibly the outputs of USAID activities in each country. The routinization and consistency available from administrative records are likely to be superior to the same information attainable from surveys. Thus, surveys are likely to be most useful in the collection of information on outcomes and impacts, even though we do not rule them out for input and output information.[11]

[11] In fact, as we noted in the section on data collection, a survey of mission officers may be an excellent way to collect information on inputs and outputs.

As we mentioned above, we can broadly divide the survey data into data that are already collected and those to be collected as part of the evaluation. Though survey data collection is not as extensive in developing countries as they are in the United States, a surprising number of surveys are available. And, many of these survey datasets are fairly accessible. The ICPSR data archive in Ann Arbor, Michigan (www.icpsr.umich.edu), the Odum Institute for Research Social Science archive at UNC-Chapel Hill (www.odum.unc.edu), and the Roper Center at the University of Connecticut (http://www.ropercenter.uconn.edu/) are among the largest social science data archives in the world. They have surveys for some countries that receive USAID funding. For example, the Roper Center maintains a Latin American collection of surveys, some of which have variables related to USAID sectors. Mitch Seligson (University of Pittsburgh) has survey data for all of Central America, Ecuador, Bolivia, Peru, and Paraguay. These surveys contain a set of questions, repeated over time, that are relevant to areas such as civil society.

Despite the accumulation of survey data on many topics and in many countries, it is certain that not all of the data needed for the evaluation will be in archived data sets. In this situation a survey could be developed and administered before and after a USAID activity. The questions could be customized to tap the information most relevant to the goals of the activity.

Examples

To illustrate the use of survey data, suppose that USAID had an activity that was oriented toward increasing the general adult population's participation in nongovernmental voluntary organizations. If a survey in Bolivia asked respondents the number of voluntary organizations in which they were active, and, if this were repeated in surveys done before and after the USAID activity, these could be a rich source of information. Such data would be even more useful if the same question were asked of a neighboring country without the same USAID activity. In this situation, we could have a quasi-experimental design (see above) where the survey responses provide information on the outcomes and impacts using an intervention and control group for comparisons.

As a second example, suppose that a USAID activity is designed to educate the poor about their political rights and we do not have the appropriate questions for the particular country to help in our evaluation. In this situation, we could design a survey that collects data both prior to the USAID education activity and after in the section of the country receiving the intervention and another part that does not. A comparison of these two parts of the country and the before and after survey measures would be informative.

Advantages and Disadvantages of Surveys

The advantages of surveys depend on the quality of the probability sampling and the consistency of questions over time and across countries. If these meet rigorous standards, then the survey data can provide information that is not readily available from other sources. This is particularly true when the goal is to measure attitudes, beliefs, and some behaviors. If the data are already collected the cost of the information will be low.

On the other hand, if the conditions in a country do not permit rigorous probability sampling, then the survey measures will be less reliable. Also many variables might not be found in archived surveys or, if present, might be for only one time period. This will necessitate the collection of new data and new surveys can be costly.

Emphasis of method (sector, program, or activity level)

Existing surveys are best designed for sector- or program-level evaluation. Nearly all existing surveys would have questions that are broader than the activity level. However, a new survey could add questions specifically designed to capture the output or outcome of an activity.

Estimated Cost of a Retrospective Analysis Using Survey Data

a. one senior analyst, half-time, 1.5 years (360 days x 0.5 x $500/day), $90,000
b. 2 research assistants, half-time,1.5 years (180 x 2 x $150/day), $54,000
c. Travel, telephone, per diems, computers, etc., $30,000

Total, $174,000

FOCUS GROUPS, INTERVIEWS, AND EXTENDED SITE VISITS

Our next method is more qualitative in its approach. Because qualitative data collection is generally less constrained by predetermined analytical categories, using qualitative methods permits inquiry into selected activities in great depth and with careful attention to detail.

The methodology we propose is a combination of focus groups, interviews, and site visits. We see the methodology as occurring in a series of steps: First, a contractor would find a series of 8-10 missions in a region that undertook relatively similar activities in a sector within the past three years. Second, the contractor would organize a series of independent focus groups of: (1) the democracy officers in charge of the activity, (2) recipients of the activity, (3) representatives from the development partners who aided in the activity, and (4) a group of knowledgeable informants who have no stake in the results of the evaluation. Third, based on the information from these focus groups, the contractor should determine two countries to examine in more depth with extended site visits. The two countries should be picked on the following criteria: (1) a country where the activity was particularly successful, and (2) a country where the implementation and results of the activity seemed to be "typical" of the missions involved in the focus groups.

In conducting the focus groups, the moderator must have no stake in the results. And, the focus groups must be taped and transcribed so that the original data is available. It is also essential that the moderator focus on open-ended questions. Open ended questions help the moderator to "understand and capture the points of view of other people without predeterimining those point of view through prior selection of questionnaire categories" (Patton 2002:21).

We see two ways the focus groups could proceed. First, each group of participants (mission officers, recipients, etc.) could have their own focus group. Or, a single focus group could combine a mix. The latter strategy may allow more cross-pollination of ideas than the former, and may also be more cost-effective.

If the site visits can be identified on other critiera (for example, to provide more information on a location undergoing a quais-experimental design) the focus groups could be bypassed…

One goal of the focus groups would be to identify issues and questions that could be better addressed with site visits. In addition, problems that were common across missions, as well as problems that were unique, could be identified and slated for further investigation. As noted above, two countries should be picked for site visits based on the following criteria: (1) A country where the activity was particularly successful; (2) a country where the implementation and results of the activity seemed to be "typical" of the missions involved in the focus groups.

The site visit will progress in two waves. First, a single individual will arrive in country approximately three months before the site visit team is scheduled to arrive. This individual will collect advance data, request necessary documents from the government, and arrange the site visit. Second, the site visit team will arrive in country for two weeks.

The advance wave individual would have five basic tasks: a) the collection of basic information about USAID programming, resulting in a timeline of USAID activities; b) the collection of quantitative information on inputs (e.g., cost, personnel involved, date of inception, implementing organization); c) the collection of quantitative information on outputs (e.g., number of participants, NGOs affected); [12] d) the collection of information on alternative explanations for outputs and outcomes, such as other donors working in the field; e) the creation of a brief report for the incoming site visit, based on the data collected earlier, as well as documents retrieved from the DEC and R4/annual reviews; f) the organization of the team's visit (to reduce the burden on the mission).

We recommend two coding templates to collect the quantitative data. First, for every activity that is part of the evaluation, the individual should categorize the activity based on the census of activities we recommended above, e.g., the Rule of Law Census appearing in Appendix B. Second, we recommend that information on inputs and outputs be collected based on the coding sheet below. Overall, the goal of this coding sheet is to make sure that most relevant information about the activities is collected in a consistent manner across countries. The contractor needs to be able to "tell the activity's story": what happened, when, to whom, and with what consequences.

Who? (implementer)	Activity	Recipient	Spent	Personnel	When?	Other donors?	Outputs/Outcomes

[12] If our previous recommendation on the collection of USAID data was followed, this individual's main task would be to fill in information that had not already been collected.

We recommend a post-doctoral researcher for this advance wave.

In the next wave, a team of researchers will visit the country for two weeks to build on the information provided by the advance wave, to consolidate information, and to make judgments about the attribution of outcomes to USAID. During their visit, the advance wave individual will act as the local host, to ease the burden of hosting on the mission. During this visit, their tasks would include: a) reviewing what the advance person has already collected; b) reviewing relevant documents; c) interviewing relevant stakeholders, including mission officers, development partners, recipients, government officials, and other knowledgeable parties without a direct stake in outcome (e.g., academics, unrelated NGOs and donors, investigative journalists).

The site visit team will focus within its country on all activities undertaken in part of a sector, with a particular focus on two specific activities: (1) an activity commonly thought to be successful by mission officials, recipients, and third parties, and (2) an activity commonly thought to be unsuccessful by the same groups. This combination of breadth and depth will allow an overview of the area, as well as an understanding of how two particularly relevant activities performed.

We recommend that the site visit teams be composed of a US academic, a local academic, a former mission officer from the region, and one DG central office staff member.

At the end of the site visit, the team would create a written summary of their findings. This would include both the information collected in the first wave, as well as qualitative information from the second wave. While quantitative measures provide succinct, systematic, standardized information about what was going on in the program, qualitative data gives longer, more detailed and variable information about the program (Patton 2002:20). Methodology should be recorded, and filled-in coding templates attached as appendixes.

This methodology will address Evaluation Problems #2 and #3 – a lack of information on what was done, and what resulted. In addition, this research design attempts to deal with USAID Structure Problem #2 – that there are problems in working too closely with missions in evaluations. To briefly reiterate these problems, missions can deny access to countries, and strongly influence site visit teams during evaluations. The focus groups will help mitigate these problems in the implementation of this research by bringing mission officers out of their countries. In addition, having the advance wave individual organize the main site visit will help alleviate the burden that a site visit places on missions, while also removing some of the control over the visit from mission officials.

An Example: Strengthening Judicial Sector Institutions in LAC

To evaluate USAID efforts to strengthen judicial sector institutions in Latin America and the Caribbean, the contractor would begin by determining a set of 8-10 countries in Latin America that had attempted to strengthen the judicial sector between 2000 and 2003.[13]

The contractor would convene a focus group of mission officers (one from each country) who had worked in related activities (either the mission director, or the democracy and governance officer) during the time USAID was active. For example, a mission officer in Panama may have worked on the training of judges. These mission officers should not be *currently* working in that class of activities, which should help reduce their personal investment in the findings of the focus groups. In fact, it would be best if the mission officers had moved on from their LAC post to another post, to further reduce their distance from the activities they undertook.

For the focus group, the mission officers would be brought together in a single location, somewhere in Latin America or the Caribbean (or, potentially, in Washington, if the focus groups could be combined with one of their bi-annual visits to the US).
The focus group would concentrate on four broad sets of questions. The contractor and moderator should make every effort to ask open-ended, rather than "leading" questions of the participants. Table 1 contains examples of the types of questions that could be asked of focus group participants.

<u>Table 1: Example Questions for Focus Groups on Judicial Strengthening</u>

1. Questions about Activities
    - What activities were undertaken to strengthen the judicial sector?
        - Training of judges?
        - Improved information dissemination to judges?
        - Enhanced judicial independence?
        - Improving judicial accountability?
        - Creation of new courts?
        - Improving court administration?
2. Questions about Outputs
    - Specific questions about outputs
        - How many judges were trained?
        - How many new courts were created?
    - Implementation Questions:
        - Did you train as many judges as you originally expected to train?
        - Did the development partner use suitable materials?
        - Were there institutional barriers to implementing these activities?
        - Were other donors active, and did this impact USAID activities?
    - Was this a good investment, for the output?

[13] This group of activities was identified by the Rule of Law area as one in which the most money has been spent over the last ten years. In addition, most missions tend to program in this area, and it is viewed by the central officers as one of the most important for democracy.

3. Questions about Outcomes
    Are judicial decisions by trained judges consistent with their training?
    Did the judges continue to use what they'd learned 6 months later?
    Did the judges help train anyone else?
    Are judges abiding by ethical standards?
    Did the training help promote more effective and fair legal institutions?
    Is the judiciary objective?
    Did respect for the court increase among clients?
    Has court efficiency improved?
    Was this a good investment, for the outcome?
4. Questions about Larger Sector
    What judicial strengthening activities have been most effective in producing outcomes?
    What judicial strengthening activities have been most effective in producing impacts?
    What judicial strengthening activities have been *least* effective in producing outcomes or impacts?
    What external factors enhance or limit the effectiveness of judicial strengthening activities?

The questions about outputs would help USAID officials with basic input-output monitoring issues. Types of problems encountered by USAID mission officials in implementing activities related to judicial strengthening across a range of countries could be documented and distributed to other mission officials undertaking similar programming. The questions about outcomes would help USAID officials understand whether judicial strengthening activities were having an effect at a higher level. This information could be used to evaluate whether this class of activities should be encouraged or discouraged by the central office.

The contractor would similarly convene focus groups of (1) judges and court administrators, (2) lawyers, defendants, government officials, and other third-parties with little to no stake in the outcome of the evaluation, (3) development partners. In each focus group, the questions would need to be modified. However, the questions should generally cover the same basic topics.[14]

Once the countries for site visits had been identified (based on the criteria outlined above), the site visits would commence as described above.

Emphasis of method (sector, program, or activity level)

Site visits are usually best for evaluating the outcomes of specific activities or programs of activities. However, the more abstract nature of any sector-level impacts might be indirectly tapped through specific questions of interviewees.

[14] If focus groups seem impractical for the activity under consideration (e.g., it is unlikely that recipients would feel comfortable providing sensitive information in a group), interviews could be substituted.

Estimated Cost of a Site Visit

1. *First Wave*
    a. post-doc researcher (60 days x $200/day), $12,000
    b. local support for first-wave researcher (60 days x $50/day), $3,000
    c. travel, $2,000

2. *Second Wave*
    a. senior analyst (20 x $500), $10,000
    b. local academic (14 x $200/day), $2,800
    c. international travel (3 x $2,000), $6,000
    d. local travel and accommodations, $5,800

    Total, $41,600

COUNTRY EXPERT SECTOR OVERVIEW

More generalized changes occurring in a DG sector are not well tapped by experimental designs, quantitative measures, or qualitative interviews. Growing confidence in the political system or more tolerance for dissent, for instance, might be occurring but difficult to measure. Heightened interest in forming NGOs, or changes in the human rights situation might not always make their way to the available statistics. Even if all such changes were recorded, there still is a need to have a description of their meaning and to interpret the changes that they suggest.

One largely *qualitative* approach for assessing the impact of USAID activities is to have a country expert provide an overview of the major developments in a chosen sector (e.g., governance; elections) over a period of years that includes the USAID activity in that sector. In this method, USAID would commission an expert in the country (and sector) to spend approximately three months writing a report. As part of the commission, the expert would be allowed a brief visit (two to four weeks) to the country to collect any necessary information or documents, and to conduct interviews. We recommend a post-doctoral researcher who recently completed a dissertation related to democracy in that country.

The final report would cover major developments and trends in the sector and would have three components. First, the report would entail a general overview of the sector in that country, covering the major developments as the expert perceived them. Second, the report would include a detailed timeline of events related to that sector (e.g., constitutional changes, creation of new laws, implementation of new laws, major challenges to the sector, and changes in public perspectives). Third, the report would answer a set of standardized questions related to the sector.

Each component of the report is necessary. Once created, the timeline could be matched up with information about USAID activities in the sector. With the timeline, USAID officials could attempt to understand the sequence of events that occurred in the country, and whether USAID actions had an impact.[15] The common questions are important to make sure the expert addresses features of the sector of interest to USAID. In addition, with multiple reports, the standardized questions could be used to systematically compare reports across countries. Finally, the more open-ended description should capture unique events and general aspects of development in the sector that are not apparent from the timeline or standardized questions.

Having one expert do this overview has the danger that we are relying too heavily on a single person. It may not be practical or affordable to hire a second expert to do the same work. However, there are at least a couple of safeguards that would lessen this problem. One possibility is to hire a person to do a check on reports, books, and articles to ascertain the consistency of the expert overview with these other sources. Any discrepancies could be pursued with the expert to resolve the differences. Another possibility is to send the expert report out for review to other experts on the topic. These would be paid, anonymous reviews where the reviewers would raise questions about the overview that the expert would need to address.

These country expert sector reports could be used in a number of ways. If USAID wanted information about the performance of a sector in just one country, a single expert could provide a "spot check." Or, if USAID wanted more comprehensive information about the success of its work in a particular sector, it could commission a number of these expert reports, as well as a general synthesis. For the second approach, there would be additional benefits to commissioning reports in countries where USAID had not been working in the sector. These would provide a comparison group, analogous to the control groups described earlier. Finally, a report could be commissioned in any country where a more detailed evaluation in a sector will take place (e.g., in preparation for a randomized experiment). .

An Example: Electoral Assistance

Consider interest in USAID elections assistance activities [electoral administration, civic and voter education, electoral oversight (domestic and international monitoring), electoral framework improvement, participation of women and disenfranchised groups, and transfer of political power].[16]

The contractor could commission eight expert reports, based on the missions spending the most money in Elections and Political Processes. See Appendix G1 for information on budget obligations across missions in this sector. Appendixes G2-G4 contain information on obligations for the other sectors (Civil Society, Rule of Law, and "other Governance").

---

[15] Toward this end, it is essential that the country expert not be given information about USAID activities in the country. Further, the general description of the task and the questions the expert has to answer should not suggest an ultimate comparison with USAID activities. This is to lessen the contamination of the expert report with knowledge of the ultimate goals.

[16] Although this is technically only one class of activities under the sector of Elections and Political Parties, we focus on it because USAID has more control over activities in this sub-sector. We thank Michelle Schimpp and Sundaa Bridgett for helping to identify the types of activities undertaken in this sub-sector.

These eight country experts would each write a report of changes in elections and participation for their country over a 10 year period that included USAID programming in that sector. Their timeline would include any major political events that occurred such as regime changes, coups, protests, or electoral violence. The timeline would also cover changes to the electoral code, the presence of domestic or international monitoring, and any other major events related to elections in that country. The overall description could describe any vote fraud, vote buying, harassment, or other factors that affect the freeness and fairness of the elections, including the expert's impression of its intensity.

To make sure that standardized information was collected from each expert, they would be instructed to answer a series of questions. For example, they might answer questions such as:

1. Were there any new laws during this period that would affect electoral administration?
2. Did the national "climate" changed in a way that would affect voter turnout?
3. What is the demographic make-up of the most likely voters?
    3b. Did this change over the period?
4. Are there regional differences in turnout?
5. Was there evidence of vote fraud during the period (provide concrete examples)?
6. Has the national economy had an effect on voter participation over this period?
7. Describe the extent of domestic monitoring over the time period.
    7b. How did it change over time?
8. Did other international or domestic groups have activities geared toward affecting elections during this period?

These focused questions, developed in conjunction with the Elections Assistance team would make sure that the expert overviews provide consistently relevant information. Obviously, a much longer list of questions would be necessary. We provide these only as an example.

Once the reports were complete, they would be sent out for "review" to two other academics, investigative journalists, or USAID officers working in the sector. Once approved, they would be sent to a single academic who would synthesize the reports, in conjunction with information from USAID about its programming, and present a final synthesis report.

Advantages and Disadvantages to Country Expert Sector Overviews

As with all of the methods that we suggest, this expert overview has both strengths and weaknesses. Among its strengths is that the overview will be done by someone already familiar with the country. This should help to provide the context of any changes that are described. In addition, the overview should report on some characteristics that are not otherwise available. Furthermore the expert can provide the "big picture" by synthesizing the facts and giving them meaning.

In addition, this method does not rely on the missions for access or information. The expert would be separately commissioned and visit the country apart from the missions. Academic documents on changes in the sector, as well as independent document reviews and interview would form the basis of the report.

The approach has limitations as well. The country expert report might not provide enough detail to identify short-term effects of USAID activities. We already have mentioned that over reliance on an expert can be hazardous, though we have mentioned some safeguards to check the expert's observations. A related point is the difficulty in qualitative or quantitative analysis of knowing whether the associations described are due to real effects or are spurious. So an expert's assessment of whether an activity has an effect need not be right.

Emphasis of method (sector, program, or activity level)

Country Expert Sector Overviews are best designed for sector- or program- level evaluation. As the name suggests, these are meant to give sector overviews. However they could be modified to assess programs by asking very specific questions of the expert.

Estimated Cost of Country Expert Sector Overviews

a. postdoc (60 days x $200/day), $12,000
b. travel and accommodation, $4,000
c. senior analyst participation, 5 days, $2,500
d. peer review of report (3 reviewers x $500) $1,500

Total, $20,000 per country

CROSS-NATIONAL LONGITUDINAL DATA COLLECTION AND ANALYSIS

The last method we suggest involves the collection and analysis of a wide range of data on democracy, USAID programming, and alternative explanations. This method involves three steps: (1) compiling existing data into three master datasets, one containing democracy data, one containing USAID programming data, and one containing data on a wide range of alternative explanations; (2) supplementing the existing data with data on USAID programming; and (3) descriptive and multivariate analysis of the data that includes controls for confounding variables.

Data Compilation

Democracy Data
There are a wide range of existing datasets with information relevant to democracy and USAID sub-sector objectives. For example, besides the widely-known Freedom House country scores on democracy, there are additional measures of **democracy** (e.g., Bollen 1998; Vanhanen 2003), **autocracy** (Polity IV), and **dictatorship** (Alvarez, Cheibub, Limongi, Przeworski 2001).

More important, there are a wide range of existing indicators related to sub-sector objectives and strategies, e.g., **legislative strengthening** (legislative effectiveness, Banks 1979), **decentralization** (subnational revenues and expenditures, World Bank 2003), **electoral competition** (voter turnout), and **independence of the media** (the presence of laws and regulations that influence media content, Freedom House). Most of this data is available for multiple years (often multiple decades) and for most countries of the world.

For some countries, there is even more specific information available. For example, the *International IDEA* has information on 85 extremely precise indicators such as "How independent are the sub-central tiers of government from the center, and how far do they have the powers and resources to carry out their responsibilities?," "how effective is civilian control over the armed forces, and how free is political life from military involvement?," and "How inclusive and accessible for all citizens are registration and voting procedures, how independent are they of government and party control, and how free from intimidation and abuse?" These 85 indicators are available in Bangladesh, El Salvador, Italy, Kenya, Malawi, New Zealand, Peru, and South Korea in some recent years.

Similarly, *The Center for Public Integrity* is collecting data on 100 indicators of democracy and corruption for thirty countries in 2003, with plans to expand to 50 countries in 2004. Indicators include measures of **civil society advocacy and oversight** "in practice, do CSOs actively engage in public advocacy campaigns?," **election administration** "is there an effective election monitoring agency?," **legislative transparency** "does the public have access to legislative processes and documents?," **anti-corruption** "are judges safe when adjudicating corruption cases?," and **strengthening of police and law enforcement** "are law enforcement officials accountable for their actions?"

We suggest that a contractor compile all existing data on democracy into a single, searchable dataset. Search criteria should include country, sector objective, and strategy. The ability to reference indicators of overall democratization and sector objectives will help USAID DG central office officials measure the performance of USAID sector programming by making comparisons across countries and within the same country over time. The indicators can also provide baseline data for countries where USAID has been active in democracy programming, and where it has not been active. Reference to indicators can help central office officials find countries, potentially from different regions, with similar democracy and sector profiles.

An example of a dataset of democracy measures appears as Appendix H. Accompanying the compilation of such data should be a discussion of any information that is available on their reliability and accuracy. This should include the primary sources for the data and any limitations that are noted. Even the best national data will have faults so it is important to state these as part of the analysis.

USAID Programming Data

As we hope we have made clear, understanding the impact of USAID programming requires information on USAID programming. To measure the presence and magnitude of Democracy and Governance sub-sector objectives in countries, we have identified two information sources: (1) the budget/obligations data of USAID programs, and (2) The Results Review and Resource Request (the R4).

The first source of data, on budget obligations, provides information on dollars invested in democracy sectors by missions. The data covers approximately 110 countries from 1990 to 2001 (from 1994 to 1996 for E&E countries), and information is classified into five categories: civil society, political process, rule of law, other democracy, and other governance. Although these categories are not exactly the same as the four sectors of the DG Office, the budget/obligations data does provide the input information of three major areas of USAID's interest in the democracy program: civil society, rule of law, and political process.[17]

A second source of data is the R4 report, an annual document submitted by a Mission office to report its Strategic Objectives (SOs) and their implementation and performance over the past year. Our review of a sample of the R4 reports and the Annual Budget Submission Reports (prior to the mid-1990s) suggests that two pieces of key information, sector objective(s) and intermediate results, could be extracted.

First, the "Agency Strategic Framework" in recent R4 reports (listed as items 2.1 to 2.4) reports information on sector objectives. These cover the four DG sectors: rule of law and respect for human rights of women as well as men strengthened, credible and competitive political processes encouraged, the development of politically active civil society promoted, and more transparent and accountable government institutions encouraged. Thus, the sectors upon which a given Mission concentrates can be coded from these reports.

Second, the Intermediate Results (IR's), provide more specific information about the programs undertaken by the mission under a given sector objective. Overall, the IR's in most of the R4 reports seem to contain information at the programming level. For example, a mission might report that it is undertaking activities in "equal access to justice" – a sub-intermediate result in the results framework. In general, the R4s most closely follow the USAID DG results framework presented in Appendix C.

Appendix I provides more detailed information about the R4 reports, including a sample coding sheet. The collection of this data could obviously occur in conjunction with the USAID data collection effort suggested earlier.

Data on Alternative Explanations

Like data on democracy, there is existing data on a wide range of alternative explanations for democracy. For this part of the data collection, we suggest that the contractor scour the academic literature on democracy for typically hypothesized explanations, e.g., industrialization, literacy, religion, inequality, ethnic homogeneity, urbanization, region, etc. The contractor can then identify and compile the existing cross-national data on these explanations.

Some other alternative explanations of democratization may require new data collection such as, say, information on other donors.

[17] The budget data also contains information on the number of activities being undertaken in its four sectors. The usefulness of these figures is a source of some debate among USAID officials. Also, to check for reliability, this budget data should be checked again the Congressional Budget Justification.

Supplementation of Existing Data

Once the three master datasets are compiled, the contractor should determine the major variables for which data are missing. For the democracy data, this can be done through a comparison of existing data sources to USAID DG activities and programs. For USAID programming data, this can be done through interviews with USAID officials. For data on alternative explanations, this can be done by reviewing the academic literature on democracy and interviewing USAID officials about reasonable alternative explanations.

Once variables for which data are needed are identified, the contractor can code the data in one of two ways: coding qualitative accounts, or using expert informants to code information.

In the first case, qualitative accounts relevant to the variable of interest would be identified, a coding scheme coded, and variables created (see Hodson 1999 for a description of documentary analysis, and Paxton, Bollen, Lee, and Kim 2003 for an example of its application to suffrage). For example, if the contractor determined that quantitative data on free and fair elections was missing for recent years, they could use qualitative accounts such as Freedom House, electoral monitoring reports, etc. to code a quantitative indicator. In this method, a single coder or a small panel of coders would code all countries on the variables of interest.

In the second case, questions related to the variable of interest would be created, and these questions would then be posed to expert informants in each country, who would independently score their own country. To use the free and fair elections example again, the contractor could determine a set of questions about elections, and contact expert informants in each country to rate elections in their country. This methodology has been used successfully by *The Center for Public Integrity* to create a wide range of indicators of democracy and corruption. In this method, a range of expert informants would each code a single country on the variables of interest.

As discussed earlier, a variant on using expert informants could be very useful in providing new information about the strategies and activities undertaken by USAID mission officials. Here, we would suggest that the contractor, in conjunction with USAID officials, develop a survey on mission activities that can be sent to mission democracy and governance officials. This method has been used successfully in the USAID LAC bureau to gather information on mission democracy sector programming.

Analysis of Cross-National Data

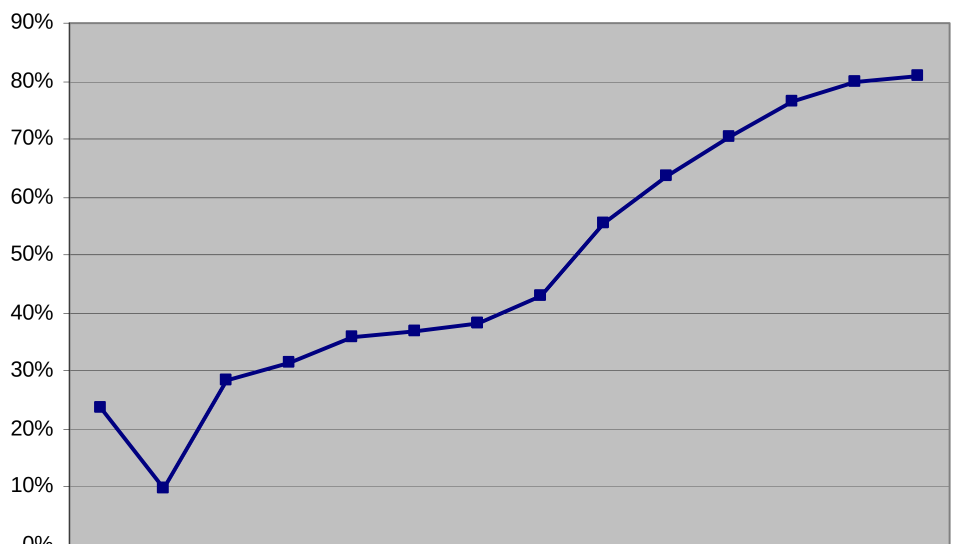
Once the data are collected and supplemented, the democracy data will be combined with information about USAID programming and alternative explanations in statistical analyses to demonstrate the impact of USAID. Several methods of analysis are appropriate. All such methods are best at assessing more general impacts of USAID activities rather than more proximate outcomes or outputs. This is true in that many of the measures (e.g., legislative effectiveness, performance of NGOs, etc.) are at the national level rather than looking at specific individuals or single groups. One broad way in which we can classify the methods is into *descriptive* and *multivariate* techniques. We explain and illustrate each of these approaches below.

Descriptive Methods

Descriptive methods are generally simple statistics or graphs that plot the values of one or two variables at a time. Though simple, they can be very powerful in communicating trends and simple relations between variables. For instance, suppose that a USAID set of activities is oriented toward increasing voter turnout in national elections. National voter turnout statistics are readily available for most countries so that a simple plot of the percent of the electorate voting over time could be graphed to determine if there is an upward trend. If there has been a change in suffrage over this period, the percent of adult population voting could be plotted as well.

To illustrate this, we take data from International IDEA on the percent of the adult population who have voted in national elections from 1945 to 1998. Figure 1 plots turnout over time in Brazil. The graph shows that over time, voter turnout has generally increased in Brazil.

Figure 1: The Percent of the Voting Age Population Who Have Voted in Parliamentary Elections, 1945-1998, Brazil (Source: International IDEA)



	1945	1947	1950	1954	1958	1962	1966	1970	1978	1982	1986	1990	1994	1998
Vote/VAP (%)	24%	10%	28%	32%	36%	37%	38%	43%	56%	64%	70%	77%	80%	81%

Another descriptive approach to these data is to compare the country of interest to other countries in the region or other parts of the world. For instance, the level of turnout in all developing countries might be analyzed to see where Brazil stands in relation to other countries at several points in time. Or a researcher might suspect that the trend in turnout observed in Brazil reflects regional trends. Plots of the percent of adult population voting over the same time period in neighboring LAC countries can reveal what part of the trend observed here is common to its neighbors and what part is unique to the particular country. More elaborate analyses that include control variables move us away from simple descriptive analysis and toward multivariate methods.

Multivariate Methods
Despite the value of the descriptive analysis, it is insufficient for systematic examination of the impact of USAID efforts. Associations between USAID activities and impact variables might be spurious or suppressed due to other factors that are related to both variables. For example, voter turnout might be reduced due to electoral violence in a country even though USAID activities made the turnout higher than it would have been in the absence of the USAID efforts. If we looked only at the bivariate relationship between USAID activities and voter turnout, we might underestimate the impact of USAID.

Multivariate techniques enable us to statistically control for confounding factors so that we can see USAID's impact net of other variables that influence the relationship. At the same time we must recognize that multivariate techniques are only as good as the correctness of the model that is used. With nonexperimental data, we must include the main variables that could confound the analysis of the relation between USAID activities and the impact variables. If we omit key factors, our analysis might lead to incorrect conclusions.

Single Country Analyses. There are a number of multivariate methods that could be applied to a single country. If over time measures for a single country are the focus, then regression methods would be appropriate.[18] Over time measures for an impact variable would be the dependent variable. Variables indicating USAID sector programming (e.g., dollars spent) would be the key explanatory variables. Additional explanatory variables would be included to capture the effects of possibly confounding factors.

An important decision at this point would be the best way to measure the USAID effort. For example, we could use either "stock" or "flow" versions of USAID dollars invested in a sector. This issue is whether the cumulative amount of resources (i.e., stock) matters more than the year to year amount (i.e., flow). In many situations, the choice is unclear and we might need to look at both types of measures.

A practical constraint on single country analyses is the number of years of data available. Suppose there are only eight national elections for which we have data on voter turnout. So few years of observations places severe limits on the number of other explanatory variables that we can consider.

Though it might not be possible to increase the number of years of turnout data, we might be able to gain useful information by forming a longitudinal data set that has more information on additional countries that can help to form a basis for comparisons. We turn now to this possibility.

---

[18] If data were available for 40 or more time periods, then we also could consider Autoregressive Moving Average (ARMA) methods of analyses. However, many fewer time-periods are likely for most of the democracy and governance variable that are of interest to USAID.

Longitudinal Cross-national Analyses. USAID evaluations sometimes look at more than one country over time. There are a variety of multivariate techniques that are suitable for analyzing such longitudinal data. We highlight two. First are techniques for estimating random and fixed effect models when pooling cross-sectional and time-series data. These techniques include a distinct intercept term for each country to capture its unique time-invariant characteristics that affect the dependent variable. In addition, it includes the effects of explanatory, time-varying variables that affect the dependent variable above and beyond the time-invariant intercept effect. These models capitalize on both the cross-national and the over-time variation in the data and hence contain more information than either the cross-sectional alone or the time-series data for a single country.

A second, less frequently used technique for longitudinal cross-national data are latent growth curve models. Latent growth curve models (henceforth, latent curve models) permit each country to have a distinct trajectory in the outcome or impact variable. Parameters are estimated that characterize these trajectories. Other variables for the country are entered in a model to predict these trajectory parameters.

Consider the voter turnout variable. In a linear latent curve model we have an intercept and slope to describe each country's trajectory in voter turnout. Countries can increase, decrease, or stay the same in voter turnout and their combination of intercept and slope will describe the trajectory.[19] We can enter the USAID dollars spent on Elections and Political Parties as determinants of the intercept or slope of these trajectories and include other variables that might confound the relationship. The result is that we can estimate whether USAID sector budget numbers influence the over-time trajectory of voter turnout as would be expected if the activities are effective. Variants of these latent curve models permit nonlinear trajectories, time-varying and time-invariant effects of variables on the repeated measures.

Advantages and Disadvantages to Cross-National Analyses

The main advantage to the cross-national method is that it can be used to gauge impact at a very high level of generality. Most of the indicators of democracy pertain to the sectors in which USAID works (e.g., civil society). Using this method can provide very general information on USAID efforts to promote democracy.

This advantage is also a disadvantage. Because the level of generality is so high, it may be difficult to "pick out" the impact of USAID programming from other effects. There are a wide range of determinants of democracy, only one of which is foreign aid. Thus, we may not be able to discern an effect of USAID, and if we do find one, its magnitude may be small.

Although information on USAID programming is generally not available, there is currently information on spending by sector (civil society, rule of law, governance, and elections and political processes). This sector level data could be used to begin cross-national analyses of USAID impact.

---

[19] As an example, returning to Figure 1, Brazil has an intercept of approximately 10%, a positive slope, and nonlinearity in the trajectory.

Regardless this method will provide the broadest possible analysis of USAID programming, across a wide range of countries and time periods.

Emphasis of method (sector, program, or activity level)

Cross-national analyses are best designed for sector-level analyses and analyses of overall democratization. It would be highly unusual to find variables that tap the activity level.

Estimated Cost of Cross-National Analyses

1. *Data compilation* - one year
    a. 2 senior analysts, half-time (240 x 2 x 0.5 x $500/day), $120,000
    b. 2 research assistants, full-time (240 x 2 x $150/day), $72,000
    c. Web site design, data entry, etc. $20,000

2. *Analysis* - two years
    a. 2 senior analysts, quarter-time (480 x 2 x 0.25 x $500/day), $120,000
    b. 2 research assistants (full-time), (480 x 2 x $150), $144,000
    c. Travel, data, computers, etc., $40,000

    Total, $516,000

Costs of Administration of Whole Project, Task Force through Multiple Methods

a. project director ($120,000/yr x 3 years)	$360,000
b. project assistant ($40,000/yr x 3 years)	$120,000
c. Travel, office, equipment, comm., etc.	$ 90,000

Total, $570,000 (note: no cost figures include institutional overhead)

EXAMPLES OF IMPLEMENTATION

The ultimate implementation of the research design depends on the decisions made by USAID and the task force on the programs and activities to prioritize for evaluation. In this section, we give some examples of possible implementation strategies for a few programs that USAID officers have identified as particularly important.

Civil Society Advocacy

Within the civil society sector, the central DG office identified civil society advocacy as a possible priority area. Civil society advocacy is viewed as theoretically essential for democratization. Civil society advocacy has the additional benefit that it crosses a number of sectors, for example, rule of law, and therefore is not constrained to a single sector. If USAID were interested in evaluating civil society advocacy, we recommend the following mix of methods:

1. *Cross-national analyses* using the NGO sustainability index. The E&E bureau of USAID has created the NGO sustainability index, made up of seven separately-coded variables: legal environment, organizational capacity, financial viability, advocacy, service provision, infrastructure, and public image. Separately, these variables provide information closely related to a range of civil society programming. Of particular interest in this case is advocacy. The index is available for six years, 1997-2002 and covers 17 countries in the first year, increasing to 28 in 2002. A pooled, cross-sectional analysis of these countries and years would allow an assessment of whether the amount of USAID spending in the sector of civil society has an impact on countries' advocacy scores.[20]
2. *Country sector expert overviews* of civil society in Eastern and Central Europe. We suggest that expert overviews be commissioned in eight countries where USAID was active in civic advocacy, and five where USAID was not active in that sector. The country sector experts could describe the situation in the country from 1990 on, which would cover the potential period of USAID activity. These overviews would provide more in-depth information on countries analyzed in the cross-national component, and allow an investigation of whether the presence of USAID aid (vs. its absence) made a difference to civil society in the region.
3. *Interviewing and site visits* to the countries deemed most successful from the cross-national analyses and the country sector expert overviews. The goals of these site visits would be to understand *why* USAID had an effect on civil society. Interviews with mission officers and implementers would help USAID understand which activities within the program of civil society advocacy had been the most successful in creating outcomes and impacts. For example, did mission officers feel that media campaigns had been the most successful types of activities, or were training activities viewed as more successful? The site visits would provide much more detailed information on the activities undertaken in the successful countries, as well as provide more information on the unanticipated consequences of activities.

Voter Education

Through discussions with Sundaa Bridgett and Michelle Schmipp, we identified voter education as an area within election assistance where USAID has been active for many years, and where missions tend to program the most. If USAID chose to evaluate this activity, we would recommend the following combination of methods:

1. *Cross-national analysis* of USAID programming in elections (i.e., money spent in the area of elections) and voter turnout. The International IDEA has voter turnout data for approximately 170 countries beginning in 1945. The growth curve models described above could be used to investigate the impact of USAID funding on voter turnout.

[20] The amount of funding is necessary, rather than simply program presence, since all but one of the countries in the sample received USAID civil society sector funding. This suggestion also assumes that information on USAID spending would only be available at the sector level. If more detailed information was available (e.g., at the advocacy programming level), it could be used instead.

2. *A quasi-experimental design* to investigate voter knowledge and attitudes is possible using the World Values Survey. The World Values Survey asked citizens a variety of political questions in 1990, and again in 1995. If USAID and the task force could identify one or more countries that had received USAID funding in elections between the two waves of the survey, and one or more countries that had not received USAID funding, a quasi-experimental design could be used to evaluate USAID impact. Countries with surveys available in each time period include: Mexico, South Africa, Argentina, South Korea, Poland, Brazil, Nigeria, Chile, Belarus, India, Slovenia, Bulgaria, China, Turkey, Lithuania, Latvia, Estonia, and Russia. The World Values Survey asks a variety of relevant questions, including:
    The importance of politics to the respondent,
    The respondent's interest in politics,
    How often a respondent discusses political matters with friends,
    Participation in unconventional political activities,
    Whether an individual has developed political opinions on a range of issues,
    Confidence in public institutions, such as parliament and the police,
    Civic responsibility, such as viewing cheating on taxes as wrong,
    Political efficacy,
    Whether the individual can name a party that they would vote for, and
    Tolerance.
3. *A quasi-experimental design* to investigate voter turnout may also be possible. If a country can be identified in which (a) local voter turnout information is available, and (b) the mission or implementer has a record of where USAID did get-out-the-vote drives, a quasi-experiment would follow.
4. *Site visits* to the countries investigated in the quasi-experiments to gain insight into *which* activities were most successful in producing results, and *why*.

MIDSTREAM ASSESSMENT

The preceding material describes our recommended initial multi-method approach to USAID evaluations. Given the number of methods, the number of activities and countries to evaluate, and the complexity of the situation, it is inevitable that the evaluation identified by the task force will require modification. This is why we propose a midterm assessment as part of the evaluation plan. This also is consistent with the aim of making USAID a learning organization.

We recommend that the first 2.5 years of the evaluation be devoted to implementing the initial mix of methods. This will give sufficient time to gain enough experience with the approach to understand its limitations and merits. We recommend a half year period for the midstream assessment. Its purpose will be to determine how well the initial plan is working and to propose revisions to the methodology.

The midterm assessment will build on annual reviews of the evaluation activities. Each annual review will include a description of the activities, countries, and years covered, a budget update including money spent and personnel involved as well as the methodology used and a summary of the primary findings. The midterm assessment will include the same information up through the first two and half years of the project. But in addition it will assess each of the methodologies employed as well as the mix of methodologies and how well they performed in this initial period. The midstream assessment also will consider whether the coverage of activities and regions is adequate or whether a different mix makes sense.

A primary result of the midstream assessment should be a set of specific recommendations on how to modify the current mix of methods, activities, and programs.

The contractor and USAID will determine the personnel most involved in the midstream assessment. We recommend that the advisory board/task force be available to review the report that is developed.

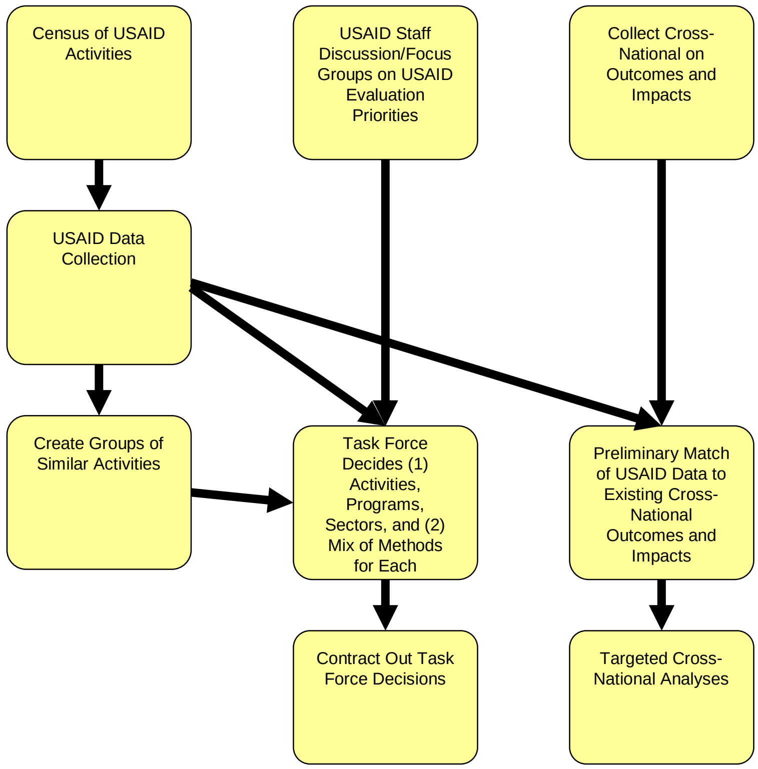
REVISED MULTI-METHOD APPROACH

The last phase of the evaluation plan is to modify the initial evaluation approach in response to the recommendations from the midstream assessment. These modifications are likely to lead to changes in budgets and personnel to enable them to be implemented. For this reason it would make sense to fund this revised approach as a separate activity from the initial evaluation and midstream assessment.

NEXT STEPS/SUGGESTED TIMELIME

We recommend a number of steps in implementation of the evaluation. These are represented schematically below as a flow chart. In the flow chart, steps that need to be taken logically prior to others are represented at the base of an arrow, while steps that follow logically from these appear at the head of the arrow. Steps on the same horizontal plane can be undertaken simultaneously. Each step is discussed in more detail below.

Figure 2: Flow Chart of Initial Evaluation Tasks



Most information on these steps has already been provided throughout this research design. Below, we briefly summarize each step in order, where the step number represents order of implementation. If two steps have the same number, then they can be done simultaneously and are distinguished by a letter. We also add a final, prospective, step that does not appear in the figure above.

Step 1a. Create a census of USAID activities. This should be a list of all USAID activities undertaken in all regions and sectors.

Step 1b: Undertake series of focus groups exploring USAID evaluation priorities.

Step 1c. Begin collection of longitudinal cross-national data on outcomes and alternative explanations. Add country code that will enable merging of variables from different data sets. Also keep record of sources and known problems of data. Devise system for naming variables. Make sure that informative labels are given to variable, including the year to which the variable refers.

Step 2. USAID data collection on activity, program, and sector inputs (money, person hours, presence of other donors). These data should be reported in as disaggregated a form as possible.

Step 3a. Form groups of activities according to similarity (e.g., strengthening judicial sector institutions vs. legal education reform).

Step 3b. Convene task force first meeting to select activities, programs, and sectors for evaluations. Consider money spent, person hours, material resources, theoretical importance, and representation of different sectors. The task force should include roughly half USAID staff. Convene task force sector second meeting to select the best mix of methods for each activity, program, and sector chosen.

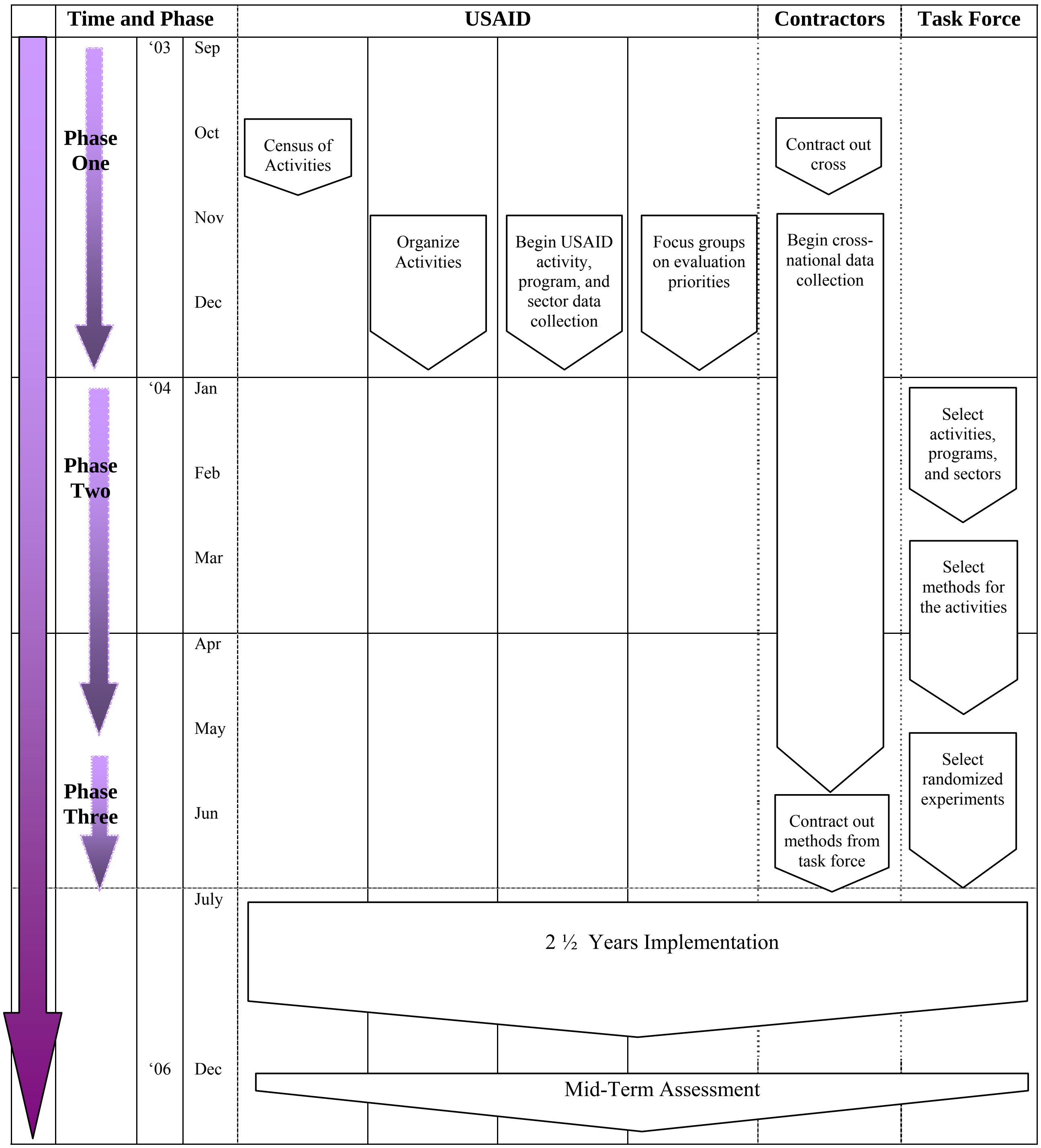
Step 3c. Make preliminary matches of cross-national variables (collected in step 1b) to input variables (collected in step 2).

Step 4a. Contract out the evaluations suggested by the task force.

Step 4b. Begin targeted cross-national analyses, based on the recommendations of the task force.

Step 5. Convene task force third meeting to select activities or cluster of activities for randomized experiments. An essential criteria for selection is the location of cooperative missions that would permit randomized experiments.

Below we include a suggested timeline for implementation of this research design, demonstrating which steps must be undertaken first, and which can be undertaken simultaneously with others.



AN ALTERNATIVE: INVESTIGATING ONLY AT THE SECTOR LEVEL

USAID staff, after seeing an earlier draft of this research design, expressed concern about undertaking a large scale USAID data collection effort. They are unsure of the amount of consistent information that could be gathered about activities and programs across countries, and therefore whether they and the task force can make informed recommendations about what activities and programs to evaluate in more depth. Given that sectoral-level data on spending already exists, they asked for explicit recommendations about how to evaluate USAID programming at the sector level (rule of law, civil society, elections and political processes, and governance). In this section we provide a recommendation for a sector level analysis. This part of the evaluation could be undertaken immediately, without the collection of new data on USAID programming, and without the need for a task force.

In making our recommendations in this section, we want to stress that we suggest that USAID pursue the full evaluation design outlined in this whole document. As outlined in the section on trade-offs, there are a number of disadvantages to focusing only at the sector, not the least of which is that it may be difficult to find effects of USAID programming at such a high level of abstraction. Thus, while we make this suggestion for immediate evaluation at the sector level, we stress that our preference is for the whole package as outlined in this document.

To evaluate its sectoral programming, we recommend that USAID undertake a combination of (1) cross-national, longitudinal analyses, and (2) country expert sector overviews.

After data on outcomes and impacts are compiled, as well as data on alternative explanations, the cross-national, longitudinal analyses could begin with the outcome variables in Appendix H. This appendix lists a selected subset of currently available cross-national measures of outcomes and impacts related to USAID.

Under the area of civil society, we see the NGO sustainability index as a good source of information on outcomes relevant to civil society programming. This index is only available for Central and Eastern European countries, however. Use of this measure is discussed in more detail in the section (below) on Examples of Implementation. Humana's measure of Assembly and Association would provide information on civil society outcomes for more countries in a few years. Similarly, his measure of the freedom to monitor human rights violations would provide some evidence of the success of civil society advocacy programming.

Rule of Law sectoral programming could be evaluated with at least two existing indicators. First, the World Bank governance indicators have a rule of law scale available every other year from 1996 to 2002. In addition, Humana has two variables that may be particularly useful: the number of extrajudicial killings, and the presence of torture by the state.

In the area of Elections and Political Processes, we recommend that cross-national analyses begin with two variables by Banks, party legitimacy, and the competitiveness of the nomination procedure. Voter turnout is another existing cross-national measure that could be used to evaluate USAID programming in this sector.

Governance programming covers decentralization, legislative strengthening, and anti-corruption. Measures of outcomes related to these programs are available cross-nationally. For example, Banks has a measure of legislative effectiveness, while the World Bank has a series of fiscal decentralization indicators, only some of which are included in Appendix H.

If USAID is interested in its impact on overall democracy, there are a number of existing indictors that could be used. First, Bollen has created a measure of democracy that combines some existing measures to avoid some known problems with bias. His measure is available from 1972 through 1988 and has been updated through 1995 by Paxton. Monty Marshall, Keith Jaggers, and Ted Robert Gurr in the Polity dataset have another measure of democracy available from 1800-2001.

We stress that these are only a small subset of the possible cross-national outcomes and impacts that could be used to evaluate USAID sector programming. There is a wide range of other variables that could be used as measures of impact. We suggest these only as a place to begin.

We also stress that the data currently available will only allow an examination of total USAID spending by sector (e.g., total spending in civil society). If USAID wants to examine more precise links between USAID programming and outcomes (e.g., the relationship between spending on civil society advocacy and a civil society advocacy outcome), new retrospective data on USAID programming must be collected. To take example: with existing data, it is not possible to distinguish mission spending on decentralization, legislative strengthening, and anti-corruption, within the sector of governance. Thus, any outcome related to legislative strengthening, such as Banks legislative effectiveness variable, could only be predicted by total USAID spending in the sector, *not* spending on legislative strengthening. Again, if USAID wants the latter investigation, new data on spending must be collected.

The country expert sector overviews could either occur simultaneously with the cross-national analyses, or after those analyses are completed. In the former case, we recommend that for each sector, eight countries be evaluated (in the two regions where most money has been spent, the top two country recipients for the 1990-1995 period as well as the top two country recipients for the 1996-2001 period).[21] This would result in thirty-two country sector expert overviews. In the latter case, we recommend that country expert sector overviews be commissioned to investigate surprising findings from the cross-national analyses, as well as regions and sectors that are determined to need further investigation after the cross-national analyses are complete. In either case, a few control" countries (where USAID had not been programming in a sector) would need to be investigated as well. This would allow an assessment of whether USAID's aid made a difference to the sector, across countries, not just across time. We recommend two control countries in each sector, for a total of eight control countries.

[21] For example, according to appendix G2, the biggest regional recipient of civil society funding over time has been Europe and Eurasia. Thus, we would recommend that country overviews be done for Serbia and Russia (the top recipients for 1996-2001) as well as Croatia and Ukraine (the top recipients in 1990-1995 excluding Russia because it is already included in the sample). In addition, we would recommend that in Asia and the Near East (the second largest recipient of USAID funding) overviews be done for Indonesia and Egypt (top country recipients in 1996-2001) as well as Cambodia and the Philippines (top recipients in the 1990-1995 period).

These country sector overviews would require two additional items. First, a single academic would need to synthesize the incoming reports for each sector. This academic would take the eight country studies, as well as the two controls, and synthesize the material to determine evidence for USAID impact. The second item is intricately related to the first. The academic will need information on USAID programming over the time period that the country experts are covering.[22] (This alternative design assumes that USAID has not already collected this information) For this, we suggest that all countries being evaluated with overviews receive some additional administrative support from the central office so they can give one of their administrators (who knows where things are) the task of collecting information on what USAID was doing in that sector in that year. This information would be consolidated and used by the synthesizing academic to determine USAID impact.

Cost of Alternative Plan

Because we are making specific recommendations related to sectoral evaluation in this section, we can give a more precise estimate of the cost of this part of the evaluation effort.

Cross-National Portion

1. Data compilation - one year
    - a. 2 senior analysts, half-time (240 x 2 x 0.5 x $500/day), $120,000
    - b. 2 research assistants, full-time (240 x 2 x $150/day), $72,000
    - c. Web site design, data entry, etc. $20,000

2. Analysis - two years
    - a. 2 senior analysts, quarter-time (480 x 2 x 0.25 x $500/day), $120,000
    - b. 2 research assistants (full-time), (480 x 2 x $150), $144,000
    - c. Travel, data, computers, etc., $40,000

    Total, $516,000

Country Expert Sector Overview Portion

- a. 40 country expert reports (40 x $17,500), $700,000
- b. senior analyst (80 x $500/day), $40,000
- c. administrative support in missions (60 x $30/day x 40), $72,000

Total $812,000

Grand Total: $1,328,000

[22] Remember that the country experts should be writing their report in the *absence* of information on USAID programming.

WORKS CITED

Banks, Arthur S. 1971. *Cross-Polity Time-Series Data*. Cambridge, Mass.: MIT Press. See also http://www.databanks.sitehosting.net/.

Bollen, Kenneth A. 1998. *Cross-National Indicators of Liberal Democracy, 1950 to 1980* [codebook]. University of North Carolina.

Bollen, Kenneth A., Pamela Paxton, and Rumi Morishima. 2003. "Review of USAID Evaluations on Democracy and Governance."

Carothers, Thomas. 1999. *Aiding Democracy Abroad: The Learning Curve.* Washington, D.C.: Carnegie Endowment for International Peace.

Campbell, Donald T. and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research.* Boston, Mass: Houghton Mifflin Company.

Center for Public Integrity. 2003. "Global Access." http://www.global-access.org/ga/default.aspx

Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings.* Boston, Mass.: Houghton Mifflin Company.

Denzin, Norman K. 1978. *The Research Act: A Theoretical Introduction to Sociological Methods.* 2nd ed. New York: McGraw-Hill.

Freedom House. 2003. "Press Freedom Survey." Available from http://www.freedomhouse.org/research/pressurvey.htm

Hodson, Randy. 1999. *Analyzing Documentary Accounts*. Thourand Oaks, Ca.: Sage Publications.

International Institute for Democracy and Electoral Assistance (International IDEA). 2003. "Voter Turnout from 1945 to Date." Available form http://www.idea.int/vt/index.cfm

Joint Committee on Standards for Educational Evaluation. 1981. *Standards for Evaluations of Educational Programs, Projects, and Materials.* New York: McGraw-Hill.

Marshall, Monty G. and Keith Jaggers. 2003. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2001." Available from http://www.cidcm.umd.edu/inscr/polity/

Patton, Michael Quinn. 2002. *Qualitative Research and Evaluation Methods*. 3rd ed. Thousand Oaks, Ca: Sage Publications, Inc.

Paxton, Pamela, Kenneth A. Bollen, Deborah M. Lee, and Hyojoung Kim. 2003. "A Half-

Century of Suffrage: New Data and a Comparative Analysis," *Studies in Comparative International Development* 38: 93-122.

Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub, and Fernando Limongi. 2000. Democracy and Development: Political Institutions and Well-Being in the World, 1950-1990. Cambridge: Cambridge University Press. (http://www.ssc.upenn.edu/~cheibub/data/Default.htm)

The United States Agency for International Development (USAID). 2003. "The Commodities Reference Guide." Available from http://www.usaid.gov/hum_response/crg/

Valadez, Joseph and Michael Bamberger. 1994. Monitoring and Evaluating Social Programs in Developing Countries: A Handbook for Policymakers, Managers, and Researchers. Washington, D.C.: Economic Development Institute of The World Bank.

Vanhanen, Tatu. 2000. "The Polyarchy Dataset: Vanhanen's Index of Democracy." Available from http://www.sv.ntnu.no/iss/data/vanhanen/

World Bank. 2003. "Data on Decentralization." Available from http://www1.worldbank.org/publicsector/decentralization/dataondecen.htm

APPENDIXES

APPENDIX A: GLOSSARY OF TERMS

A *strategy* is defined as the overall efforts of USAID in a region or country. A country may have a number of strategic objectives as part of its strategy.

Each strategy consists of several *agency goals*. The agency goals correspond to major areas in which USAID operates such as the environment, population, or democracy and governance.

A *strategic objective* is an abstract statement of a mission's objective in a particular area. Most often a strategic objective corresponds to an agency goal, such as "strengthening democracy and good governance" or "protecting the world's environment of long-term sustainability." Thus, looking at strategic objectives can identify certain missions as operating or not operating in democracy.

Each agency goal consists of several *agency objectives* that fall under it. In the case of DG, the agency's objectives refer to one or more of: rule of law (2.1), political processes (2.2), civil society (2.3), and governance (2.4). Thus, looking at what agency objectives a mission is pursuing lets us know in which democracy sector they are working.

An *intermediate result* is a component of a strategic objective. These can be viewed as strategies to attain a strategic objective which in turn promotes the agency objective. For example, to promote "more genuine and competitive political processes" (agency objective 2.2), missions may work on the "inclusion of women and disadvantaged groups" (intermediate result 2.2.6).

Activities are specific development interventions undertaken by USAID or one of its development partners. These are the most basic level at which USAID works. For example, a mission may work with NGOs to increase awareness of their rights and responsibilities regarding women. Or, a mission may train voter registrars in techniques for increasing the number of registered ethnic minority voters.

Activities can be classified in a number of ways. USAID uses *sub-intermediate results*, which are analytical classes of activities, to classify activities. The activities above can be classified into these sub-intermediate results. For example, training voter registrars could be classified under sub-IR "increased participation by women and disadvantaged on elections day." The Office of Democracy and Governance has also worked on a classification scheme for activities which is similar but not identical to the general USAID classification. The USAID classification appears in appendix A, while the DG classification appears in appendix B.

Inputs are the resources (e.g., dollars, personnel, equipment) needed to implement a USAID activity.

Implementation refers to "the transformation of activity inputs, through a set of technical and organizational systems and procedures that produce a specified volume and quality of activity outputs." (Valadez and Bamberger 1994)

Outputs are the direct products of an activity and are typically communicated in terms of quantity. For example, the number of legislators trained or the number of conferences held. USAID defines an output in part as, "A tangible, immediate, and intended product or consequence of an activity within an organization's control or manageable interest." (USAID Commodities Reference Guide 2003)

Outcomes and Impacts: are the benefits of an activity on a group of participants or a target population. "Impact" and "outcome" tend to be used interchangeably in the literature. In this document we distinguish the outcomes and impacts based whether they correspond to short- or long-term results.

Outcomes are short-term results of activities, partially generated by the activity's outputs. For example, an output of a judicial activity might be the number of judges who attended a training session; an outcome of the same activity might be reduced backlogs in the courts of the newly-trained participants.

The next, and more difficult step is to assess the overall or longer-term *impacts* of activities. Impact evaluation focuses on this task and therefore faces all the challenges of assessing and attributing change in overall democratization.

Outcomes and impacts are measured by identifying characteristics that tell whether change has occurred. These measures are called *indicators* and they are concrete, observable phenomena that indicate that a specific change has taken place.

APPENDIX B: RULE OF LAW ACTIVITY CENSUS

Rule of Law Assessment Activity Focus Topics

IMPROVING THE LEGISLATIVE FRAMEWORK

A. CONSTITUTION DRAFTING

1. Draft/Review Constitution
2. Draft/Review Constitutional Amendments

B. BASIC LAWS

1. Draft/Improve Civil Code
2. Draft/Improve Civil Procedure Code
3. Draft/Improve Criminal Code
4. Draft/Improve Criminal Procedure Code
5. Draft/Improve Administrative Code
6. Draft/Improve Human Rights Laws/Treaties
7. Draft/Improve Other Legislation, e.g., media legislation, NGO legislation

C. LEGISLATIVE PROCESS

1. Improve legislative procedures, e.g., committee system/review process in Parliament
2. Improve legislative drafting skills
3. Create/Upgrade Legislative drafting center (institution building)
4. Develop/Strengthen Legal think-tanks
5. Improve transparency and consensus building/legislative advocacy

D. CODIFICATION AND ORGANIZATION

1. Develop/Upgrade Computer data base

E. ENFORCEMENT

1. Train judges, lawyers, other officials, and public on improved enforcement
2. Develop/Improve other enforcement mechanisms

STRENGTHENING JUDICIAL SECTOR INSTITUTIONS

A. LAW RE-ORGANIZING THE JUDICIARY

1. Draft/Improve Laws on Organization of Judiciary

B. CREATION OF NEW COURTS

1. Create/Improve Constitutional Court
2. Create/Improve Supreme Court (court of cassation)
3. Create/Improve Specialized commercial or administrative courts

C. SELECTION AND APPOINTMENT OF JUDGES

1. Develop/Upgrade Judicial qualification examinations

D. TRAINING OF JUDGES

1. Improve/Provide training before appointment
2. Improve/Provide Pre-appointment training
3. Improve/Provide on-going after appointment training
4. Develop/Upgrade Judicial Training Center (institution building)

E. INFORMATION DISSEMINATION TO JUDGES

1. Develop/Upgrade Law libraries
2. Disseminate Laws Disseminate
3. Other decisions
4. Improve access to computers

F. ENHANCING JUDICIAL INDEPENDENCE

1. Improve security and length of tenure
2. Ensure sufficient salary
3. Strengthen immunity
4. Increase budgetary control
5. Develop self-governing bodies
6. Increase protection for judges

G. BUILDING JUDICIAL ACCOUNT ABILITY

1. Develop/Disseminate ethics code
2. Institute/Upgrade disciplinary proceedings

H. TRANSPARENCY

1. Initiate/Increase the frequency of open hearings
2. Initiate/increase the accessibility of written decisions

I. JUDICIAL POWERS

1. Shift arrest and search warrants from pro curacy
2. Review only through appellate procedures
3. Strengthen finality of judgments

J. ENFORCEMENT POWERS

1. Introduce/strengthen contempt powers
2. Improve enforcement agency (bailiff s service)

K. COURT ADMINISTRATION

1. Improvement concepts/approaches introduced to judiciary/needs assessed

2. Develop New filing systems
3. Increase Computerization
4. Develop/ Introduce new clerical positions
5. Introduce/Upgrade Alternative Dispute Resolution (ADR) (court annexed or not)

L. INFRASTRUCTURE

1. Undertake Courthouse reconstruction

M. JUDICIAL ASSOCIATIONS

1. Strengthen institution building
2. Expand Advocacy
3. Improve Media relations
4. Increase membership in international organizations
5. Improve Member services, e.g., publications, education, etc.

N. BAR REFORM (LAWYERS AND NOTARIES)

A. NEW LAW ON THE BAR

1. Draft/pass law on privatization of the bar
2. Increase Independence of the bar (no MOJ control)
3. Expand freedom to represent clients

B. ADMISSION STANDARDS

1. Develop/Upgrade admissions standards

C. INSTITUTION BUILDING

1. Create/Develop Bar Associations
2. Increase Transparency
3. Improve/introduce Membership services including education programs, e.g., how to establish/run a private practice
4. Increase membership in international organizations and links to bar associations elsewhere
5. Introduce dues structure/practices

D. ETHICS

1. Ensure new standards adopted

E. CONTINUING LEGAL EDUCATION (CLE)

1. Development/implementation of continuing education programs and publications

F. PUBLIC SERVICE

1. Establish/Expand Pro bono representation (access) Introduce/Improve Legal Information Centers/Programs

PROCURACY REFORM

A. POWER SHIFTING

1. Shift "Supervisory" powers
2. Improve Relations with the judiciary
3. Improve Relations with the police

B. TRAINING

1. Train Concerning new commercial crimes
2. Training on Anti-corruption
3. Train on Human rights
4. Train on International standards

LEGAL EDUCATION REFORM

A. NEEDS ASSESSMENT

1. Assess needs for curricula and methods reform

B. ACCREDITATION STANDARDS

1. Creation of law school associations
2. Work with Ministry of Education

C. OBJECTIVE ADMISSION STANDARDS

1. Develop/apply objective law school graduation standards

D. CURRICULUM REFORM: SUBJECT MATTERS TAUGHT

1. Develop/Expand Human rights curricula
2. Enhance Modem commercial law curricula
3. Expand Ethics curricula

E. CURRICULUM REFORM: TEACHING METHODOLOGY

1. Introduce/Expand Interactive methods in the classroom
2. Introduce/Expand objective testing and standardization of testing
3. Create/Upgrade Internships
4. Develop/Expand Clinical programs

F. INFRASTRUCTURE

1. Create/Expand Libraries
2. Improve Computer resources
3. Build/Upgrade Moot courtrooms

G. LAW STUDENT ASSOCIATIONS

1. Form Law Student Associations
2. Increase/Improve Competitions
3. Expand/Upgrade Publications
4. Increase Membership in international organizations

PUBLIC AWARENESS

A. PUBLIC EDUCATION PROGRAMS

1. Introduce Advertising regarding specific laws
2. Develop/Distribute Pamphlets: "Your Rights"
3. Institutional strengthening of relevant NGOs
4. Enhance Public school curriculum

JOURNALIST TRAINING

1. Train on Public education role
2. Train on "Watchdog" role
3. Improve Relations with Government

ACCESS TO JUSTICE

A. PUBLIC DEFENDERS OFFICES (Government supported)

1. Improve Work on Criminal matters
2. Improve Work on Civil matters

B. LEGAL RESOURCE CENTERS (REPRESENTATION AND CONSULTATION) (NGOs)

1. Develop/Expand Legal Information Center /Programs
2. Develop/Expand Environmental Law Center
3. Develop/Expand Gender/Children Law Assistance
4. Develop/Expand Minority representation resources
5. Develop/Expand Human Rights Assistance

C. OMBUDSMEN

1. Introduce/Upgrade Ombudsman

D. LAW SCHOOL CLINICS

1. Introduce/Expand Law Clinics

APPENDIX C: USAID DG RESULTS FRAMEWORK

To view the results framework, please go to:

http://www.usaid.gov/our_work/democracy_and_governance/publications/pdfs/pnacc390.pdf

and click on "Rule of Law Index – hyperlinked"
    or, go to page 17

then click on "Elections Index – hyperlinked"
    or, go to page 57

then click on "Civil Society Index – hyperlinked"
    or, go to page 116

then click on "Governance Index – hyperlinked"
    or, go to page 151

APPENDIX D: ACTIVITY TYPOLOGY DEVELOPED BY DG OFFICE

Civil Society					
Advocacy & Oversight	**CSO Capacity Strengthening**	**Labor & Trade union**	**Civic / Minority Participation**	**Media**	**Conflict prevention, reconciliation**
Policy Analysis	Resource mobilization (fundraising, recruitment)	Anti-Sweatshop legislation	Civic Education	Print, Broadcast journalist training	Inter-faith/ ethnic dialogue
Legislative Drafting	CSO Organizational Development	Trade Union strengthening	Voter Education		
Lobbying	Legal Framework (law drafting, advocacy)		Legal rights assistance		
Creation of consultative mechanisms	Intermediate service provision		Grassroots mobilization		
Support for think tanks			Rights awareness campaigns		
Budget analysis and review			Women's political participation campaigns		
Human Rights Campaigns					

Elections and Political Parties		
Organizational Development	**Electoral Competition**	**Legislative Party Building**
Strategic Planning & Resource Development	Campaign Development, Candidate Selection	Inter-party relations
Research Capacity (surveys)	Voter registration	Legislative Drafting
Local/Regional party organization strengthening	Election Administration	Party transitions in government
Media Relations & Message Development	Poll Watching	Support for Effective Opposition
Membership recruitment & Fundraising	Partisan Codes of Conduct	Constituency development services
	Platform Development	Coalition Building
	Voter Outreach Initiatives	Caucus development (women, etc
	Conflict Mediation & Negotiations	
	Public Opinion Polling	
	Youth and Women participation	
	Electoral law reform	

Rule of Law			
Human Rights Protection	**Legal Frameworks**	**Access to justice**	**Justice institutions**
CSO support for human rights monitoring and advocacy	Support for drafting and amending Constitutions	Legal Education	Improvement of court administration
Truth and reconciliation commissions	Development of Criminal & Civil procedure codes, Commercial laws	Legal services for disadvantaged	Training Judges, Magistrates, Lawyers, Prosecutors
		Alternative dispute resolution	Support for Public Defenders
			Strengthening police & law enforcement

Governance		
Decentralization & Local Governance	**Legislative Strengthening**	**Anti-Corruption**
Local/municipal finance, financial management	Technical Support to Legislative Members and Staff	N/A
Decentralization implementation support	Constituency Relations	N/A
Decentralization policy analysis	Oversight Hearings and Investigations, Legislative and Committee Processes	N/A
Support to associations of sub-national public officials	Legislative Transparency & Information Flow	N/A
Information flow to enhance transparency and accountability	Legislative drafting and research	N/A
Material support to local governments	Budget and Fiscal Processes	N/A
	Public Hearings or Forums	N/A

APPENDIX E: STANDARDIZED CODING FORM FOR EVALUATIONS

Below we list a coding form that we recommend for use in completing evaluations. The sheet provides a format to guide the evaluators in giving basic information on the activities and the information used to assess activity/program success. We recognize that these sheets do not cover all aspects that are needed in an evaluation. We encourage that those doing an evaluation separately include material that is not covered in the items in this form.

A.		**BASIC INFORMATION**			
1		Institutional Author(s) of Evaluation (List all)			
2		Individual Author(s) of Evaluation (List all)			
3		Year of Program Evaluation			
	a.	Program Evaluation Start Date	Mo=	Yr=	No info
	b.	Program Evaluation End Date	Mo=	Yr=	No info
4		Unit of study			
	a.	Country/Countries in study (List all)			
	b.	Sub-national unit(s) in study (List all)			

Comment:

B.		**SECTOR CHARACTERISTICS**			
1		DG sub-sectors (Circle most applicable one)	1. General 4. Rule of Law	2. Civil Society 5. Governance	3. Election 6. No info

Comment:

2	Activities (List all, e.g., training of judges, public opinion poll)	

Comments:

3	Main *target* groups (List all)	

Comments:

4		Input to activities: (list separately for each activity if possible)	
	a.	Funding provided	
	b.	Personnel involved (person hours)	

c.	Time of activity implementation	Mo=	Yr=
d.	NGOs that undertook activities (List them all)		

Comments:

5	What outputs, outcomes, and impacts were considered? (list separately for each activity if possible)	
a.	Individual (e.g., judges, leaders, voters)	
b.	Group (e.g., political party, NGOs, etc)	
c.	Institutional (e.g., offices in government, ministries, rules and laws)	
d.	National level (i.e., democracy)	

Comments:

e.	Long-term (longer than 1 yr)	1. Yes (a. Success; b. Failure)	2. No
f.	Short-term (shorter than 1 yr)	1. Yes (a. Success; b. Failure)	2. No

Comments:

C.	**SAMPLE SELECTION**	
1	Why were country or area selected for activity? (Circle applicable ones)	1. Values of outcome 2. Unusual case 3. Importance of case 4. Variety in cases 5. Data availability 6. Random Sampling 7. Familiarity (e.g., country expert evaluators, AID's history in the country, etc) 8. Other (name) 9. No info

Comments:

D.	**RESEARCH DESIGN**	
1	Give baseline characteristics/values for outputs, outcomes, or impacts that are relevant to aim of activity. (list for each activity)	
2	Give same characteristics during and after activity. (list for each activity)	

Comments:

3	Research method for evaluation (Circle applicable one(s)).	1. Interviews (How many?:) 2. Documents (Types?:) 3. Field Work (Time in country: Mo= Da= Yr= ; Place in country:) 4. Quantitative (describe) 5. Other (describe)

Comment

4	Method of Control (Circle applicable one(s))	1. Case selection	2. Other methods (describe)	3. No controls (explain why)
	List other groups (not receiving USAID funding) that were considered in the evaluation			

Comments:

5	What factors (or variables) are controlled when making the evaluation? (describe)	
6	Country context (briefly describe most relevant aspects)	

Comments:

7	Does country-context influence USAID activity? (if so, describe how)	

Comments:

8	Alternative explanations? (describe)	
a.	Within country	
b.	Foreign	

Comments:

9	Did prior change in sector bring in USAID activity? (explain answer)	
10	Were there any other donors or NGOs that were working on the same or similar activities? (describe)	

Comments:

E.	**MEASUREMENT**	
1	Define objects that are targeted by the activity.	
2	Give measures used to indicate these.	
3	Are there any alternative measures of these same outputs, outcomes, or impacts?	
4	Same variables assessed in each country? (if multiple countries)	
5	Assess the quality and accuracy of the quantitative or qualitative information that you are using in the evaluation? Any reason to suspect error or bias? If so, any estimates on its magnitude?	

Comments:

Coder's overall comments:

1.

APPENDIX F: DRAFT SURVEY OF MISSION OFFICERS

Annual Trends Analysis Questionnaire

1. What was your DG budget for the following years? Please specify funding source as appropriate:

	DA	ESF	INC	SEED/FSA
FY2002 Actuals	___________	___________	___________	___________
FY2003 OYB	___________	___________	___________	___________
FY2004 Request	___________	___________	___________	___________

2. In your best estimation for FY2003, what is the relative mix of DG programming in your Mission (please express in dollar terms, totaling 100% of the total from the last question)? Also, please indicate whether you view the sector as increasing, decreasing or staying the same in terms of the share of your Mission's budget:

	Amount	Trend
Rule of Law/Justice	________	_____________
Human Rights	________	_____________
Elections Administration & Monitoring	________	_____________
Political Party Development	________	_____________
Civil Society	________	_____________
Labor	________	_____________
Media	________	_____________
Legislative Strengthening	________	_____________
Decentralization/Local Government	________	_____________
Anti-corruption	________	_____________
Other/Specify	________	_____________

3. In the areas that you indicated above, please describe (in a few words) the primary focus areas of the program (i.e. “Rule of Law: court administration, access to justice and alternative dispute resolution”).

4. Does your Mission have any cross-cutting programming that includes democracy and governance, or that contributes to DG objectives? If so, briefly elaborate.

5. What has been your Mission’s greatest DG success of the past year?

6. What has been your Mission’s greatest DG disappointment of the past year?

7. In your view, thinking of the country in which you work, what are the greatest challenges to the consolidation of democracy?

8. In terms of their influence on democracy and good governance, how would you rank order these underlying dynamics that frequently characterize developing countries? Please rank for your country, from one (greatest impact) to five (least impact):
    - Lack of consensus on the democratic rules of the game? ______
    - Lack of inclusion of significant portions of the population in the country's political life? ______
    - Lack of competition, both in political and social terms (competition of ideas/interests) ______
    - Lack of effective rule of law ______
    - Lack of good governance ______

9. For your country, please respond to these statements about the quality of some key aspects of democracy and good governance on a scale from one from one (terrible) to five (superb). Then, please indicate, in your estimation, whether the trend in recent years, for this category, has been for conditions to worsen, stay the same, or improve. Use the blanks at the end of this question to add any additional areas that you think are significant.

	Quality (1-5)	**Trend (worse, same, better)**
"Advocacy NGOs are effective."	_______	_______
"Governance is transparent and accountable."	_______	_______
"Government institutions are effective."	_______	_______
"Human rights are respected."	_______	_______
"The justice system works."	_______	_______
"Political parties aggregate citizens' interests."	_______	_______
"There are checks on the power of the executive."	_______	_______
________________________________	_______	_______

10. Beyond statements in the MPP and Mission documents, do you feel that democracy promotion is a high priority in your Mission? Is it a high priority for the U.S. Embassy? Please rank your estimation for each as low, medium or high.

    U.S Embassy priority on democracy promotion (low, medium or high): _______

    Mission priority on democracy promotion (low, medium or high): ____________

11. Are there any important issues you face that have not been elicited by this questionnaire or sufficiently explored?

Appendix G1: Size of the Program in the Past 10 Years, Political Processes (Indicated by Budget Obligation Amount)

	Program Size World Ranking		Program Size		Number of Activities	
Country	96 to 01 (Avg)	90 to 95 (Avg)	96 to 01 (Avg)	90 to 95 (Avg)	96 to 01 (Avg)	90 to 95 (Avg)
AFR Region						
Nigeria	4	31	3988	740	2.50	1.00
Liberia	8	29	2555	774	2.33	1.50
DR Congo	18	.	1050	.	2.00	.
Zimbabwe	18	.	1050	.	1.50	.
Mozambique	22	13	965	2231	2.00	2.40
Kenya	27	66	830	48	1.00	1.00
Malawi	29	33	749	595	1.25	2.00
Ghana	31	5	683	4300	1.33	3.00
Mali	37	19	525	1426	1.50	2.00
Uganda	40	52	500	304	1.00	1.33
Guinea	44	68	391	25	1.33	1.00
Benin	47	.	361	.	1.00	.
Burundi	48	32	358	674	1.00	2.33
Gambia	49	18	300	1454	1.00	3.00
Ethiopia	52	21	253	1278	1.33	2.25
Senegal	53	39	231	445	1.00	1.00
Guinea-Bissau	54	59	220	208	1.00	1.00
Niger	58	.	156	.	2.00	.
Angola	59	3	150	5214	2.00	2.00
South Africa	60	4	149	4992	1.00	2.00
Somalia	61	.	104	.	1.00	.
Sudan	61	.	104	.	1.00	.
Zambia	.	36	.	500	.	1.00
Tanzania	.	45	.	359	.	1.00
Eritrea	.	47	.	350	.	1.00
Cote D'ivoire	.	50	.	328	.	1.00
Madagascar	.	65	.	57	.	1.00
Cameroon	.	69	.	16	.	1.00
ANE Region						
West Bank/Gaza	1	1	11892	10004	1.00	1.50
Indonesia	3	11	4904	2746	3.00	1.33
Cambodia	9	30	2516	752	1.60	4.00
Mongolia	36	.	544	.	1.40	.
Bangladesh	46	54	372	291	1.00	2.00
Philippines	49	35	300	520	1.00	1.00
Thailand	.	16	.	1581	.	1.00
Egypt	.	20	.	1386	.	1.00
Sri Lanka	.	23	.	1086	.	2.00
Lebanon	.	25	.	983	.	2.00
Nepal	.	26	.	957	.	1.33
Afghanistan	.	27	.	896	.	2.00
Yemen	.	40	.	431	.	2.00
E&E Region						
Bosnia	2	.	11055	.	1.50	.
Ukraine	5	17	3812	1471	1.83	2.00
Russia	6	9	3724	3137	1.67	2.00
Armenia	10	48	2337	329	2.40	2.00
Kosovo	11	.	1911	.	1.50	.
Georgia	12	58	1576	217	3.25	2.00

Albania	13	24	1425	1083	2.00	3.00
Azerbaijan	14	56	1405	248	1.80	2.00
Serbia	16	.	1275	.	2.00	.
Bulgaria	20	44	1047	364	1.33	1.00
Latvia	21	43	984	380	3.00	1.00
Romania	23	12	960	2341	1.25	2.00
Belarus	25	62	931	99	1.80	1.50
Croatia	28	46	792	351	1.50	1.00
Macedonia	30	.	712	.	2.25	.
Kazakhstan	33	37	632	466	2.17	1.50
Moldova	34	60	631	177	1.50	2.00
Kyrgyz Republic	38	49	520	329	2.00	1.50
Tajikistan	39	64	515	71	2.00	1.50
Turkmenistan	42	67	475	40	2.00	2.00
Uzbekistan	43	61	413	145	1.25	2.00
Slovakia	45	41	385	407	1.33	1.00
Hungary	51	.	260	.	2.00	.
Poland	55	55	219	286	1.00	1.00
Lithuania	61	63	104	89	1.00	2.00
LAC Region						
Haiti	7	2	3283	5823	2.20	3.60
El Salvador	15	8	1319	3591	2.50	2.17
Nicaragua	17	14	1054	1705	2.25	2.67
Paraguay	24	.	953	.	1.75	.
Peru	26	10	867	3109	1.60	3.00
Mexico	32	.	672	.	2.60	.
Guatemala	35	42	591	398	1.50	2.20
Dominican Republic	41	22	487	1093	1.00	1.50
Bolivia	56	38	174	446	1.75	1.80
Guyana	57	34	166	555	1.25	1.25
Colombia	.	6	.	4250	.	1.00
Panama	.	7	.	3964	.	1.60
Honduras	.	15	.	1692	.	2.25
Costa Rica	.	28	.	831	.	1.60
Chile	.	51	.	304	.	1.00
Jamaica	.	53	.	300	.	1.00
Ecuador	.	57	.	243	.	1.33

Appendix G2: Size of the Program in the Past 10 Years, Civil Society (Indicated by Budget Obligation Amount)

	Program Size World Ranking		Program Size		Number of Activities	
Country	96 to 01 (Avg)	90 to 95 (Avg)	96 to 01 (Avg)	90 to 95 (Avg)	96 to 01 (Avg)	90 to 95 (Avg)
AFR Region						
South Africa	10	2	5761	12932	2.50	7.00
DR Congo	14	.	4160	.	2.00	.
Nigeria	22	50	3001	750	3.50	2.50
Burundi	23	46	2741	826	1.67	3.40
Angola	24	20	2622	2235	2.50	2.00
Mali	38	54	1659	642	3.67	2.50
Ethiopia	44	24	1412	1980	2.20	2.75
Kenya	45	79	1355	160	2.00	2.00
Mozambique	50	23	1017	2169	1.40	3.00
Liberia	52	55	983	624	2.50	4.50
Zimbabwe	57	.	937	.	1.60	.
Malawi	58	57	912	610	1.83	1.67
Zambia	59	26	869	1570	1.50	2.75
Guinea	61	49	796	755	1.80	1.75
Rwanda	62	14	773	4045	1.17	3.50
Tanzania	65	38	634	1000	1.80	1.67
Namibia	66	71	627	348	1.17	1.67
Somalia	67	.	608	.	1.50	.
Madagascar	69	41	565	960	1.25	1.00
Senegal	70	52	534	716	1.33	1.00
Niger	72	68	483	397	3.00	2.00
Ghana	73	67	434	400	1.33	2.00
Sudan	74	.	416	.	1.00	.
Uganda	76	58	372	600	1.33	2.33
Benin	78	12	326	4363	1.17	3.00
Guinea-Bissau	80	60	312	555	1.00	1.75
Gambia	81	34	300	1167	2.00	2.00
Eritrea	83	35	199	1100	1.33	2.00
Sao Tome/Principe	.	74	.	300	.	3.00
Togo	.	76	.	278	.	1.00
ANE Region						
Indonesia	3	27	12399	1513	3.17	1.40
East Timor	4	.	11342	.	1.00	.
Egypt	6	17	9237	3072	2.17	2.25
West Bank and Gaza	8	9	7192	4615	1.17	3.00
Burma	17	.	3680	.	1.33	.
Philippines	18	15	3529	3685	2.00	3.00
Cambodia	39	13	1603	4344	2.83	3.60
India	46	.	1200	.	1.00	.
Bangladesh	47	39	1103	990	2.40	2.20
Mongolia	54	.	972	.	1.60	.
Nepal	64	56	639	617	1.50	2.00
Sri Lanka	77	37	330	1029	1.00	3.00
Jordan	79	.	320	.	2.00	.
Morocco	84	77	190	266	1.00	1.00
Afghanistan	.	36	.	1085	.	2.00
Lebanon	.	44	.	831	.	2.00
Yemen	.	70	.	383	.	1.50
Tunisia	.	73	.	322	.	1.00
Thailand	.	75	.	289	.	2.00

E&E Region						
Serbia	1	.	14459	.	2.25	.
Russia	2	1	13209	24048	3.17	5.00
Kosovo	5	.	10475	.	3.00	.
Bosnia	7	11	8848	4424	1.00	2.00
Ukraine	9	5	6142	6908	3.67	5.50
Bulgaria	11	22	4577	2174	5.33	5.00
Azerbaijan	12	43	4445	882	2.50	4.00
Montenegro	13	.	4228	.	1.50	.
Croatia	19	4	3403	8093	2.17	3.00
Slovakia	20	7	3239	6302	3.67	4.00
Georgia	21	28	3227	1406	3.33	4.50
Kazakhstan	25	18	2406	2946	2.17	4.00
Poland	26	6	2379	6743	2.50	5.00
Armenia	27	16	2331	3461	2.17	8.00
Albania	28	29	2241	1395	2.33	4.00
Romania	29	10	2173	4568	1.67	4.00
Uzbekistan	31	45	1986	829	2.33	4.50
Macedonia	32	21	1981	2216	3.00	2.00
Kyrgyz Republic	37	30	1673	1353	2.33	4.50
Hungary	42	8	1500	5850	2.00	2.00
Belarus	43	53	1452	683	2.00	7.00
Tajikistan	49	40	1042	982	1.83	4.00
Moldova	51	59	1004	562	1.60	4.00
Turkmenistan	55	42	972	887	1.60	6.00
Cyprus	60	.	857	.	1.00	.
Lithuania	68	51	596	736	1.25	4.00
Czech Republic	82	32	284	1252	2.00	3.00
Latvia	86	47	59	824	1.00	2.00
Estonia	.	63	.	485	.	2.00
Slovenia	.	80	.	120	.	2.00
LAC Region						
Colombia	15	.	4000	.	1.00	.
Haiti	16	31	3855	1337	2.17	4.00
Bolivia	30	48	2057	802	3.20	2.40
Guatemala	33	62	1978	512	2.17	2.60
El Salvador	34	3	1848	9214	3.00	6.80
Dominican Republic	35	33	1805	1203	2.17	3.40
Cuba	36	.	1750	.	1.00	.
Nicaragua	40	19	1589	2440	4.20	4.20
Peru	41	25	1574	1728	3.33	5.20
Ecuador	48	64	1052	451	1.40	2.20
Honduras	53	69	982	396	2.40	1.80
Panama	56	61	962	554	3.50	1.75
Paraguay	63	82	696	96	1.50	1.00
Jamaica	71	.	500	.	1.00	.
Guyana	75	78	388	220	1.40	1.00
Mexico	85	66	94	405	1.00	1.00
Costa Rica	.	65	.	418	.	1.50
Belize	.	72	.	325	.	2.00
Chile	.	81	.	117	.	2.00

Appendix G3: Size of the Program in the Past 10 Years, Rule of Law (Indicated by Budget Obligation Amount)

	Program Size World Ranking		Program Size		Number of Activities	
	96 to 01 (Avg)	90 to 95 (Avg)	96 to 01 (Avg)	90 to 95 (Avg)	96 to 01 (Avg)	90 to 95 (Avg)
AFR Region						
South Africa	6	7	3683	4316	2.33	1.83
DR Congo	15	.	2250	.	1.50	.
Rwanda	18	37	2149	547	1.83	2.50
Nigeria	23	65	1275	24	1.50	1.00
Madagascar	26	62	1163	57	1.00	1.00
Senegal	36	.	872	.	1.67	.
Tanzania	39	21	733	917	1.00	1.67
Kenya	40	63	718	40	1.33	1.00
Liberia	44	.	634	.	2.40	.
Angola	46	.	566	.	2.40	.
Mozambique	47	18	525	1058	1.00	1.67
Sierra Leone	48	.	500	.	1.00	.
Ethiopia	50	47	473	303	1.20	1.25
Zimbabwe	52	.	406	.	4.00	.
Guinea-Bissau	53	31	404	631	1.00	1.00
Zambia	54	44	387	362	1.00	1.00
Burundi	56	25	358	803	2.00	2.67
Malawi	57	38	341	525	1.60	1.33
Uganda	59	51	265	254	1.00	1.33
Eritrea	61	36	248	550	1.50	1.00
Mali	64	61	117	90	2.00	2.00
Namibia	65	53	95	220	1.00	1.50
Benin	66	.	88	.	1.00	.
Somalia	.	4	.	8500	.	3.00
Ghana	.	26	.	780	.	2.00
Cote D'ivoire	.	32	.	630	.	1.00
Gambia	.	39	.	522	.	1.67
Togo	.	67	.	4	.	1.00
ANE Region						
Egypt	3	1	8953	21789	1.00	1.00
Indonesia	9	16	2873	1226	3.00	1.25
Cambodia	14	24	2300	828	2.00	4.00
Mongolia	29	.	1004	.	1.50	.
Bangladesh	41	58	696	141	1.00	1.75
Nepal	58	19	299	1003	1.00	2.33
Vietnam	63	.	160	.	1.00	.
Afghanistan	.	14	.	1496	.	3.00
Sri Lanka	.	27	.	723	.	2.00
West Bank/Gaza	.	28	.	705	.	2.00
Philippines	.	45	.	351	.	2.00
E&E Region						
Kosovo	1	.	39043	.	2.50	.
Russia	4	3	5427	12592	2.17	3.00
Ukraine	5	11	4339	2663	3.00	3.00
Kazakhstan	8	13	3218	1536	2.50	2.00
Azerbaijan	10	56	2648	162	1.60	1.00
Kyrgyz Republic	11	22	2505	910	2.67	1.50
Croatia	13	64	2330	36	1.83	1.00
Georgia	16	42	2221	394	2.83	2.00

Armenia	17	54	2212	194	2.33	4.00
Bosnia	19	.	2100	.	1.25	.
Bulgaria	20	20	1761	952	1.80	2.00
Belarus	24	41	1255	405	1.80	1.50
Albania	27	35	1149	593	1.60	1.00
Uzbekistan	32	50	960	255	1.83	1.50
Macedonia	34	23	912	846	1.83	1.00
Turkmenistan	37	60	840	102	2.00	1.00
Moldova	42	43	692	370	1.17	1.50
Romania	45	.	616	.	1.40	.
Slovakia	49	46	497	313	1.00	1.00
Tajikistan	51	55	421	172	1.60	1.00
Lithuania	62	48	162	280	1.00	2.00
Poland	.	10	.	3858	.	1.00
Estonia	.	52	.	250	.	1.00
Czech Republic	.	59	.	131	.	1.00
LAC Region						
Haiti	2	2	9774	13777	3.00	3.00
Guatemala	7	15	3402	1268	2.50	2.67
Dominican Republic	12	57	2410	150	1.50	1.00
El Salvador	21	8	1753	4284	2.50	2.00
Cuba	22	.	1750	.	1.00	.
Nicaragua	25	9	1244	3975	2.67	2.60
Peru	28	30	1048	652	2.33	2.67
Panama	30	6	990	5676	4.00	2.60
Honduras	31	17	987	1149	2.50	1.50
Mexico	33	.	957	.	2.80	.
Ecuador	35	49	887	260	1.60	1.00
Bolivia	38	12	800	1866	2.00	1.67
Guyana	43	40	676	467	1.00	1.00
Paraguay	55	66	386	11	1.50	1.00
Jamaica	60	33	250	622	1.00	1.25
Brazil	67	.	13	.	1.00	.
Colombia	.	5	.	6770	.	2.60
Costa Rica	.	29	.	656	.	1.25
Chile	.	34	.	599	.	1.00

Appendix G4: Size of the Program, Other Governance (Indicated by Budget Obligation Amount)

	Program Size World Ranking	Program Size	Number of Activities
Country	97, 99, and 01 (Avg)	97, 99, and 01 (Avg)	97, 99, and 01 (Avg)
AFR Region			
South Africa	10	5448	1.00
Nigeria	14	4059	2.00
Zimbabwe	20	3180	2.00
Kenya	28	2329	2.00
Guinea	38	1481	1.00
Mozambique	41	1334	1.50
Benin	42	1243	1.50
Ethiopia	43	1237	1.00
Senegal	44	1041	1.50
DR Congo	45	1025	1.00
Rwanda	49	749	1.00
Mali	50	730	1.50
Uganda	52	686	1.00
Ghana	53	676	1.00
Namibia	55	647	1.00
Malawi	56	633	1.00
Angola	57	500	1.00
Tanzania	61	350	1.00
Eritrea	62	312	1.00
Zambia	67	50	1.00
ANE Region			
West Bank/Gaza	1	25078	1.00
Egypt	4	12094	2.00
Indonesia	13	4681	5.00
Burma	16	3492	1.00
Lebanon	23	2841	1.00
India	30	2193	1.00
Mongolia	36	1568	1.00
Cambodia	47	850	1.00
Bangladesh	59	414	1.00
Philippines	60	400	2.00
E&E Region			
Bosnia and Herzegovina	2	23661	3.00
Cyprus	3	12342	1.00
Russia	5	10945	2.00
Ukraine	6	9276	3.00
Serbia	7	8450	2.00
Montenegro	8	7000	1.00
Kazakhstan	9	6294	4.00
Armenia	12	4855	2.00
Kyrgyzstan	15	3989	4.00
Bulgaria	18	3282	3.00
Georgia	19	3223	2.67
Macedonia	21	3092	3.00
Albania	22	3026	2.50
Romania	27	2397	1.50
Croatia	33	1919	1.50
Azerbaijan	34	1718	1.00
Moldova	35	1677	1.50

Uzbekistan	37	1527	2.50
Belarus	40	1420	2.00
Tajikistan	48	778	2.00
Turkmenistan	64	290	1.00
Kosovo	65	200	1.00
LAC Region			
Cuba	11	4989	1.00
Guatemala	17	3470	2.00
Haiti	24	2800	1.00
Bolivia	25	2478	1.00
Mexico	26	2452	2.00
Honduras	29	2223	1.50
Paraguay	31	2065	2.00
Ecuador	32	1997	2.00
El Salvador	39	1425	1.00
Panama	46	1000	1.00
Jamaica	51	697	2.00
Guyana	54	670	1.00
Peru	58	429	1.00
Dominican Republic	63	311	1.00
Nicaragua	65	200	1.00

APPENDIX H: SAMPLE OF CROSS-NATIONAL DATA ON DEMOCRACY

Data Source Name	Author	Pub Yr	Variable	Year	Country	Region	Measure
Overall Democracy Indexes							
"Liberal Democracy: Validity and Method Factors in Cross-National Measures." American Journal of Political Science 37: 1207-1229.	Bollen, Kenneth, A.	1993	Liberal democracy index	72-97	Most	General	
Polity IV project: Political Regime Characteristics and Transitions, 1800-2001	CIDCM, Monty G. Marshall and Keith Jaggers (principal investigators) & Ted Robert Gurr (founding director)	2003	POLITY	1800-2001	Most (except for small countries)	General	Range = -10 to 10 (-10 = high autocracy; 10 = high democracy) Combined Polity Score: Computed by subtracting AUTOC (autocracy score) from DEMOC.

Governance Variables							
Political Handbook of the World; Cross-National Time-Series Data Archive Users' Manual	Banks, Arthur	annual (the latter one in 1979)	Legislative effectiveness	50-	Most	General	0 = No legislature exists, 3.3 = Ineffective. There are three possible bases for this coding: 1) legislative activity may be essentially of a "rubber stamp" character; 2) domestic turmoil may make the implementation of legislation impossible; 3) the effective executive may prevent the legislature from meeting, or otherwise substantially impede the exercise of its functions, 6.7 = Partially Effective. A situation in which the effective executive's power substantially outweighs, but does not completely dominate that of the legislature, 10 = Effective. The possession of significant governmental autonomy by the legislature, including, typically, substantial authority in regard to taxation and disbursement, and the power to override executive.

Fiscal Decentralization indicators	World Bank		Vertical Imbalance (the degree to which subnational governments rely on central government revenues to support their expenditures)		about 3-60, overall 40-60 range	General	% (intergovernmental transfers as a share of sub-national expenditure)
Fiscal Decentralization indicators	World Bank		Sub-national expenditure		about 3-60, overall 40-60 range	General	% of GDP
Fiscal Decentralization indicators	World Bank		Sub-national revenues		about 3-60, overall 40-60 range	General	% of GDP
Fiscal Decentralization indicators	World Bank		Tax revenue		about 3-60, overall 40-60 range	General	% of total sub-national revenues and grants
Fiscal Decentralization indicators	World Bank		Transfers from other levels of Government		about 3-60, overall 40-60 range	General	% of total sub-national revenues and grants
Database of Political Institution	Keefer, Philip	2002	AUTHOR	75-00	Most	General	Federalism, Do Sub-national govts. have extensive taxing, spending or regulatory authority?

Elections Variables							
Political Handbook of the World; Cross-National Time-Series Data Archive Users' Manual	Banks, Arthur	annual (the latter one in 1979)	Party legitimacy	50-	Most	General	0 = No parties, or all but dominant party and satellites excluded, 3.3 = Significant exclusion of parties (or groups), 6.7 = One or more minor or "extremist" parties excluded, 10 = No parties excluded
Political Handbook of the World; Cross-National Time-Series Data Archive Users' Manual	Banks, Arthur	annual (the latter one in 1979)	Competitiveness nomination processes	50-	Most	General	0=non-competitive, 5=partially competitive, 10=competitive
IDEA Voter Turnout from 1945 to Date	IDEA		Presidential Elections: Total votes	45- (vary across countries)	171	General	
IDEA Voter Turnout from 1945 to Date	IDEA		Presidential Elections: Registration	45- (vary across countries)	171	General	
IDEA Voter Turnout from 1945 to Date	IDEA		Presidential Elections: Vote/Registration	45- (vary across countries)	171	General	

IDEA Voter Turnout from 1945 to Date	IDEA		Parliamentary Elections: Total votes	45- (vary across countries)	171	General	
IDEA Voter Turnout from 1945 to Date	IDEA		Parliamentary Elections: Registration	45- (vary across countries)	171	General	
IDEA Voter Turnout from 1945 to Date	IDEA		Parliamentary Elections: Vote/Registration	45- (vary across countries)	171	General	
Rule of Law Variables							
Governance matters III: governance indicators for 1996 to 2002	World Bank	2003	Rule of Law	96, 98, 00, 02	Most	General	the extent to which agents have confidence in and abide by the rules of society. -2.5 to 2.5
World Human Rights Guide	Charles Humana	86, 92	EXJKIL (Human extrajudicial killings)	85, 91	Most?	General	Measure 0 = Indicates a constant pattern of violations of the freedom from extrajudicial killings or 'disappearances.'; 3.3 = Frequent violations of the freedom from extrajudicial killings or 'disappearances.'; 6.7 = Occasional breaches of respect for the freedom from extrajudicial killings or 'disappearances.'; 10 = Unqualified respect for the freedom from extrajudicial killings or 'disappearances.'
World Human Rights Guide	Charles Humana	86, 92	TORTR (Torture by state)	85, 91	Most?	General	0 = Indicates a constant pattern of violations of the freedom from torture or coercion by the state.; 3.3 = Frequent violations of the freedom from torture or coercion by the state.; 6.7 = Occasional breaches of respect for the freedom from torture or coercion by the state.; 10 = Unqualified respect for the freedom from torture or coercion by the state.
Civil Society Variables							
The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Legal Environment	97-02	17-28	E&E	The legal status of non-governmental organizations.
The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Organizational Capacity	97-02	17-28	E&E	The operation of NGOs.

The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Financial Viability	97-02	17-28	E&E	NGOs' resource mobilization capacity.
The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Advocacy	97-02	17-28	E&E	NGOs' record in influencing public policy.
The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Service Provision	99-02	17-28	E&E	NGOs' ability to efficiently provide services that consistently meet the needs, priorities and expectations of their constituents.
The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Infrastructure	99-02	17-28	E&E	The strength of sectoral infrastructure that can provide NGOs with broad access to Intermediary Support Organizations that provide local NGO support services.
The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Public Image	97-02	17-28	E&E	Public awareness and credibility of NGOs
The NGO Sustainability Index for Central and Eastern Europe and Eurasia	The DG Office, Bureau for Europe and Eurasia, USAID	annual	Composite measure of all the dimensions	97-02	17-28	E&E	
World Human Rights Guide	Charles Humana	83, 86, 92	ASSC (Assembly and association)	82-83, 85, 91	Most?	General	0 = Indicates a constant pattern of violations of the freedom of peaceful assembly and association; 3.3 = Frequent violations of the freedom of peaceful assembly and association; 6.7 = Occasional breaches of respect for the freedom of peaceful assembly and association; 10 = Unqualified respect for the freedom of peaceful assembly and association.
World Human Rights Guide	Charles Humana	86, 92	MONRHT (Freedom to Monitor Human Rights Violations)	85, 91	Most?	General	Measure 0 = Indicates a constant pattern of violations of the freedom to monitor human rights violations.; 3.3 = Frequent violations of the freedom to monitor human rights violations.; 6.7 = Occasional breaches of respect for the freedom to monitor human rights violations.; 10 = Unqualified respect for the freedom to monitor human rights violations.

APPENDIX I: CODING R4 REPORTS

The Results Review and Resource Request (the R4) is an annual document where a Mission office describes its Strategic Objectives (SOs) and summarizes their implementation and performance over the past year. After reviewing a sample of the R4 reports and the Annual Budget Submission Reports (prior to the mid-1990s), we conclude that by constructing a systematic coding protocol and template based on a clear concept system, two pieces of key information about the USAID programming after the mid-1990s—the DG sector and programs undertaken under each agency objective—could be extracted. In this section, we present our brief review of the R4 and recommendations about future data compilation.

Four items in the R4 labeled the “Agency Strategic Framework” (listed as 2.1 to 2.4) are identical to the DG Office’s “Agency Objectives,” “Strategic Support Objectives,” and the four “DG sectors” i.e., rule of law, elections/political processes, civil society, and governance. We recommend (1) that all the USAID documents follow the same terminology for referring to the group of these four sector areas; and (2) that the Agency Strategic Framework be coded for information about the DG sectors in which a given Mission is actively involved. Earlier R4 reports may not explicitly provide information on the Agency Strategic Framework. In that case, the DG Agency Objective information would need to be coded based on the information provided under “Intermediate Results,” which are subsumed under the Agency Strategic Framework.

Intermediate Results (IR’s) provide more specific information at the program level under a given Agency Objective. Unfortunately, the way that information is presented does not exactly follow the USAID Results Framework (Appendix C of this document), and is not consistent across different R4 documents. In some cases, Intermediate Results and sub-Intermediate Results are reported as the same level, whereas in others, the key terms (e.g., Agency Objective, Strategic Objective, Intermediate Result, indicators, and so forth) are conceptually mixed up. Further, some (sub-) Intermediate Results presented in the R4 are multiple-barreled in that they may belong to multiple Intermediate Results or Agency Objectives.
Problems are more frequent in earlier reports. They would demand a clear, consistent coding protocol guided by the Results Framework, with special care on the part of coders. We recommend that all program level information be coded at the Intermediate Result (2.x.x.) level (and sub-IR’s be coded as the IR’s under which they are subsumed). By so doing, the IR’s in most of the R4 reports seem to provide reasonably reliable information about the program undertaken. We also suggest that multiple-barreled IR’s be dissected and coded in line with the Results Framework. Finally, coders need to be informed of possible inconsistency in the use of the key terms and the need to read the program and key result descriptions to glean information.

The followings are two examples:

(1) Mongolia 2003 R4: One of the IRs states the improvement of “the effectiveness of Parliament and the judiciary.” This IR clearly covers two Agency Objectives (i.e., elections and rule of law) and description that follows suggests that it incorporates five IR’s in the Results Framework (IR 2.1.3. “Equal access to justice,” 2.1.4. “Effective and fair legal sector institutions,” 2.2.4. “Effective oversight of electoral process,” 2.2.5.

“Representative and competitive multiparty system,” and 2.2.6. “Inclusion of women and disadvantaged groups”).

(2) Ukraine 1996 R4: One of the “Strategic Objectives” (S.O. 2.3 “More effective, responsive and accountable local government”) corresponds to one of the Intermediate Results in the Results Framework (IR 2.4.1. “Increased government responsiveness to citizens at the local level”). Meanwhile, some IR’s listed as 2.x.x.x in the report (e.g., 2.1.2.1. “Greater transparency in government decision-making”) actually correspond to the 2.x.x. (program) level in the Results Framework (IR 2.4.2. “Increased citizen access to improved government information).

Cases such as above should be reorganized and coded according to the Results Framework.

In addition to these two key pieces of information, the R4 also reports information of the program implementers (contractors and grantees). We propose that the number and the names of contractors and grantees, their country base, and if available, the Agency Objectives and the Intermediate Results in which they are active, be coded. As just implied, the information about the Agency Objectives and the Intermediate Results that contractors and grantees worked on is not consistently provided across the R4 documents.

This coding activity would focus on the R4 report; the Annual Budget Submission Reports are essentially a project proposal statement, and as such, do not seem to serve for the data collection purpose here.

Finally, it is strongly recommended that USAID require Missions to report this information according to a systematic format so that future data coding/compilation would be made efficient and reliable. Specifically, future R4 reports should follow the exact conceptual hierarchy formulated in the “Results Framework.” A template for this part of the report would solve this problem both for data compilation and reporting. Attached is a suggested coding template for the existing R4 reports.

Coder Name Coding Date

Country/Organization

1. Sector and Intermediate Results

Agency Strategic Framework (Agency Objective Level)	Intermediate Results (Program Level)
Instruction: According to "Primary Links to Agency Strategic Framework," enter numbers (1 to 4) in the order of Mission priorities. If no % figure is provided or links are not mentioned, leave it blank.	Instructions: Check applicable IR's the Mission seeks to gain to achieve the SO. If the report lists sub IR's, check the IR's under which the sub IR's are subsumed.
___ 2.1 Rule of Law: Rule of Law and respect for human rights of women as well as men strengthened	________ 2.1.1.Foundations for protection of human rights and gender equity conform to international commitments ________ 2.1.2. Laws, regulations, and policies promote a market-based economy ________ 2.1.3. Equal access to justice ________ 2.1.4. Effective and fair legal sector institutions
___ 2.2 Elections and Political Parties: Credible and competitive political processes encouraged	________ 2.2.1. Impartial electoral framework ________ 2.2.2. Credible electoral administration ________ 2.2.3. An informed and active citizenry ________ 2.2.4. Effective oversight of electoral process ________ 2.2.5. Representative and competitive multiparty system ________ 2.2.6. Inclusion of women and disadvantaged groups ________ 2.2.7. Effective transfer of political power
___ 2.3 Civil Society: The development of politically active civil society promoted	________ 2.3.1. A legal framework to protect and promote civil society ensured ________ 2.3.2. Increased citizen participation in the policy process and oversight of public institutions ________ 2.3.3. Increased institutional and financial viability of CSO's ________ 2.3.4. Enhanced free flow of information ________ 2.3.5. Strengthened democratic political culture
___ 2.4 Governance: More transparent and accountable government institutions encouraged	________ 2.4.1. Increased government responsiveness to citizens at the local level ________ 2.4.2. Increased citizen access to improved government information ________ 2.4.3. Ethical practices in government strengthened ________ 2.4.4. Strengthened civil-military relations supportive of democracy 2.4.5. More effective, independent, and representative legislatures 2.4.6. Enhanced policy processes in the executive branch

2. Contractors/grantees info

Organization Name	Contractor or grantee?	US-based?	Direction: List up the Agency Objective and IR's (following the Result Framework) in which this contractor or grantee was active (Note: this information may be provided in the result and/or performance section of R4 reports. If no information is provided, circle the "No Info."			
			Agency Objective	IR	IR	
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
____	C / G	Y / N	( )	( )	( )	No Info
	C / G	Y / N	( )	( )	( )	No Info

2. Other assistance agency info.

	Donors					
___	No other donors than USAID					
	Name: ____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
___	Other US government agency					
	Name: ____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
___	Other government(s) or their agencies					
	Name: ____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
___	Trans-Regional organizations (e.g., EU, ASEAN)					
	Name: ____	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	____					
	____					
___	International organizations (e.g., UN, UNDP, WB)					

Name:					
Name:	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info
	( ) ROL	( ) ELE	( ) CS	( ) GOV	( ) No info

APEENDIX J: REPORT ON USAID DATA

Bruce Kay, a USAID staff member, investigated possible internal data sources that could be used in the evaluations we propose in this document. We thank Kay for preparing this memo for us and attach it here without making any changes. Margaret Sarles responded to his memo, and we again reproduce it here without changes.

MEMO

To:	USAID Staff, Advisory Board
From:	Bruce Kay
Subject:	INVENTORY of USAID DEMOCRACY PROGRAM DATA
Date:	August 27, 2003

Introduction:
At the last meeting, the Advisory Board members suggested that information on the availability and reliability of data collected by USAID on its democracy programs would partially shape its recommendations for the selection of methods that comprise the research design being developed by SSRC. At the Board's request, I put together some basic information on these data to help advance the design effort. I investigate data in the following categories: money, people, time, other donors, partners, geographic location.

I believe that already collects a great deal of potentially useful information in one form or another, though much of that data do not exist in a useful format for analysis and would have to be adapted and cleaned -- an onerous, labor-intensive job in some instances; easier in others. Where possible, I have estimated the effort that would be required to collect 22 country/years of data, from 1980-2002 at the lowest possible unit of disaggregation, the so-called "activity" level. I would hope that the Board could consider these estimates and weight them against the potential utility of the data to be collected.

The 3 levels, from lowest to highest:
Activity: A description of assistance within sector with at least these characteristics:

1. Type (training, technical assistance, financial/material assistance);
2. Recipient (organization, institution);
3. Goal: ("capacity building", etc.)

Sector: All activities grouped by four areas: Governance, Civil Society, Rule of Law, Elections and Political Processes.

Program: All Sectors of Democracy & Governance portfolio for country/year

Country: All Programs in the USAID mission portfolio, including DG (Economic Growth, Trade, Health, Education, etc.)

I. Money (e.g., DG funding appropriation and outlays in USAID presence countries)

Source: USAID Budget Office; Office of Resource Allocation
Contact: Jonathan Richter jrichter@usaid.gov
Website: http://inside.usaid.gov/M/B/eacoding_page.htmBruce's data.doc
Data Link: http://inside.usaid.gov/M/B/ProgramPivot.xls
Codebook: http://inside.usaid.gov/M/B/emcodes4.doc
Coverage: 1990-2002 in electronic form (Excel worksheet); 1980-1989 in hard copy in W253, USAID library
Unit[1]: Activity: "Code" allegedly corresponds to contracts in each sector
Collection: 1980-1989 data for all USAID countries.
Estimate: 20 person-days to collect, clean and organize 1980-89 data (in conjunction with time data, see below)

Description:

USAID Management Bureau's budget office maintains yearly spending figures for all country/program areas of USAID going back to the first year of the Agency. As far as Democracy and Governance is concerned, however, there are limitations. "DG" did not exist as a budget category until 1990, so spending data for democracy per don't exist for the period 1980-1989, when democracy spending was lumped under "Human resources" and "Other."

To get access to data prior to 1990, we'd have to consult the attractively titled "W253", a report issued annually by the Budget Office from the 1960s onward. This is available in the USAID library and compiles democracy activity data under the two general headings ("human resources" and "other"). Data are not electronic; they must be extracted from hard copies. I estimate that time needed to collect an additional 10 years of data for all USAID countries would be about 30 person-days.

Our financial data from 1990 to 2002 exist in electronic form, organized by country and year and then disaggregated by the following elements:

1. Sector (democracy and governance health, etc.)
2. Bureau (USAID division; Europe & Eurasia, Latin America and Caribbean, Africa, Asia & Near East Democracy Conflict & Humanitarian Assistance);
3. Emphasis Area Code (codebook provided). This category is broader than sub-sector and may include multiple sub-sectors. The Emphasis Areas include:

a) DRHR Human Rights
b) DEPP Political Party Support
c) DCCS Civil Society-General
d) DGAC Anti-Corruption/Govt. Transparency & Accountability
e) DRLJ Legal and Judicial Development
f) DEEA Elections Assistance
g) DCLA Labor

h) DGOV Other Governance

4. Sub-Sector (Rule of Law, Civil Society, Governance, Political Processes);
5. Funding Source (USAID, State Department, specially earmarked funds) Codes are as follows:
   a) DA=USAID funding;
   b) ES=State Department ("ESF") Funding;
   c) Congressional earmarks: SS=Sub-Saharan Africa; SH-Shale; AI=Support for Eastern European democracy (SEED); NI-Newly Independent States (Freedom Support Act)
6. Activity (coded: training judges, etc.) These codes are unique; there is one code per activity per country. A codebook gives a short description of the activity. We need to examine it to see exactly what is meant by "activity."

For information on the codes, see the Budget Office Codebook:
http://inside.usaid.gov/M/B/eacoding_page.htm

I have attached data to this correspondence. Data are updated annually, are searchable.

Note: People split hairs over the accuracy of these sums and what they represent; i.e., whether they reflect appropriation, "obligation", and outlay. There are a dozen different ways of defining the term "spent" in USAID. According the USAID Budget Office, the figures refer to "the amount of money in thousands of dollars that the United States made available for implementation of DG programs in each country-year, with the presumption that those programs were in fact implemented. In other words, it reflects the appropriation, or level of commitment of the United States to democracy activities in the country. It does not represent the exact amount disbursed by the United States for democracy assistance in that country."

With Budget Office data, it is also possible to collect and/or compute the following:

Measure		**Utility**
Sub-sectors or emphasis areas that comprise DG program	Code sub-sector as either present or absent for each country/year	Measure of complexity of DG program
Number of Sectors or emphasis areas engaged in DG program	Count the number of sectors engaged in DG portfolio per country years	Measure of complexity of DG program
Numbers of Activities per a) sub-sector b) emphasis c) activity	Count the number of activities sponsored by USAID listed at each level, for each country/year	Gross measure of extensiveness of DG program
Average spending in a) sub-sector b) area of emphasis c) activity per country/year	Divide total $ by # of a) sub-sectors, areas of emphasis, activities per country/year	Measure of concentration or diffusion of spending in DG portfolio (lots of $s on few sectors/activities, etc.)

II. People (e.g., DG/Mission Staff in all of the USAID presence countries)

Source: USAID Personnel Office; Mission Staffing Pattern System (MSPS)
Website: http://inside.usaid.gov/M/HR/wwsp/userm.htm
http://inside.usaid.gov/M/HR/wwsp/
http://inside.usaid.gov/M/HR/wwsp/wwsprpt/
Coverage: 1994 to present
Unit: Country
Collection: 1994-Present: data are collected in electronic form, organized by country/quarter but last quarter (September) could be used as annual report
Estimate: 2 person days to clean and organize.

1980-1993 data are in the intriguingly titled "W490", and are organized by country/quarter - but there is little variation between quarters.

Estimate: 10 person days to collect clean and organize.

Total: 12 person days

Description:

USAID Management Bureau's Office of Human Resources publishes a quarterly World Wide Staffing Pattern report. These quarterly reports are based on the Mission Staffing Pattern System (MSPS), a mechanism by which the USAID missions report on the status of the workers at their missions, by employment category. It shows the total number of employees in USAID by funding source, employment type, broken out by region and also shows totals by countries within each region.

Significantly, these data do not break down employees into program or sector but by country, then by employee category (USAID direct hire, contractor, etc.). The data do not indicate how many staff works on democracy support activities in the country. That number would have to be estimated based on some function of the size of the country's DG program (say 2 individuals per million of program dollars) and/or other factors that determine the staffing patterns of USAID missions.[2]

III. Time (e.g., Duration of implementation of DG program, sector, activity, in months or years from first funded year)

Source: USAID Budget Office; Office of Resource Allocation
Website: http://inside.usaid.gov/M/B/eacoding_page.htm
Coverage: 1990-2002 in electronic form (Excel worksheet); 1980-1989 in hard copy in W253, USAID library
Unit: Activity: "Code" corresponds to contracts in each sector
Collection: 1990-2002 data collected. $ figures recoded, then variables created indicating time in years of each sector for each country year. One person-day of effort.

1980-1989 data for all USAID countries.
Estimate: 20 person-days to collect, clean and organize 1980-89 data

Total: 21 person-days

Description:

There is no reason why we cannot use financial information to provide information about implementation time. The Budget Office maintains records on spending for each country going back to 1961, the year the Agency was formed. The year a country Democracy program first appears on USAID's accounting sheets, a DG program has begun. It is really that simple. The same is true for the initiation of a program, a sector and an activity. These data of course are reported in country/years; there is no way to determine whether an activity was in place for a partial year, in other words.

There are a few anomalies in these data. For example, there are instances in which funding is reported for one year drops to zero from some level and then resumes in a subsequent year. These "spell gaps" may indicate a suspension in the program, or they may be an accounting glitch. We'll need to investigate and fill in the blanks.

IV. Other Donors

Source: USAID Congressional Budget Justification (CBJ)
USAID publications (online, 1997-present) The Activity Data Sheet "Other Donor Programs"
See the USAID public websites below.
http://www.usaid.gov/pubs/cp97/
http://www.usaid.gov/pubs/cp98/
http://www.usaid.gov/pubs/cp99/
http://www.usaid.gov/pubs/cp2000/
http://www.usaid.gov/pubs/cp2001/
http://www.usaid.gov/pubs/cp2002/

Coverage: 1997-2002 in electronic form (not in spreadsheet)
Unit: Activity Level (spotty); Sector (inconsistent); Program (Consistently reveals presence of other donors)
Collection: Unknown.

If we need actual financial data, collection will be onerous, and depend on the response time of other donor organizations or countries. If we need to know only whether there were other donor organizations present in democracy assistance efforts, this will be easier to collect. We could use the CBJ source first and then fill in the gaps later.

Estimate: 10 person-days, but this would be done in conjunction with data collection on partners (See V. below)

Description:

The Congressional Budget Justification for USAID does reveal the presence of other donor organizations within the countries in which USAID also maintains a presence. These sources also tend to name the donor organizations that are working in the democracy and governance area. However, USAID does not collect the detail on other donor programs in any systematic way: expenditures, implementation time, personnel, etc, so there is no way of knowing from our internal data what level of effort was sustained by how many other organizations or countries in which USAID is implementing a democracy and governance program. These data would likely have to be obtained from the donor organizations themselves.

To get general data on the existence of other donors, use the website above, select region and country. There is both an entry for Other Donors at the DG program level, which reports whether other countries have assistance efforts in place in that country for that year.

As the Albania CBJ (below) indicates, there is also a separate entry for "Other Donor Programs" by activity. It reports which donors are also involved in that particular democracy support activity during the current year. These data are suspect in terms of reliability, though; it isn't clear for example whether omitting the mention of a particular donor in a particular year means that the donor didn't have a program in place at the time.

These summaries are too vague to be of much use beyond recording the presence of another donor or donors. It is hard to determine the number of organizations that maintain support programs in that sector for that year. It is noted, for example, that "some [Albanian] parties receive support and assistance from their equivalent party organizations in Western Europe."

PROGRAM: ALBANIA
TITLE: Increased, Better-informed Citizens' Participation in Political and Economic Decision-Making, 180- SO02.1
PROPOSED OBLIGATION AND FUNDING SOURCE: FY 2000: $3,260,000 SEED Act
INITIAL OBLIGATION: FY 1991 ESTIMATED COMPLETION DATE: FY 2002

Other Donors Programs: Some of the political parties are receiving support and assistance from their equivalent party organizations in Western Europe. Denmark's Dialogue for Development is supporting the Albanian Civil Society Foundation which, in turn, is supporting Albania's NGO community. The Organization for Security and Cooperation in Europe has been very active in the past year in a range of programs supporting electoral processes and the constitutional referendum and in the promotion of political stability. Other donors include the Council of Europe (EU), Soros Foundation, the World Bank, and the European Bank for Reconstruction and Development EBRD.

V. Partners (organizations, institutions that receive USAID grants/contracts and with whom USAID collaborates in the implementation of its DG programs in USAID presence countries)

Source: USAID Congressional Budget Justification (CBJ); USAID publications (online, 1997-present); The Activity Data Sheet "Principal Contractors, Grantees, or Agencies" item[3]
See the USAID public websites below.
http://www.usaid.gov/pubs/cp97/
http://www.usaid.gov/pubs/cp98/
http://www.usaid.gov/pubs/cp99/
http://www.usaid.gov/pubs/cp2000/
http://www.usaid.gov/pubs/cp2001/
http://www.usaid.gov/pubs/cp2002/

USAID publications (hardcopy 1980-1996)
Coverage: 1997-present in electronic form 1980-1989 in hard copy in W253, USAID library
Unit: Activity
Collection: 1997-2002- Coding & counting from website
Estimate: 10 person-days to collect and clean 1980-1996-coding /counting from hardcopy
Estimate: 15 person-days to collect and clean
Total: 25 person-days to collect and clean

Description:

USAID maintains information on its activities in the Annual Report (2001-present); the R4 (1995-2000), and, finally, the Congressional Budget Justification. It is hard to say which is the authoritative source, but the CBJ maintains a consistent reporting format, and it covers the 1980-2002 period. The last several years (since 1997) the CNJ has been posted in electronic form to the USAID website that is maintained by the publications office.

To get the electronic data on partners, go to the website for each CBJ for every year, and then click on the appropriate region. At the regional site, you then click on the country link, and scan the Activity Data Sheet. The Activity Data Sheet contains an entry every year for "Principal Contractors, Grantees, or Agencies" item. These items may be coded as follows: 1=US Organization;

2=Host country Organization;
3=3rd country Organization;
4=host government institution.

Because partners are listed under each activity, they may be counted, yielding a count of implementing partners for each activity. These might be aggregated to sector and program level. In the Kyrgyzstan example below, there are 8 organizations listed. These lists of partners are mostly complete since missions are loath to exclude a partner.

(Sample) ACTIVITY DATA SHEET

PROGRAM: KYRGYZSTAN
TITLE AND NUMBER: Increased Citizens' Participation in Political and Economic Decision-making, 110-SOO2.1
STATUS: Continuing
PROPOSED OBLIGATIONS AND FUNDING SOURCES: FY2000: $2,750,000 Freedom Support Act
INITIAL OBLIGATION: FY93; ESTIMATED COMPLETION DATE: FY 2004

Principal Contractors, Grantees or Agencies: USAID-funded US PVOs include the American Bar Association/CEELI, Internews, the National Democratic Institute (NDI), and the International Foundation for Electoral Systems (IFES), are playing a role in promoting democracy in Kyrgyzstan. Counterpart International, Aid to Artisans, and the International Center for Not-for-profit Law, which together form the Counterpart Consortium, also participate in local NGO development. Initiative for Social Action and Renewal in Eurasia (ISAR) will also participate in the development of local environmental NGOs.

Other data on partners that might be collected using this source include:

1. Whether or not a host government entity is a collaborator or interlocutor in the assistance program in any given activity. Activity data indicates whether some branch of government or government personnel were involved in training or technical assistance, or whether government personnel are deliberately excluded from democracy support activities by policy or by mission decision.

2. What branch or level of government is engaged in the effort
   a) Executive, Legislative, Judicial branch
   b) Central government, local (regional, city) government

I include these measures only because there seems to be a lot of interest in the agency in testing the proposition that engaging with "key government entities" in the DG assistance efforts tends to advance the democratization process.

VI. Geographic Location (e.g., local vs. national)

Unfortunately, USAID does not do a good job of collecting data on the geographical location of its programs in each country, and I have identified no good source of data for this information as yet.

The Congressional Budget Justitification and the R4 both report in a general way whether certain DG activities are being implemented in selected regions or municipalities. It is usually (but not always) the case, for example, that where USAID is implementing a local government and decentralization program, USAID is spending money or providing assistance to organizations or institutions outside the capital. Hese reports fail to mention consistently at least the exact localities or the number of localities outside the capital.

These sources may also mention a program or set of programs that correspond to a sector (legal aid clinics in a Rule of Law program) that are being implemented on a "pilot" basis or in a sub-set of regions or municipalities. But these data are not collected or reported to Washington in any systematic way, to my knowledge.

This latter question will require some more investigation.

That's all for the moment. I do hope this little inventory was helpful.

Bruce Kay

[1] Refers to the smallest possible unit at which reliable data are obtainable

[2] Margaret Sarles informs me that the overall DG budget has little relation to the staff size in some missions. DG funding alone may not be a valid means of estimating the size of the DG staff.

[3] (Example: Nepal http://www.usaid.gov/pubs/cp2000/ane/nepal.html)

Below is a comment by a USAID staff member, Margaret Sarles, on Bruce Kay's data memo. We thank Sarles for sharing this information based on her experience as a USAID staff member and attach it without making any changes.

Bruce, This is by far the most comprehensive and thorough work that has been done to see what is available in USAID/Washington, the best thing we have in the Agency so far. Congratulations.

At the same time, I think we need to be very careful not to convey (even implicitly) that the data are more available and more useful than is likely to be the case, and therefore give outsiders a mistaken impression. This is particularly true at the activity level, where it is easy to raise false expectations. We haven't yet even analyzed the "codebook" for activity-level information, but it is possible that it is nothing but identification numbers of contracts.

The chart on page 3 highlights the difficulty I have with the activity level data. It states that the number of activities would be a good "gross measure of the extensiveness of a program." Unless this is a definitional tautology, the number of activities is not usually any measure of how extensive a program is. It would be misleading to use it in this way. Activities are often not comparable across missions, or even within missions. Is one trip to the US by the contractor or head of an NGO equivalent in any way to, for example, a one-day training conference or a one year training of judges? This measure strikes me as unacceptably vague, and not measuring anything real.

People: The discussion of people that begins on page 3: The heart of the matter, which is found embedded in the discussion is that USAID does not collect data on people broken down by technical area, so we have no data on how many people are in a democracy office in any mission. As we discussed, there is no legitimate way to guess what that number would be from program size or any other variable we have so far thought of.

But by beginning with "how we could clean up the data" and "how to use it," the memo conveys the idea that it would be worthwhile to do so. Why not simply state that so far we have not been able to find the data we need, and cannot find anyplace it is compiled within the agency?

Other Donors: I concur with you on the difficulty of getting useful data on other donors from published sources or websites. I know from personal experience in rule of law in Latin America that actually getting the information from just two other donors - World Bank and the Inter-American Development Bank-was not easy at all. And particularly for looking at "when their program started" in comparison to our own (to answer the interesting question "Did our early work lead to the large donors taking over? Can we be said to "leverage" the other donors' funds?) turned out not be readily available on websites, or indeed, at the easy fingertips of specialists in these agencies. Basically I succeeded by interviewing enough people who had worked in LAC long enough on these issues to compile their individual knowledge.

Query: Are PMPs from each country collected somewhere in PPC? The Performance Measurement Plan, with indicators at every level, is required for each mission after approval of a strategy, so it may be that they are sent to Washington. Since the PMPs include all the indicators of a given mission and for each sector and probably multiple indicators in each sector for particular activities, they could be a rich source of data. This might actually amount the best descriptive material available at the "activity" level. Even if not available in Washington, if there are electronic copies in each mission, getting hold of them might not be very difficult. Have you explored this possibility?