USAID
FROM THE AMERICAN PEOPLE

EVALUATION OF THE SERBIA GOOD GOVERNANCE MATRIX

DECEMBER 2010

This publication was produced for review by the United States Agency for International Development. It was prepared by Social Impact, Inc..

EVALUATION OF THE SERBIA GOOD GOVERNANCE MATRIX

DECEMBER 2010

James Fremming, Team Leader
Andrew Green, Senior Technical Adviser

Social Impact, Inc.

2300 Clarendon Blvd, Suite 300
Arlington, VA 22201

This publication was produced for review by the United States Agency for International Development under contact number RAN-I-00-09-00019.

ACKNOWLEGEMENTS

The evaluation team is especially grateful to Ellen Kelly, USAID/Serbia Rule of Law Adviser and Contract Officer's Technical Representative for this evaluation, for her willingness to share her deep knowledge of the GGM and the Serbian good governance context with us. She also was exceptionally helpful to the team in many ways as we came to familiarize ourselves with Serbia and to enjoy the setting as our time allowed.

In addition we wish to thank all those who took their time to talk with us or to provide us with information as we requested it. Outstanding in this regard is Milos Mojsilovic of the Center for Democracy and Free Elections, who was fully responsive to our many requests and cooperative in our probes for documentary evidence. His assistance was critical to the successful execution of this review.

Our interpreter and logistics specialist, Masa Matijassevic-Simic, who has served on other Social Impact activities in Serbia, was expert and diligent in her double role.

The technical content of the report remains the responsibility of the authors.

1 Contents

ACRONYMS

ABA-CEELI	American Bar Association-Central Europe and Eurasian Law Initiative
ADS	Automated Directives System
CBJ	Congressional Budget Justification
CeSID	Center for Democracy and Free Elections
CSO	Civil Society Organization
DG	Democracy and Governance
GGM	Good Governance Matrix
M&E	Monitoring and Evaluation
RFA	Request for Assistance
RFP	Request for Proposals
SI	Social Impact
TS	Transparency Serbia
UNDP	United Nations Development Programme
USAID	United States Agency for International Development
USG	United States Government

EXECUTIVE SUMMARY

Evaluation Objectives and Methodology

USAID/Serbia selected Social Impact, Inc., to conduct an evaluation of the Good Governance Matrix (GGM), a multi-dimensional data set and data collection process that provides document-confirmed data on the extent to which Serbian central and municipal government institutions comply with relevant Serbian law and internationally recognized good practice in transparent and accountable governance. The evaluation was to carry out a systematic assessment of GGM data quality, recommend strategies for improvements to data quality and efficiency in data collection, and offer an approach to the sharing of GGM results with participating public institutions and the Serbian public at large. In addition, USAID requested an analysis of the estimated costs of future administration of the GGM, including needed methodological refinements.

The evaluation was conducted in October and November of 2010 by a team of two international consultants. Evaluators reviewed relevant USAID and GGM documents; utilized focus groups and semi-structured personal interviews to gather data from GGM research staff and others knowledgeable of GGM data collection and analysis; reviewed GGM scoring approaches and associated documentation; reviewed training materials and approaches used to prepare and support GGM research staff; and gathered information pertaining to the costs of conducting a research endeavor such as GGM in Serbia.

Key Findings

Conceptually inspired by the World Bank Institute's *Governance Matters* and other leading professional resources for professionals working to promote and support governance, the GGM has been administered twice – in 2008 and 2009 -- by a consortium of Serbian pro-democracy CSOs with funding and programmatic support from USAID/Serbia. Both administrations of the institutional survey engaged samples of approximately 20 central agencies and municipalities. The Matrix covers five broad dimensions of governance (structure, accountability, management, open entry and competition, and public participation). Data are collected through 25 thematically focused questionnaires that guide enumerator questions, as well as requests for documentary evidence of governance behaviors.[1] After enumerators have collected data from the institutions, a small group of Serbian analysts allocate scores of zero to five for each of 266 detailed "checks" for each institution. While analysts were allowed broad discretion in the scoring criteria they chose to apply, typically an institution received a zero score if supporting documentation of governance behavior was not produced by the institution. Scores at the check level were aggregated up a conceptual hierarchy (of sub-indicators, categories and dimensions) so that summary scores were produced at institutional and dimension levels. For example, the

[1] In its current form, utilized for the 2007 round, the GGM comprises:

- Five broad *dimensions* (Structure, Accountability, Management, Open Entry & Competition, and Public Participation);
- Twenty-five *categories* (five per dimension);
- Fifty-three *sub-indicators* (two to four per category); and
- Two hundred sixty-six "*checks*" with direct references to questionnaire and documentary evidence sources (four to six per sub-indicator).

scores for two "checks" are added together to produce a score for the sub-indicator called "Existence and Implementation of a Procurement Plan." The score for this sub-indicator is added to those of three others to form a score for the category called "Procurement Procedures and Enforcement." This score is averaged with those for 4 other categories to derive the score for the dimension called "Open Entry/Competition." Finally, this score may be averaged with those of four other dimensions to arrive at a summary score for a public institution.

The GGM is a distinctive and potentially valuable tool for assessing and monitoring the quality of governance, in that it depends not on the transient opinions of business professionals or academics, but on intensive, structured documentation of particular behaviors of governments in five major dimensions of governance. It currently carries notable weaknesses, however, in that the engagement of Serbian institutions or other external stakeholders is very limited to date, GGM data mostly have not been shared with institutions participating in the survey, and no systematic review of GGM data quality has, up to this point, been carried out.

The GGM features significant limitations along five dimensions of data quality. First, data validity is constrained by inconsistent scoring across analysts and institutions. In addition, because there is no behavioral "anchoring" of institutional scores, the scores are not readily interpretable as particular types or levels of good governance.

Second, reliability of data is undermined by inconsistent scoring as noted above, and by the absence of formal guidance to analysts on how to score institutions. Scoring is not cross-verified among analysts—a critical omission. In addition, the evaluation team found that weak data management has resulted in widespread errors in the calculation of scores.

Third, GGM scores at the institutional level are presented at a precision level of an integer and a decimal place (for example, 2.0, 2.1, etc.). The evaluation team found this to be unduly precise, since a great deal of inference and analysis contributes to these overall scores, and the substantive significance of decimal-point differences in scores is not clear.

Fourth, timeliness of GGM data generally is also weak. Both administrations of the GGM thus far have featured data collection more than a year following the subject year of data collection (for example, the 2009 administration collected data on status of governance in 2007). USAID has thus far regarded administrations of the GGM as test runs. In addition, the Serbian elections of 2008 resulted in considerable changes in staff in public institutions, therefore making a data collection effort for that year unlikely to be productive.

On a more positive note, the integrity of GGM is quite high overall; the reliance upon document-based evidence reduces opportunities for inappropriate data manipulation. Data archiving, however, was shown to be quite weak, threatening data integrity.

Training of GGM researchers featured a single, day-long session of preparation for enumerators. No formal training session was conducted for analysts.

Conclusions

1. **The overall approach to GGM holds potential but also significant weaknesses**:
    a. The potential is great for the GGM to serve as a constructive, evidence-based tool for institutional diagnosis and a platform for civil society and citizen input to the monitoring of performance of government institutions. From a technical, quality-of-measurement perspective, however, the GGM is "not ready for prime time" as a basis for public dialogue. The almost complete absence of substantive involvement by the monitored institutions in GGM planning and data utilization notably weakens enumerators' ability to obtain full and accurate relevant data. Because public institutions have not been involved in GGM planning and have not been offered the opportunity to provide input to a plan for utilization of GGM data, there is very little sense of "ownership" of the GGM among the public institutions. The evaluation team's interviews show that this results directly in a general low level of responsiveness to GGM data requests. More active engagement by Serbian public institutions in the GGM's design, implementation and results reporting is critical to maintaining and increasing the Matrix's value for the future. Pre-administration consultations and follow-through presentations of results to participating institutions will be key ingredients for the future success of GGM administrations. Effective utilization of GGM data by the institutions providing the data is an essential step in eventually "opening up" information on government performance to Serbian society at large.
    b. It is not clear that annual surveys are either needed or cost-effective; opinions of interviewees differ on this point. The evaluation team concludes that there is not an overriding need to conduct the GGM annually: biannual surveys would be adequate.
    c. Finally, the process of selecting institutions to participate in the GGM should be reconsidered in light of (a) the challenges to date in eliciting institutional responses to the GGM survey and (b) evolving shifts in relevance of various institutions to USAID's strategic priorities. A focus in future GGM rounds on municipalities, to the exclusion of central institutions, may prove more cost-effective than the current half-and-half mix of central and local institutions.

2. **GGM features strengths and weaknesses in scoring, especially regarding extent of consistency in methodology:**

The five broad dimensions of governance included in the GGM are well-founded in the governance literature and the categories are reasonable and logical expressions of the dimensions. There is a critical need, however, for a simpler, more readily understandable, consistent system of weights to be applied across all checks, and for stronger data reliability assurance practices to be put into place.

3. **Additional data quality assessments and a preferred approach to data quality assessment use are needed:**

Effective data quality assurance can be built into the ongoing structure of GGM administration, supplemented by short-term training and technical assistance to get it started along the right path.

4. A structure is available to validate the methodology in the future:
Improving data quality and sustaining it into the future is quite feasible within the existing implementation structure. Focused training and technical assistance from Serbian or international experts will be required to institute new practices, including integration of scoring templates with protected areas of data and calculation, and consolidation of data for verification, analysis, and creation of tables and charts.

5. Additional tools are needed to make GGM data collection, analysis and reporting more efficient, reliable and consistent:
There clearly is a need to establish better data management at data collection, analysis and storage stages. Field pre-testing of data collection tools and processes, along with use of written protocols and cross-validation of scoring by analysts would produce significant improvements.

6. Training to members of the GGM research team has been limited and not sufficient:
Training of enumerators who gather data from the institutions was modest and incomplete, and training of analysts was essentially absent. Modest short-term training and technical assistance (particularly in guidance of pre-testing and cross-validation) are needed for future rounds of the GGM.

7. Opportunities available for sharing GGM results with institutions and the public:
Active engagement by Serbian public institutions and leadership will by necessity call for their involvement in identifying the approach to sharing and publicizing results. This means that, regardless of what this evaluation recommends regarding the publicizing of GGM results (see Recommendation "n" below), it should be considered only the beginning point for a discussion by the GGM implementing organization with Serbian public institutions. This discussion ideally would focus on how GGM data may be used to support both (a) broader societal engagement in improvements to governance and (b) internal improvements by the institutions to their own policies and operational processes. Since the GGM's potential strength is as a diagnostic tool, it may best be utilized within a context of quiet consultation, at least at first.

Recommendations

1. **Overall approach**
    a. **USAID and the GGM implementation team need to invest effort in more actively and substantively engaging Serbian public sector institutions in all phases of the GGM effort**. While USAID, through its implementing partner, needs to maintain ultimate control of the structure of the GGM and dissemination of results, details of implementation and sharing of results should be determined through structured consultations with leadership of the participating institutions.
    b. **The GGM should be conducted biannually, for example, in 2011, 2013 and 2015.**
    c. **Focus on municipalities as sample units in future GGM rounds.** If the GGM were to be more closely linked to an ongoing project supported by USAID or another donor, sampled institutions should include some that are not part of an

international assistance project in that sector, so that quasi-experimental comparisons could be made, thus supporting impact evaluation.

2. Specific recommendations

Please note that unless USAID is specified directly as recipient of a recommendation below, the recipient should be understood as the GGM implementing organization.

1) *Scoring should be improved by taking up several opportunities for simplification, consistency and overall data quality.* These include:

a) Reliability checks on scoring and documentation should be coordinated and executed by a qualified, local monitoring and evaluation (M&E) specialist who handles the spreadsheets and questionnaires. If necessary, a staff member of the partner organization could be trained in these skills by an external consultant.

b) Data quality assurance should include analyses of the differences within and between analysts' scoring, with a formalized in-house "mediation" procedure for resolving differences that may arise from cross-validation.

c) Documents supplied by participating institutions should be electronically scanned (only the first identifying pages are needed) and linked to the category scoring spreadsheet templates. This will reduce paperwork and increase efficiency in analysis and archiving.

d) To strengthen validity:
   i) Routinely conduct a field pre-test of the GGM in advance of full enumeration, and revise the data collection approach in light of pre-test results;
   ii) Remove sub-checks and conditional scoring criteria from checks; revise categories to more accurately reflect how issues relate to different types of institutions;
   iii) Design and apply a standard weighting scheme across all checks, with a simplified range of weight values; and
   iv) The GGM project should reflect the implicit priority of practice (de facto) over obligation (de jure). Whether this means that practice is 50% more important or twice as important is not the point; there simply needs to be a formalized priority to guide weighting of sub-indicators within categories.

e) To strengthen reliability: Use standard patterns of scoring values; as noted, scan documents to validate possession of them; establish and maintain a practice of cross-validation of scoring.

f) To strengthen integrity: Carry out the reliability support measures itemized above, and conduct selective analyses to check for patterns indicating bias in scoring.

g) To strengthen timeliness: Collect data for the most recent year available, not two years previous to the time of data collection. Provide feedback of results to institutions within two months of completion of enumeration.

h) To strengthen precision: Utilize decimal-point precision only at the check level; use integers or grouped scores ("high –medium-low") at other levels.

2) *There should be additional data quality assessments and a plan for data quality assessment use:* Some short-term, expert guidance will likely be necessary to build the data quality assurance capacity of any implementing organization, but in the longer term, the bulk of data quality assurance should be a routine part of GGM operations. The short-term consultation should include provision of, and training in the use of, a data quality assurance work plan (including a task checklist) for future rounds of the GGM.

3) *There should be specific procedures to validate the methodology in the future:* As the GGM moves into a phase of broader engagement and active feedback and dialogue, it will become evident that validation of the methodology will increasingly be "another part of the conversation" about how to improve public governance in Serbia. Formal requirements for validation of the methodology will vary based on the GGM's loci of ownership and use. For USAID, current ADS requirements call for formal assessment of data quality within three years of commencement of reporting, but this requirement only applies to USAID data that are reported to USAID/Washington. If we focus on Serbian institutions as the prime users, the practical approach would feature a selective review of data quality by external reviewers (less ambitious than this assessment but utilizing the same quality criteria), focusing on the key issues identified in this report, within a year after the next GGM administration. Following this, routine data quality assurance practices should be in place within the GGM, and external reviews would not be needed more frequently than every second or third GGM.

4) *Additional tools are needed to make GGM data collection, analysis and reporting more efficient, reliable and consistent:*

a) In selective (and rare) situations, utilize freedom of information rights as a method to reduce the workload of GGM enumerators and their working contacts in the public institutions. FOI requests may be most useful when the direct data gathering is particularly sensitive, as in information on procurement processes and oversight.
b) Apply written protocols for scoring by the GGM analysts and spreadsheet templates for category scoring, with protected layouts and calculation formulae.
c) Build summaries of institution performance, including graphs, upon cell references in spreadsheets so that summary reports can be developed with speed and accuracy.
d) For USAID, procurement announcements for future GGM rounds should include requirement of a detailed plan for data management, quality assurance and archiving.
e) Finally, in addition to switching from averaging to addition of dimension scores, to avoid unnecessary effort to achieve precision in scores the GGM should use ranges or bands of scores, rather than specific point scores, when reporting. For example, the GGM report may describe Accountability scores as "low," "medium" or "high" instead of using numerical scores with two digits to the right of the decimal point.

5) *Adequate training of members of the GGM research team should be ensured:*

Training may be delivered as a one-time effort that builds sustained staff learning into the overall GGM approach. It should be improved by:

a) Including role-plays of interviews, with attention to different situations in different kinds of responding institutions (especially national agencies compared to municipalities);
b) Utilizing a pre-test of the next GGM round with a few institutions, and using an after-activity review of the pre-test to serve as a training opportunity for enumerators along with analysts;
c) Systematically collecting and documenting feedback from enumerators and analysts regarding ways in which GGM design and data collection may be improved;
d) Expanding training for enumerators to two days; and

e) Adding new training for analysts, including having them work with a GGM dimension with which they are relatively unfamiliar, to enhance cross-learning among analysts and to refresh analytical perspectives.

6) *Sharing GGM results with institutions and the public:*

a) Rather than immediately publicizing institutional scores, use broader participation by civil society and the general citizenry when it can be best applied – when institutions have some minimum level of comfort with the GGM and the data it produces – so that they are prepared to respond to public input. In the next round of the GGM, make the circle of engagement much broader, utilizing informal technical consultations with institutions first, then exposure via professional forums such as the Standing Conference of Towns and Municipalities, and soon thereafter sharing with the public via web site summaries and town hall meetings.

b) To the extent feasible, coordinate with GGM participating institutions and senior leadership in the Government of Serbia in refining the approach to broader public sharing of GGM results. The more that the GGM may come to symbolize an attempt by USAID or by Serbian civil society organizations to be the "cops on the street," monitoring the quality of governance, the less likely it is for Serbian leaders to be able to comfortably make a case for constructively participating in the GGM data collection process and utilizing GGM data as a basis for institutional improvements.

Cost Analysis

The potential constructive contribution of the GGM to improved governance in Serbia is currently inhibited, as we have shown, by significant current weaknesses in technical design and data quality assurance. These limitations are reparable. Future rounds of the GGM need external expertise to guide the implementing organization in reforming its data collection and analysis processes to improve both data quality and efficiency of data collection and analysis. The evaluation team's review of GGM implementation costs, and consideration of options for quality improvements, results in a proposed cost structure that features a modest infusion of short-term, external expertise to the GGM implementing organization for training and technical assistance, as well as additional level of effort from the GGM implementer dedicated to cross-checking of scores and to consultations with institutional stakeholders. Overall, the estimated cost of a future GGM round with quality refinements is approximately 23 percent greater than the cost of the most recent administration. The funding level for the most recent round was seen by the implementing organization as adequate, even though the time required for data collection and analysis was larger than anticipated. While the proposed approach includes modest design revision and a "beefing up" of data quality control, it is also expected to deliver increased efficiencies in collection and analysis.

1 INTRODUCTION

In this report, we first present an overview of the program intentions behind the GGM's construction as well as basic elements of Matrix's technical design and implementation. Section 2 (GGM's Purpose, Design and Implementation) describes the design and methodology of the evaluation, while Section 3 establishes the evaluation's objectives and methodology. Section 4 (Findings: Data Quality Assessment) contains the analytical core of the assessment, structured according to five key dimensions of data quality. We provide a review of the quality of training provided to GGM implementation staff in Section 5 (Findings: Quality of Training of GMM Staff). The evaluation's conclusions and recommendations for improvement follow in Section 6. Finally, Section 7 utilizes the conclusions and recommendations to inform an analysis of the cost of implementing GGM, with emphasis on an efficient approach to delivering future versions of the Matrix with improved data quality.

2 GGM'S PURPOSE, DESIGN AND IMPLEMENTATION

The Good Governance Matrix was developed in two parts: its conceptual framework, by program monitoring and good governance/rule of law specialists at USAID/Serbia, and then operationally by a consortium led by the Center for Democracy and Free Elections (CeSID, a Serbian civil society organization based in Belgrade). Inspired in part by the World Bank Institute's *Governance Matters*, USAID saw the GGM as a contextual indicator of the quality of governance by central and local governmental institutions. The GGM was conducted in 2008 (for data as of 2006) and 2009 (for data as of 2007).[2]

The GGM was designed as a multi-dimensional measure with an extensive hierarchy. The purpose of multiple levels was not only conceptual, but also methodological: by disaggregating concepts to a low level and requiring documentation evidence, the GGM would allow analysts to assign scores in a more objective and defensible manner.

In its current form, utilized for the 2007 round, the GGM comprises:

- Five broad *dimensions* (Structure, Accountability, Management, Open Entry & Competition, and Public Participation);
- Twenty-five *categories* (five per dimension);
- Fifty-three *sub-indicators* (two to four per category);[3] and
- Two hundred sixty-six "*checks*" with direct references to questionnaire and documentary evidence sources (four to six per sub-indicator).

Please see Figure A for an illustration of GGM's structure and scoring.

[2] For summaries of the GGM methodology and results, see the *USAID Good Governance Matrix Final Report* for the 2006 round (dated June 2008) and the 2007 round (dated August 2009). The first report was done jointly by CeSID, Transparency Serbia and ABA/CEELI Serbia while the second report was completed by CeSID. Both reports are currently restricted to internal USAID use only.

[3] There is not a level known as "indicator" in the GGM; there are only sub-indicators. For clarity of presentation, future applications of the GGM should refer to these simply as "indicators."

Figure A - GGM Structure and Scoring

LOZNICA				3.10
Level		**Relevance Coefficient**	**Assessment Score**	**Total**
Dimension	**1 Structure**			**3.94**
Dimension	**2 Accountability**			**1.79**
Dimension	**3 Management**			**2.89**
Category	3.1 Comprehensive Budgets			5.70
Category	3.2 Transparent Budgets	1.00		3.85
Sub-Indicator	3.2.1 Public institution provides public access to budget infor	0.60		2.35
Checks	Institution provides the Ministry of Finance with the mac sector-specific assumptions to be used in developing th *Yes if there is evidence of such a document*	0.10	0	0.00
	Institution publishes the macroeconomic and sector specific assumptions of the budget in a format which is easy for users to understand *Yes if there is evidence of such a document. Maximum points if it is easy for users to understand*	0.10	0	0.00
	Institution makes information about the amount and structure of its budget widely and easily available to the public *Y if both budget and financial plan is collected. Greater score if information is available on-line, lesser if only through requests*	0.15	5	0.75
	Institution includes income/expenditure information for donor-funded projects in its fiscal reports *Y if any of it is*	0.10	7	0.70
	Institution makes budget narrative easily and widely available to the public *Y if evidence is available*	0.15	6	0.90
Sub-Indicator	3.2.2 Public institution supports citizen participation in budget preparation	0.40		1.50
Checks	Institution has mechanisms for public participation in the budget/financial plan drafting process. *Y if mechanisms/records exist. [...] Bear in mind specific deadlines in internal regulation of institution.*	0.05	6	0.30
	Institution makes supporting documentation available to potential participants before public debate/consultation *Y if documents were identified in public announcement and available for free when announcement was posted.*	0.05	2	0.10
	Institution considers budget suggestions received from beneficiaries and citizens *Y if institution accepted any or answered to all suggestions. Score reflects percentage of suggestion accepted or answered and clearly stated why the suggestion is not accep*	0.10	2	0.10
	Institution makes information about substantial changes to the budget during the budget year widely and easily available to the public *Y if the institution has visibly posted changes in the budget. Maximum points if it explained the implications*	0.10	5	0.50
	Institution makes its annual financial reports widely and easily available to the public *Y if available for 2007. Maximum score if provided in a user friendly/easy to understand format*	0.10	5	0.50
Category	3.3 Personnel Recruitment and Appraisal			2.16
Category	3.4 Ethical Code of Conduct			1.44
Category	3.5 Internal Audit Function			1.30
Dimension	**4 Open Entry/Competition**			**2.00**
Dimension	**5 Voice/Participation**			**4.85**

Sum of Sub-Indicator Relevance Coefficients, should be 1.00

Sum of Relevance Coefficients (weights) of checks

Institution Score = Average of Dimension Scores

Dimension Score = Average of Category Scores

Category Score = Sum of Sub-Indicator Scores

Sub-Indicator Scores = Sum of Total Check scores

Total Check Score = Multiplication of Relevance Coefficient and Assessment Score

Relevance Coefficients were assigned by analysts during the GGM design stage, and ranged from .04 to .25

Analyst assigns **Assessment Score** based on scoring guidance. Many checks provide limited or no guidance for assigning scores other than 0 or 10.

Note calculation error caused by manual input instead of cell formula.

The evaluation used the 2007 round as the focus for its assessment of data collection. During this round, specific dimensions or categories were assigned to an individual analyst, who developed weightings and scoring criteria for checks and sub-indicators. CeSID and its partner, Transparency Serbia, employed a total of 17 enumerators to gather information for the matrix, and five analysts were responsible for scoring at the check level based on the data. Most of the enumerators had worked with CeSID before on election-related surveys. The analysts for 2007 were specialists, with experience from the first GGM round and with working roles in CeSID, Transparency Serbia or the American Bar Association's Central European and and Eurasian Law Initiative's office in Serbia (ABA/CEELI). Enumerators received one day of training on the 25 questionnaires developed by CeSID and Transparency Serbia – the questionnaires were used for gathering information on a category-by-category basis from the target institutions. There was no formal training for the analysts, although most of them were in close, ongoing communication with the CeSID project coordinator, enumerators and each other as the scoring and analysis proceeded.

There were 21 institutions in the sampling frame for the 2007 round, including 10 central agencies, one court and 10 municipalities. Eighteen of the institutions provided at least partial data, resulting in an institutional response rate of 85.7 percent. This marked an improvement over the 66.7 % response rate in the 2006 round.

The project first sent letters on CeSID letterhead to the relevant administrative director at each institution, explaining the purpose and general process of the GGM, as well as requesting a working contact. Once a working contact was identified, enumerators set up meetings to distribute the questionnaires for response by the institution. In some instances information was not forthcoming via this method or it was not available on the institution's web site. In these cases, the GGM implementer utilized selective freedom of information requests to obtain the information (FOI requests were also used as "test applications" to gain direct information on Sub-Indicator 4.2 on FOI responsiveness). Working contacts would often identify information sources in various administrative units and further distribute questionnaires for response. Enumerators would follow up with their working contacts and attempt to answer any questions the institution had about the GGM.

After gathering information from an institution, the enumerators turned the packet of materials over to the analysts for scoring. Analysts would assign raw points at the check level, which were then weighted to produce a check score; check scores were aggregated to produce a sub-indicator score; sub-indicator scores were aggregated to produce a category score; category scores were averaged to produce a dimension score; and dimension scores were averaged to produce an institution score.

The project offered to present results to institutions participating in the 2009 round, but only the National Bank of Serbia responded positively to the offer.

For the 2007 round, CeSID engaged an independent consultant who had been closely involved with USAID's design of GGM to produce a final report summarizing the survey methodology and presenting an analysis of results from both rounds. The report thus far has not been shared outside the USG, and, with the exception of the National Bank of Serbia, participating institutions have received neither the GGM data nor the diagnostic presentations that were based on their institution's data.

3 EVALUATION OBJECTIVES AND METHODOLOGY

The purpose of this evaluation is to assist USAID/Serbia in refining the methodology of and data collection for the GGM. The evaluation also is intended to help USAID/Serbia identify approaches to releasing and publicizing future GGM results in a way that would encourage Serbia's local and national government institutions to provide better governance, including the enhanced transparency, effectiveness and efficiency of government operations. Finally, the evaluation should help the Mission determine when the next edition of the GGM should be conducted.[4]

The evaluation team specifically was asked to:

1. Assess the strength and weaknesses of the scoring in the current methodology and identify any proposed improvements, including consideration of whether the methodology gives appropriate relative weight to the individual checks and if the appropriate coefficients are used in determining scores;
2. Identify whether the methodology and scoring are applied consistently across checks, researchers, analysts and institutions, and identify any proposed improvements of their application;
3. Identify whether additional data quality assessment(s) (DQAs) should be done for future editions of the GGM methodology, and how results of such assessments should be used;
4. Provide a detailed cost estimate for future editions of the GGM, and for all related DQA;
5. Identify what procedures should be used, and by whom, to validate the methodology, data collection and analysis to ensure that future editions of the GGM are as consistent and accurate as possible;
6. Identify what additional tools, if any, would make GGM data collection, assembly and analysis more efficient, reliable and consistent;
7. Assess whether GGM researchers are adequately trained and prepared for their interaction with the participating institutions and compiling data; whether training materials are complete and appropriate; and identify what additional training and materials would be necessary. The estimated cost of any such training and materials, including any tools, should be included in the detailed cost estimate for future editions of the GGM; and
8. Advise on how the results of the GGM should be released to participating institutions and the public to encourage better performance by government institutions in Serbia.

Social Impact (SI) was selected to provide this evaluation. The methodological approach included:

[4] See Annex I, "Evaluation Statement of Work".

- Review of relevant GGM, USAID and other documents, including approximately 60 Excel spreadsheets that detailed the GGM scoring methodology and resulting data at the various analytical levels;
- Twenty-one personal interviews—including one focus group—with USAID staff, GGM analysts and enumerators, others familiar with the GGM methodology, and GGM contact points (including key informants such as Heads of Administration in municipalities as well as working contacts at staff levels) in an illustrative sample of government institutions (three municipalities and two central institutions). All interviews were conducted jointly by the team leader and senior technical adviser, with interpretation and translation assistance utilized when needed;
- Analyses of GGM scoring procedures and data, checking for consistency, patterns of missing data, etc. This involved extensive consultations with the CeSID Project Coordinator, who also has served as a GGM analyst. In addition, the evaluation team conducted independent analyses of the quantitative data to carry out consistency checks and analyses of the impact of missing data on GGM results;
- Collection and review of training materials provided to enumerators; and
- Review of the cost structure of GGM implementation, utilizing GGM cost documents and interviews with persons knowledgeable of the costs associated with GGM implementation.

This methodology has *strength* in triangulation of data among interviewees and across interview and documentary sources. It is *limited* in that:

1. Review of a large, complex data collection and analysis effort by a small team in a short period required drawing samples of institutions, interviewees and data components. Since this evaluation is largely an assessment of the quality of data gathered from a small number of institutions, carrying out informed spot-checking was as effective (and considerably more cost effective) than a comprehensive sample of all relevant institutions, persons and documents. While the sample of institutions visited was small, this sample allowed sufficient variety in experiences with GGM for the team to address the key evaluation questions and to confirm patterns of findings across data sources. We interviewed all GGM analysts and six enumerators. In the document review, we tracked GGM scoring for four institutions in seven categories to PDF facsimiles of original paper documents acquired from GGM institutions. When these facsimiles were not available, we tracked to the original, category-level questionnaires and triangulated key data quality issues with analysts and enumerators; and

2. As we detail below, some GGM documentation was not available for us to review.

The evaluation team's review of GGM data quality applied the criteria featured in USAID's Automated Directives System (ADS), Chapter 203 (Assessing and Learning).[5] It was further informed by the more detailed Data Quality Checklist, found in USAID's Performance Management Toolkit.[6]

[5] See http://www.usaid.gov/policy/ads/200/203.pdf
[6] See http://www.usaid.gov/policy/ads/200/200sbn.pdf

Field research in Serbia was carried out from October 11 to October 27, 2010. The evaluation team included James Fremming (Team Leader), Andrew Green (Senior Technical Advisor), Masa Matijasevic-Simic (Logistics Coordinator and Interpreter); and Richard Blue (SI Home Office Technical Advisor.[7]

4 FINDINGS: DATA QUALITY ASSESSMENT

The quality of GGM data, as reported in the two GGM reports, is at the heart of this evaluation. As noted in this report's section on evaluation objectives and methodology, this review utilizes the broadly applied and recognized criteria as delineated in USAID's Automated Directives System (ADS) 203.3.5 ("Data Quality"). In this context, it should be kept in mind that the ADS recommends that performance data should be as complete and consistent as management needs and resources permit. The criteria include five elements:[8]

1. **Validity.** Data should clearly and adequately represent the intended result. Another key issue is whether data reflect a bias such as interviewer bias, unrepresentative sampling, or transcription bias.
2. **Reliability.** Data should reflect stable and consistent data collection processes and analysis methods over time. The central issue is whether different analysts would come to the same conclusions if the data collection and analysis processes were repeated. USAID/Serbia and the GGM implementation team should be confident that changes in GGM data from one round to another, or differences across institutions, reflects real changes rather than variations in data collection methods.
3. **Precision.** Data should be sufficiently precise to present a fair picture of performance and enable management decision-making at the appropriate levels. One key issue is whether data are at an appropriate level of detail to influence related management decisions. A second key issue is what margin of error (the amount of variation normally expected from a given data collection process) is acceptable, given the management decisions likely to be affected.
4. **Integrity.** Data that are collected, analyzed, and reported should have established mechanisms in place to reduce the possibility that they are intentionally manipulated for political or personal reasons. Data integrity is at greatest risk of being compromised during data collection and analysis.

[7] For further reference, a list of persons interviewed and a moderator's guide for the focus group with GGM enumerators are provided in Annexes II and III.

[8] Language in this section is informed by and paraphrased from ADS 203.3.5.

5. **Timeliness.** Data should be timely enough to influence management decision-making at applicable project, program and other levels. One key issue is whether the data are available frequently enough to influence the appropriate level of management decision. A second key issue is whether data are current enough to inform management and operational actions.

The following sections present the evaluation team's summary of the status of GGM data along these criteria, based on a review of the data collection process and existing available documentation for the 2007 round.

4.1 Validity

Overall, the conceptual basis of the GGM appears to be sound at the dimension and category levels. These parts of the Matrix are readily understood and widely recognized as significant elements of good governance, and they are of practical relevance to the Serbian public sector. On the other hand, the evaluation team found a number of issues that significantly affect the GGM's validity. All of these, however, can be corrected relatively easily before the next iteration of the Matrix.

The most serious issue is ***whether the GGM is actually measuring governance***. While the overall design is conceptually sound, operationally there are problems distinguishing among:

1. A *true deficiency* in institutional performance; versus
2. A situation under which the GGM scoring criterion *does not apply* to a particular responding institution's context or operational setting (the expectation that a municipality reports to an international oversight or coordinating body, for example); versus
3. A *lack of response* by the institution to a particular data request from the GGM enumerator; this may be a matter of not answering at all, providing some but not all the documentation needed to respond to GGM questions, or answering without adequate documentation.

GGM's scoring structure is unable to distinguish between scores that truly measure institutional performance and what, for validity reasons, should be treated as missing data. This is reflected in the non-systematic treatment of zero scores in checks and of not-applicable ("N/A") scores in calculations of checks into the sub-indicator level. The effect on the GGM data is that they may, for some measurement components in the Matrix and for some institutions, represent artifacts of the data collection process itself – such as the institution's administrative efficiency in providing the required documentation or even the level of motivation of an individual clerk in responding to the enumerator's requests – rather than actual, institutional good governance. The issue of zero scores versus missing data is a common one for datasets, and the common approach is to count zeroes and not count missing data in calculations. One example of how improper handling of missing data can distort scores is the Municipality of Loznica's score for the fourth dimension, Open Entry/Competition, in which 2 of the 5 categories were non-responses: counting those category scores as zeroes generated a dimension score of 2.00, while excluding them generates a score of 3.33, a recognizable difference.

A score of zero for a *true deficiency* would be valid, but perhaps not for the other two scenarios. *Lack of response* is particularly problematic, as it may be that an institution is

simply not responding because the administration is inefficient or not motivated to respond, or perhaps because the institution does not practice that good governance aspect. Similarly, an institution may explicitly report that it does not practice a specific aspect, but even then this response may actually be one of inefficiency and low motivation.

In general, while some may argue that the failure to fully respond to GGM information requests does in fact signal poor responsiveness of the institution overall, the evaluation team finds that from a measurement standpoint it is not appropriate to "punish" an institution by counting missing data as zeroes in the scoring. "Truer" scores, in which missing data at the check or sub-indicator levels are not treated as zeroes, likely would result in substantially higher scores at the dimension and institution levels. The evaluation team was not able to calculate the actual size of this difference, or to recommend check-by-check adjustments in scoring, since this would require extensive discussion with analysts of hundreds of scoring decisions. The GGM implementer must grapple with these issues for future rounds, however, and adopt a standard approach before the GGM is conducted again.

Of the three scenarios, ***inapplicability*** is the easiest to resolve. Respondents from local government reported that not all questions were appropriate to their operations, but generally were unable to provide concrete examples due to the passage of time since the GGM was conducted. One example that was given to the evaluation team is outside oversight, which examines how an institution responds to domestic oversight bodies, donors and other international organizations, as well as civil society organizations. Local governments do not normally have official oversight from international organizations or may not even be involved in donor programs, and so these aspects of the question are not relevant to them. Similarly, for institutions at the central level, questions about public-private partnerships typically would not apply to their operations. While it is clear that some individual 'checks' may not be applicable in certain circumstances, the GGM must address such instances by not penalizing institutions with a zero score; the proper solution would be to remove the effect of that question's score from the calculation for the GGM level above it

A True Difference in Governance Practice -- Or Something Else?

The municipality of Vrsac is an example of a GGM participating institution for which scores changed between the 2006 and 2007 GGM rounds, but the reasons for this change may include a simple difference in the approach of the institution to the GGM survey.

The city's GGM score decreased from 2.23 in 2006 to 1.68 in 2007. This is the largest reduction in institutional scores among GGM municipalities. While the GGM enumerator was the same person for both rounds, the city context changed between rounds: Local elections took place shortly before the 2007 round, and many of the city's staff were newly in place. The Head of Administration assigned a new public relations officer to coordinate the response to the GGM; this was her first assigned task as a city employee. She reported to the evaluation team that while some of the city's technical offices responded well to her requests for information others had little to no interest in responding. In addition, she offered that as a new employee she felt she was not in a position to pressure department heads to respond.

The 2007 data for Vrsac contains more zero scores for checks than in 2006, suggesting that the city's reduced score may be attributable to the city's logistical arrangement for responding to GGM in the 2007 round (leading to weak documentation) rather than an actual change in city good governance practice.

Second, weakness in the ***clarity of questions*** in a questionnaire can often undermine validity of survey results; if respondents misunderstand what a question is calling for or referring to, they may provide data that "miss the target" in addressing the issue intended to be addressed by the survey designers. Of the institutional representatives we interviewed, most noted that a small minority of questions were not clear. Our interviews with enumerators confirmed the presence of this problem. The limitations of respondents' memories, as well as the large number of GGM questions overall, prohibited the evaluation team from identifying exactly which questions were unclear for respondents. Field pre-testing of questionnaires in future rounds of the GGM should adequately address this problem.

A third validity issue is the GGM's ***inconsistent approach to weighting and scoring***. While the general weighting approach (as applied with the "relevance coefficients" in the scoring; see Figure A) has been to prioritize the status of practice (or *de facto*) over that of meeting regulatory obligation (*de jure*), there was no standard scheme in the GGM for either weighting or scoring.[9] Nor was there a GGM project effort to review and align weighting and scoring schemes: analysts created their categories based on a shared, general understanding of priorities, but there was no feedback process of reviewing and confirming the approaches taken by individual analysts. Nor was there a standard scoring model applied at the check level, as some checks had what amounted to conditional or sub-check scoring; one analyst later explained that for some sub-indicators, more than the designed number of five contributing checks were identified, so the 'extra' items were compressed into a single check. The result is that some sub-indicators feature considerably greater complexity in scoring than others. This variation in complexity in itself can undermine the intended balance of importance among sub-indicators and categories, thus challenging the validity of scores within and across institutions.

Continuing with the issue of validity of scoring, a fourth problem is that the overall ***numeric results lack a conceptual reference point***. Since the overall institution score is an average of averages of sums of sums of weighted scores, it is not possible to determine the substantive significance of a difference between, say, an institutional score average of 3.8 and a score average of 4.0 (see also the related discussion of precision below). The GGM scores can be characterized as simply "high" (i.e., something relatively closer to 10 than other scores) or "low" (closer to zero), which is not precise enough to determine the actual relevance of a score. Compare this to a well-marked thermometer displaying zero degrees Celsius: we know from experience that at zero degrees water will freeze so the mark of 'zero' is a meaningful reference point. Without substantive reference points, the overall GGM scores deliver little diagnostic utility. However, there is potentially useful data, with substantive reference points, deeper within the hierarchy of the GGM, at the category and sub-indicator levels where detail in measurement is higher.

We also note that if the conceptual argument of the GGM is that the five dimensions together comprise good governance, then the dimension scores should be added, not averaged; it makes no sense to average the concepts of 'accountability' and 'management' as the basis

[9] A weighting scheme would simply apply a rule for quantitatively allocating proportions of the total weight across the weighting criteria. For example, the scheme could specify that within sub-indicators the status of practice in an institution is always weighted twice as much as that of meeting legal/regulatory obligation. The actual proportions to be applied may be selected based on GGM designers' knowledge and experience; there is no authoritative professional literature to guide such weighting. The important aspect for data validity is that the weighting scheme be done systematically across the full GGM data set and that the scheme be presented in a fully transparent fashion in reporting of GGM results.

for any conclusions about good governance. Better still would be to focus attention on dimension and category scores, rather than institution-level comparisons, in GGM reports.

Finally, the GGM originally was intended to be a contextual indicator, but the Matrix's ***set of surveyed institutions cannot be generalized to all Serbian public institutions.*** Generalizing GGM data to good governance performance at a country level, for the reasons stated below, is not justified from a statistical standpoint. This means that, under its existing sampling approach, the GGM does not provide a valid country-level context indicator. Therefore, the GGM's usefulness as a source of USAID/Serbia Performance Monitoring Plan data is contingent upon what level of generalization the Mission expects from the data. If the Mission plans to use GGM data at the Intermediate Result level and lower, and utilizes GGM data only from USAID-assisted institutions in its reporting, the GGM may be a source of valid data for such a purpose.

GGM's overall results should not be considered a proxy for broader country context in good governance performance because not all central agencies, and only a very small proportion of municipalities, are included in the sampling frame. The sampling frame of institutions in GGM (20 institutions in the full sample, with eight central agencies and 10 municipalities responding for the 2007 round) was selected because they were USAID's counterparts in various programs and/or because of the scopes of authority of the institutions meant their actions could have substantial impact on the future of Serbia. Most of the local governments are participating in current USAID local government or economic development programs, or at least had participated in such programs over the previous 10 years. Due to the statistical impossibility of (a) generalizing from a non-probability selection of central agencies to the full central government and (b) generalizing from a very small portion of municipalities selected on a non-probability basis, comparing the performance of the central agencies as a group with the municipalities as a group (as has been done in the GGM reports thus far) is therefore not statistically valid.

While the GGM is not a valid country-level indicator, however, it is potentially quite useful at the institutional level. For example, an assistance project could use it to diagnose governance strengths and limitations of the agencies or municipalities with which it is working.[10]

4.2 Reliability

Largely basing the GGM on the review of documents boosts data reliability when compared to indicators based on experts' considered opinions. Nevertheless, the evaluation team has identified some serious reliability issues. As with validity, however, the issues uncovered by the evaluation team can be remedied relatively easily before another round of GGM data collection.

First, as noted above, ***the practice of weighting and scoring varied*** by analyst. This presents a challenge to reliability as well as to validity of data. A single check in the GGM Matrix can reflect slightly different relevance coefficients (weights) across institutions and analysts. The

[10] The GGM's usefulness as a source of USAID/Serbia Performance Monitoring Plan data is contingent upon what level of generalization the Mission expects from the data. If the Mission plans to use GGM data at the Intermediate Result level and lower, and utilizes GGM data only from USAID-assisted institutions in its reporting, the GGM may be a source of valid data for such a purpose.

scoring of assessment points was also somewhat problematic, in terms of both guidance and application. For example, the guidance for scoring category #1.2 (Outside Oversight), advised scoring the institution points of 0 or 10, yet the analyst for this category assigned points ranging between these two values. Applications of scoring criteria also varied somewhat across the GGM analysts. For example, one analyst would penalize an institution's lack of documentation by assigning a zero score, while another would look past such gaps to reward interview evidence of institutional effort or activity. The inconsistent assignment of zero scores weakens data reliability and, as we have noted, validity as well.

Second, analyst ***scoring was never reviewed or cross-validated.*** Each analyst scored the categories, sub-indicators, and checks that she or he had created, and there were no inter-analyst reliability checks, or even a project-wide discussion or defense of the scores. Analysts reported that they did not even review their own scores for consistency over a period of time. Given the close connection between the creator of the scoring method and the scorer (in some instances this was the same person), it is quite conceivable that another analyst would score the same collected material very differently.

Finally, the implementer's ***data management practices were poor***. This problem extends from the archiving of documents to the calculations in spreadsheets and tables. When the evaluation team requested hard-copy documentation of the scoring paper trail, CeSID no longer had any of the paper documentation because the boxes had been thrown away during their last office move. In addition, CeSID had electronic documents (lists of collected documents and detailed scoring spreadsheets, for example) for only five of the 18 institutions responding for the 2007 round. The evaluation team was thus unable to verify the existence of hard-copy source data for the GGM, although CeSID did maintain category-specific, detailed questionnaires for each institution; this provided evidence that such documentation was held by the GGM team at some point. More seriously, category-scoring templates did not automatically calculate total check scores based on assessment scores; rather, analyst calculated the score separately and then entered the result in a cell. The scores were then entered by hand in dimension- and institution-level tables. As one might expect, the evaluation team identified many calculation errors: 49 check scores were incorrect, affecting 16 of the 18 institutions and 17 of the 25 categories, with errors being made by all analysts. Although some errors washed out against others in the same sub-indicator or were relatively insignificant, a small number of errors were significant enough to have a noticeable effect on the overall institution score, even after all the averaging and weighting.

The evaluation team carried out analyses of the quantitative impact of these internal calculation errors, summarized in Table 1.

Table 1. Original and Corrected GGM Averages for Dimensions

DIMENSION	ORIGINAL GGM AVERAGE	CORRECTED AVERAGE (ACCOUNTING FOR DATA MANAGEMENT ERRORS)11	DIFFERENCE
Central Institutions			
Structure	4.63	4.58	(.05)
Accountability	2.38	2.53	.15
Management	3.83	3.80	(.03)
Open Entry/Competition	2.10	2.19	.09
Voice/Participation	3.20	3.50	.30
Municipal Institutions			
Structure	3.01	3.16	.15
Accountability	1.36	1.46	.10
Management	3.05	3.13	.08
Open Entry/Competition	2.81	3.21	.40
Voice/Participation	2.93	3.16	.23

When averaged at the dimension level for the two levels of government, the differences attributed to data management errors are not huge; in only two dimensions (Accountability and Voice/Participation for central institutions) would a difference signify a change in score if the quantities were rounded to whole integers. Nevertheless, the errors are present in scoring for all dimensions, and additional analyses show that differences can be quite large (of more than one integer, for example) at the category and sub-indicator levels. This latter weakness is particularly important if data from the dimension or category levels of the GGM are to be used in the future to support dialogue and technical assistance with Serbian public institutions.

4.3 Precision

The use of multiple levels in the hierarchy that in turn contain multiple contributing elements in the design of the GGM was intended to provide a reasonable amount of objective information at a fine level of disaggregation. This "emphasis on numbers" was in turn designed to offer (a) the opportunity to carry out numerical analyses across the substantial breadth of good governance practices covered, and (b) protection against prospective claims by stakeholders or external observers that the project was biased or could not substantiate scores. In one sense, the GGM structure succeeds in protecting objectivity by focusing on

[11] Corrections account for effects of non-automatic calculations in spreadsheets. Inconsistent treatment of "N/A," "0" and no data in a cell is a complex issue with varying potential effects by check in the GGM; it is not included in this analysis.

documentation as the main form of evidence. In another sense, due to erratic scoring and weighting, inference and subjectivity were simply "pushed down" to the check and sub-indicator levels. In the end, the use of a large group of numerical scores, weights and averages in the GGM delivers a ***false sense of objectivity and precision***, given the findings detailed in the validity and reliability sections above.

Actual margins of error, as might be calculated classically in a probability survey, are not calculable for the GGM, since the range of values for GGM variables in the universe of all Serbian public institutions is not known. Ultimately, the preferred level of precision in GGM data should be informed by the *needs of users* of GGM data and the *costs* of developing data at various levels of precision.

Simplifying the methodology for the GGM would reduce the workload involved without sacrificing meaning. As noted above, the difference in the GGM between institutional scores of, for example, 3.8 and 4.0 is arguably meaningless. Looking at the structure of the GGM, there is little point in an assessment score of 5 versus 7, or a relevance coefficient of .10 versus .12 for a check. By the time the total check scores are summed to sub-indicators, in turn summed to categories, then averaged for dimensions, and then finally averaged for an institution, the fine distinctions made for assessment scores or relevance coefficients have all but disappeared.

This is demonstrated in Table 2, in which our analysis compares the corrected dimension scores (as shown in Table 1) with those that would result if scoring utilized a simplified, three-level scoring arrangement and an equal weighting of checks within a sub-indicator.

Table 2. Corrected and Simulated GGM Dimension Averages

DIMENSION	CORRECTED GGM AVERAGE[12]	CORRECTED AVERAGE UTILIZING A SIMPLIFIED SCORING SCHEME[13]
Central Institutions		
Structure	4.58	4.42
Accountability	2.53	2.46
Management	3.80	3.65
Open Entry/Competition	2.19	2.18
Voice/Participation	3.50	3.51
Municipal Institutions		
Structure	3.16	3.19
Accountability	1.46	1.62
Management	3.13	3.21
Open Entry/Competition	3.21	3.15
Voice/Participation	3.16	3.08

Table 2 shows how close the simplified scoring matches results from the (corrected) original scheme (note that this table continues the practice of reporting averages solely for the sake of comparison to Table 1). On the whole, the matching is quite close. Out of the 10 dimension comparisons (five dimensions compared once for each level of government) eight result in a difference of .10 or less, with the largest difference being .16. Differences again are larger at the category and sub-indicator levels. Overall, however, the simulation indicates that a simplified scoring scheme has very little effect on scores at the dimension or even institution levels of the Matrix.

The cost of obtaining the considerable existing detail of scoring at the one-decimal-point level is a greatly increased workload, and the GGM analysts noted repeatedly that ***much more work was required than anticipated***. The effort needed to analyze data for each check, decide the raw points, and then make the category spreadsheet entry was laborious. Since the major portion of the cost of conducting the GGM lies in compensation for enumerator and (especially) analyst time, this added effort translates to increased cost as well as increased propensity for error if the workload does not mesh well with the calendar for GGM deliverables.

4.4 Timeliness

[12] With scores corrected to remove effects of data management errors, as in Table 1.

[13] These are scores resulting from consolidation of the corrected scores ranging from 0 to 10 into a three-level range of 0, 5 or 10 points, and an equal weighting of checks within each sub-indicator.

The second iteration of the GGM in 2009 was based on the situation in institutions as of 2007. For participating institutions, the ***two-year lag was difficult for accurate data collection*** because some aspects are based on memories rather than documents; this is a particularly thorny problem in cases where there was a change in power in some local governments in the 2008 elections. It was also difficult for the evaluation team to verify evidence, or observations of data collected by the institutions, approximately 20 months earlier about specific items of governance in 2007. Furthermore, interviewees in GGM participating institutions offered to us that a time lag in feeding back GGM results to them would also reduce the potential utility of the GGM data for correcting governance problems within participating institutions.[14]

Respondents and analysts generally, but not unanimously, thought that the ***GGM does not need annual updating***, given the generally modest pace of change in laws, regulations, and process implementation. A biennial survey should be sufficient and a more efficient investment of resources.

4.5 Integrity

This is an area of only potential risk for the GGM. While there was ***room for personal bias*** by analysts in terms of how they created the categories and scored the checks, we found no evidence that bias was in fact a problem. Similarly, the project's ***focus on documentation, while costly, reduces the risk of inappropriate data manipulation***. Finally, the ***limited archived documentation*** as noted above does present a data integrity problem; data should be stored in a secure and reliably accessible way, with clear allocation of institutional responsibility for maintaining the data. The effects of this problem are mitigated partially by the availability of completed electronic questionnaires for each institution; nevertheless, proper archiving needs attention.

5 FINDINGS: QUALITY OF TRAINING OF GGM STAFF

CeSID used a four-page guidance document for enumerators as they prepared for data collection, and held a one-day training session for enumerators to present the approach and address enumerators' questions.[15] The data collection instruction document, while detailed and clearly presented, was nevertheless the only piece of training material used. The training did not feature role-plays applying any of the 25 GGM questionnaires. And while at one point the instructions note that "not all data relates to all institutions", there was no evidence the training alerted enumerators to the possibility of differential response across institutions, or how enumerators should respond if GGM questions were of limited or unclear relevance to a given institution or situation. Finally, while informal communication among the project coordinator, enumerators and analysts was considerable, there was no pre-testing, formal enumerator feedback to the project coordinator, systematic interim review of the data

[14] While delivery of GGM results to USAID took longer than originally expected, close communication existed between CeSID and USAID on the timeline of deliverables, and this dimension of timeliness appears not to have emerged as a problem.

[15] CeSID, "Data Collection Instructions: Good Governance Matrix Project, 2007 Phase," n.d.

collection approach, nor a formalized, post-activity review of survey implementation and opportunities for improvement.

6 CONCLUSIONS

For ready reference, the evaluation's conclusions are presented here in an order approximately matched to the order of evaluation issues identified in the Evaluation Statement of Work.

1. **The overall approach for GGM holds potential but also significant weaknesses:**

   The GGM, designed in USAID/Serbia, its implementation approach developed by a small consortium of local NGOs, and the resulting data from two administrations almost completely withheld from the participating institutions and broader Serbian society, has suffered from weak stakeholder engagement throughout its existence. Strengthened involvement by external stakeholders would reinforce both the utilization of GGM data for improved governance and the active participation by government institutions in the data collection process itself. Arguably, the technical strength of the GGM lies not in the overall institutional scores, but in the more detailed scoring done at lower levels of the Matrix. It is here that information is sufficiently detailed to support institutional diagnostics and consideration of further reform and efficiency improvements.

   Administration and reporting of the GGM should be done every two years. The attempt to collect data on an annual basis has not been successful, and in fact it is not at all clear that annual surveys are either needed or cost-effective.

   Finally, the process of selecting institutions to participate in the GGM should be reconsidered in light of (a) the challenges to date in eliciting institutional response and (b) evolving prospects for utilization of GGM data. We have noted that the GGM does not represent a reliable summary of the status of overall public governance in Serbia. The GGM may prove more relevant to USAID programs in coming years if applied to a sample of local institutions only. Regardless of the scope of forthcoming rounds, it is critically important that they feature significantly greater involvement by the GGM participating institutions in data collection planning, support to survey implementation, and use of the resulting data.

2. **GGM features strengths and weaknesses in scoring, especially regarding the extent of consistency in methodology:**

   The GGM is conceptually strong at the broader, higher levels of the Matrix hierarchy; that is, the five dimensions are well founded in the governance literature and the categories are reasonable and logical expressions of the dimensions. Overall it is a very special data collection effort, in that it combines traditional survey-style data collection with assertive, detailed documentation and analysis. There is a critical need, however, for a simpler, more readily understandable, consistent system of weights to be applied across all checks, and for the establishment of stronger reliability assurance. The complex hierarchy of scoring and averaging can be simplified with no significant loss in the resulting information.

3. **Additional data quality assessments and a preferred approach to DQA use are needed:**

This evaluation is an adequate data quality review under the current circumstances, when the GGM may be considered in transition from pilot phase to active application, with structured stakeholder involvement in refinement of administration and design. However, USAID should provide short-term training and technical assistance to ensure effective data quality assurance is built into the administration of future editions of the GGM.

4. **A structure is available to validate the methodology in the future:**

Improving data quality and sustaining it into the future is quite feasible within the existing implementation structure. The evaluation team finds that USAID's ADS data quality criteria will continue to be useful for GGM data quality assurance in the future. Focused training and technical assistance from Serbian or international experts will be required to institute new practices, install data management templates and train staff in data quality assurance practices, after which the implementing organization should have the capability to continue at a high level of quality assurance.

5. **Additional tools are needed to make GGM data collection, analysis and reporting more efficient, reliable and consistent:**

There clearly is a need to establish better data management at data collection, analysis and storage stages. Reporting in the GGM final reports has generally been of high quality, considering the data quality limitations the authors have worked with. Field pre-testing of data collection, use of written protocols and scoring cross-validation by analysts, and use of spreadsheet templates for analysis and reporting would all produce significant improvements in data collection and reporting.

6. Training to members of the GGM research team has been limited and not sufficient:

As we have noted, training of enumerators was modest and incomplete, and training of analysts was essentially absent. More effort needs to be invested by the GGM implementing organization in this area (with USAID support), with an eye toward leveraging data quality improvement innovations into training opportunities for GGM research staff.

7. Opportunities are available for sharing GGM results with institutions and the public:

If the core observation in point number 1 above regarding GGM data quality is acknowledged, the implications regarding the sharing of GGM results are straightforward. Active engagement by participating Serbian public institutions in the GGM process will by necessity call for their involvement (but not control) in identifying the approach to sharing and publicizing results. Since the GGM's strength is as a diagnostic tool, it may best be utilized within a context of quiet consultation, at least at first. The significant limitations we have noted in the technical quality of existing GGM data preclude the option of widely publicizing results of the previous two rounds. Similarly, broadly based civil society use of GGM data must await completion of a GGM round that produces coherent, reliable data.

Once higher-quality GGM data are produced, the GGM implementing organization can work with civil society organizations such as the Standing Conference of Towns and Municipalities and with the GGM participating institutions to publicize results at Standing Conference forums. In addition, the implementing organization may hold its own public release of the GGM report, in conjunction with a roundtable of commentaries by experts, civil society leaders and Government of Serbia representatives.[16] All of this would be supported by extensive release of GGM data and methodology through web sites of the implementing organization and the participating institutions.

7 RECOMMENDATIONS

7.1 Overall approach

1. **USAID and the GGM implementation team need to invest effort in more actively and substantively engaging Serbian public sector institutions in all phases of the GGM effort.** A focus on prospective users of GGM data and analyses (beginning with the current GGM participating institutions, but also moving beyond this group) is the foundation of the Matrix's improvement. These institutions are unlikely to provide direct, short- to mid-term input on improvements to GGM technical methodology, but in the longer term they would "make or break" GGM's success. A key technical barrier has been the GGM survey's less-than-clear, less-than-authoritative method of entry to the participating institutions. With institutional support at top levels, both the data collection method and the critical phase of data utilization can be matured and leveraged so that the level of technical effort provided to GGM design and administration can be matched by a commensurate level of utilization and public recognition in the good governance sphere of Serbia.

   An especially effective method of focusing institutions' attention on the need to respond to GGM information requests would be for a figure of central authority, such as the General Secretary of the Government, to issue a memorandum instructing agency heads and heads of municipalities to participate in the GGM. An approach by the implementing organization to someone at this level for support to the GGM would involve various kinds of (political and reputational) risks, so it would need to be weighed carefully. In the end, however, GGM data are unlikely ever to be "owned" by Serbian civil society or individual citizens until and unless there is a fundamental "ownership" – a basic understanding of the GGM's purpose and utility -- first on the parts of participating public institutions. Details of implementation and sharing of results should be determined through constructive, low-key dialogue with the institutions (rather than any public "naming and shaming" of institutions that display weak performance). Other potential sources of collaboration and support, such as the Standing Conference of Towns and Municipalities, should similarly be involved.

[16] The Office of EU Integration, for example, may be especially interested in being involved with GGM data utilization and follow-through, given the close relevance of GGM data to integration objectives.

2. **The GGM should be conducted biannually**, for example, in 2011, 2013, and 2015, for performance in 2010, 2012 and 2014, respectively. Annual surveys are unlikely to produce new data of sufficient significance to warrant such frequency.

3. **Refine selection of participating institutions to include only those at the local level**. National-level institutions generally are not comparable to local institutions in their operations, and it may be more meaningful to focus on governance at the local level, whether by focusing on local government exclusively, or in conjunction with central institutions that provide social services directly at the local level. The sample of institutions also should include some that have not been beneficiaries of any international assistance projects within (for example) five years of the GGM administration, so that quasi-experimental comparisons could be made, thus supporting impact evaluation.

7.2 Specific Recommendations

Please note that unless USAID is identified as recipient of these recommendations they are intended for action response on the part of the GGM implementing organization.

1. **Scoring should be improved by taking up several opportunities for simplification, consistency and overall data quality.** These include:

a) Reliability checks on scoring and documentation should be coordinated and executed by a qualified, local monitoring and evaluation (M&E) specialist who handles the spreadsheets and questionnaires. If necessary, a staff member of the partner organization could be trained in these skills by an external consultant.

b) Data quality assurance should include analyses of the differences within and between analysts' scoring, with a formalized in-house "mediation" procedure for resolving differences that may arise from cross-validation.

c) Documents supplied by participating institutions should be electronically scanned (only the first identifying pages are needed) and linked to the category scoring spreadsheet templates. This will reduce paperwork and increase efficiency in analysis and archiving.

d) To strengthen validity:
   i) Routinely conduct a field pre-test of the GGM in advance of full enumeration, and revise the data collection approach in light of pre-test results;
   ii) Remove sub-checks and conditional scoring criteria from checks; revise categories to more accurately reflect how issues relate to different types of institutions;
   iii) Design and apply a standard weighting scheme across all checks, with a simplified range of weight values; and
   iv) The GGM project should reflect the implicit priority of practice (de facto) over obligation (de jure). Whether this means that practice is 50% more important or twice as important is not the point; there simply needs to be a formalized priority to guide weighting of sub-indicators within categories.

e) To strengthen reliability: Use standard patterns of scoring values; as noted, scan documents to validate possession of them; establish and maintain a practice of cross-validation of scoring.

f) To strengthen integrity: Carry out the reliability support measures itemized above, and conduct selective analyses to check for patterns indicating bias in scoring.

g) To strengthen timeliness: Collect data for the most recent year available, not two years previous to the time of data collection. Provide feedback of results to institutions within two months of completion of enumeration.

h) To strengthen precision: Utilize decimal-point precision only at the check level; use integers at other levels.

2. **There should be additional data quality assessments and preferred approach to data quality assessment use.** Some short-term, expert guidance will likely be necessary to build the data quality assurance capacity of any implementing organization, but in the longer term, the bulk of data quality assurance should be a routine part of GGM operations. The short-term consultation should include provision of, and training in the use of, a data quality assurance work plan (including a task checklist) for future rounds of the GGM.

3.*)* **There should be specific procedures established to validate the methodology in the future.** As the GGM moves into a phase of broader engagement and active feedback and dialogue, it will become evident that validation of the methodology will increasingly be "another part of the conversation" about how to improve public governance in Serbia. Formal requirements for validation of the methodology will vary based on the GGM's loci of ownership and use. For USAID, current ADS requirements call for formal assessment of data quality within three years of commencement of reporting, but this requirement only applies to USAID data that are reported to USAID/Washington. If we focus on Serbian institutions as the prime users, the practical approach would feature a selective review of data quality by external reviewers (less ambitious than this assessment but utilizing the same quality criteria), focusing on the key issues identified in this report, within a year after the next GGM administration. Following this, routine data quality assurance practices should be in place within the GGM, and external reviews would not be needed more frequently than every second or third GGM.

4.*)* **Additional tools are needed to make GGM data collection, analysis and reporting more efficient, reliable and consistent:**

a) In selective (and rare) situations, utilize freedom of information rights as a method to reduce the workload of GGM enumerators and their working contacts in the public institutions. FOI requests may be most useful when the direct data gathering is particularly sensitive, as in information on procurement processes and oversight.

b) Apply written protocols for scoring by the GGM analysts and spreadsheet templates for category scoring, with protected layouts and calculation formulae.

c) Build summaries of institution performance, including graphs, upon cell references in spreadsheets so that summary reports can be developed with speed and accuracy.

d) USAID procurement announcements for future GGM rounds should include requirement of a detailed plan for data management, quality assurance and archiving.

e) Finally, in addition to switching from averaging to addition of dimension scores, to avoid unnecessary effort to achieve precision in scores the GGM should use ranges or bands of scores, rather than specific point scores, when reporting.. For example, the

GGM report may describe Accountability scores as “low,” “medium” or “high” instead of using numerical scores with two digits to the right of the decimal point.

5. **Adequate training of members of the GGM research team should be ensured.**

Training should be improved by:

a) Including role-plays of interviews, with attention to different situations in different kinds of responding institutions (especially national agencies compared to municipalities).

b)

b) Utilizing a pre-test of the next GGM round with a few institutions, and using an after-activity review of the pre-test to serve as a training opportunity for enumerators along with analysts.

c) Systematically collecting and documenting feedback from enumerators and analysts regarding ways in which GGM design and data collection may be improved.

d) Expanding training for enumerators to two days.

e) Adding new training for analysts, including having them work with a GGM dimension with which they are relatively unfamiliar, to enhance cross-learning among analysts and to refresh analytical perspectives.

6. **GGM results should be shared with institutions and the public.**

a) Rather than immediately publicizing institutional scores, use broader participation by civil society and the general citizenry when it can be best applied – when institutions have some minimum level of comfort with the GGM and the data it produces – so that they are prepared to respond to public input. In the next round of the GGM, make the circle of engagement much broader, utilizing informal technical consultations with institutions first, then exposure via professional forums such as the Standing Conference, and soon thereafter sharing with the public via web site summaries and town hall meetings.

b) To the extent feasible, coordinate with GGM participating institutions and senior leadership in the Government of Serbia in refining the approach to broader public sharing of GGM results. The more that the GGM may come to symbolize an attempt by USAID or by Serbian civil society organizations to be the “cops on the street,” monitoring the quality of governance, the less likely it is for Serbian leaders to be able to comfortably make a case for constructively participating in the GGM data collection process and utilizing GGM data as a basis for institutional improvements.

8 COST ANALYSIS

In Table 3 below, we present a proposed breakdown of estimated costs for a future round of GGM data collection, incorporating the recommendations offered in the section above. The estimated cost figures are informed by discussion with sources at the current implementing organization, as well as general data collection and project management experience of the evaluation team.

Table 3. Estimation of Costs of a Future GGM Round with Methodological Enhancements

ITEM	QUANTITY	ESTIMATED AMOUNT ($)
1. Data collection and assembly	Enumerator days: 12 days x 17 enumerators @ $ 120 per day (includes two days per enumerator for field pre-testing and follow-up)	24,480
2. Analysis	Analysts: 12 days x 7 analysts @ $ 280 per day (includes two days per analyst for work with consultants on revising scoring design)	23,520
3. Refine and validate methodology (including technical assistance, training and consultations with stakeholders)	Local consultants: 30 person-days @ $ 350 per day	10,500
4. Conduct field pre-test of revised GGM data collection methodology	Local consultants: 10 person-days @ $ 350 per day	3,500
4. Presentation of results to institutions	–Analysts: 2 x 10 days @ $ 280 per day –Local consultants for preparation of materials: 2 x 5 days @ $350 per day	5,600 3,500
5. Preparation of reports	–1 consultant: 5 days @ $ 350 per day –Editing, materials and production	1,750 1,000
6. Publicizing of results	–Analysts: 2 x 4 days @ $ 280 per day –Editing, materials and production	2240 400
1. Project management	1 manager	7000
2. Administrative support	Staff time and ancillary materials	3,800
3. Travel: o Field pre-test to 3 municipalities o Enumerators in municipalities o Meals and refreshments o Analysts for municipality	3 municipalities @ $ 140 each 20 municipalities @ $ 140 each	420 2,800 1,000 500

ITEM	QUANTITY	ESTIMATED AMOUNT ($)
presentations		
TOTAL		**$ 92,010**

Assumptions included in this analysis are that:

- The sample of the GGM would be revised to include twenty municipalities and no central institutions;
- The field data collection would include a total of 40 person-days of training and oversight by the external consultants, including guidance with the field pre-test and after-activity review of the pre-test; and
- Enumerator and analyst daily rates would not substantially increase from those applied in the 2009 round of the GGM.

The total estimated cost here is 23 percent greater than that of the $75,000 budget of the 2009 round of the GGM. The funding level for the most recent round was seen by the implementing organization as adequate, even though the time required for data collection and analysis was larger than anticipated. While the proposed approach includes modest design revision and a "beefing up" of data quality control, it is also expected to deliver increased efficiencies in collection and analysis.

USAID
FROM THE AMERICAN PEOPLE

EVALUATION OF THE SERBIA GOOD GOVERNANCE MATRIX - ANNEXES

DECEMBER 2010

This publication was produced for review by the United States Agency for International Development. It was prepared by Social Impact, Inc..

EVALUATION OF THE SERBIA GOOD GOVERNANCE MATRIX - ANNEXES

DECEMBER 2010

James Fremming, Team Leader
Andrew Green, Senior Technical Adviser

Social Impact, Inc.

2300 Clarendon Blvd, Suite 300
Arlington, VA 22201

Contracted No.: RAN-I-00-09-00019

USAID/Lebanon PMPL Project
Beirut, Lebanon

Table of Contents

ANNEX I. EVALUATION STATEMENT OF WORK

(Text begins on following page)

ATTACHMENT 1

STATEMENT OF WORK

GOOD GOVERNANCE MATRIX EVALUATION

I. TITLE OF THE ACTIVITY

"Good Governance Matrix Evaluation"

II. PURPOSE

The purpose of this evaluation is to assist USAID/Serbia in refining the methodology of and data collection for the Good Governance Matrix (GGM) now and in the longer term. The evaluation should also help USAID/Serbia determine how releasing and publicizing future GGM results can encourage Serbia's local and national government institutions to provide better governance, and enhance the transparency, effectiveness and efficiency of government operations.

III. BACKGROUND

The Good Governance Matrix was developed by USAID/Serbia to track the performance of selected government agencies as a tool to encourage better governance by Serbian institutions. It currently includes 240 "checks," selected to determine if the institutions being examined (1) comply with specific provisions of Serbian laws, and/or (2) comply with specific international standards of good governance.

The Good Governance Matrix has now been performed twice: once in 2008 to examine data from 2006 (the "2006 edition")[1], and once in 2009 to examine data from 2007 (the "2007 edition")[2]. The detailed methodology for scoring the checks, attached as Attachment VIII, was developed by a research team, led by the Center for Free Elections and Democracy (CeSID), a Serbian non-governmental organization (NGO). Another Serbian NGO, Transparency Serbia, worked closely with CeSID on both GGM editions. The Serbia office of the American Bar Association's Central European and Eurasian Law Initiative (ABA-CEELI) also contributed to the first edition. The CeSID team was chosen by the Mission to complete the GGM methodology and gather and analyze data through a purchase order, following a full and open competition for a local implementer.

For the 2006 and 2007 editions, CeSID and Transparency Serbia trained a team of researchers, who then performed interviews of government workers and examined official records to see how

[1] See **Attachment VI:** USAID Good Governance Matrix Final Report, June 2008
Center for Democracy and Free Elections (CeSID) Transparency Serbia
American Bar Association Central Europe and Eurasian Law Initiative Serbia

2 See **Attachment VII:** USAID Good Governance Matrix Report, August 2009
Center for Democracy and Free Elections (CeSID)

selected entities were performing against each of the 240 checks. Most of the elements of the GGM were drafted directly by USAID/Serbia, although some have been modified based on advice and suggestions from the teams performing the checks and reporting the results. The GGM is an evidence-based tool since researchers check agency records for information, instead of relying solely on the opinions of officials or other experts to determine if institutions are meeting their obligations.

The 2006 edition attempted to examine the operations of 10 municipalities and 11 republic-level agencies. These institutions were chosen by USAID/Serbia because they were USAID's counterparts on other programs and because their missions could have substantial impact on Serbia's future. For 2006, 14 of the 21 institutions responded to the survey, at least to some extent. For 2007, 18 of the same 21 institutions responded. This was an increase over the previous survey, although some institutions failed to respond even after formal requests for the relevant information were made pursuant to Serbia's freedom of information law, the Law on Access to Information of Public Importance.

To date the results of the GGM have not been released publicly, since USAID considered the 2006 and 2007 to be field tests of the GGM. This evaluation activity would be conducted before the next edition of the GGM, to validate the GGM methodology and system of data collection and analysis, make any needed refinements and adjustments, and establish procedures for the methodology, data collection and analysis to be adjusted as needed in the future.

IV. RELATIONSHIP TO USAID/SERBIA'S STRATEGY

The Good Governance Matrix Evaluation will contribute to several of USAID/Serbia's Assistance Objectives and Intermediate Results (IRs).

Through its Amended Mission Strategy (2006- 2015), USAID/Serbia has refocused its democracy program to reflect the country's current commitment to a pro-EU course, while at the same time addressing key areas where democratic development has stalled. The amended strategy features a single Assistance Objective for the democracy program.

The GGM links to the Mission Strategy and feeds into its Results and are supported through implementation of the matrix to identify areas for improvement of accountability, responsiveness, and transparency in participating government institutions and by the participating agencies implementing changes based on the findings of the GGM. Several measures within the GGM examine the provision of government services as well as government's responsiveness to citizens

V. EVALUATION ACTIVITIES

The contractor shall complete a Good Governance Matrix analysis and evaluation based on the 2006 and 2007 editions to help the Mission determine the best means of publicizing GGM results to stimulate USAID implementers and government counterparts to improve their performance. The evaluation should also help the Mission determine when the next edition of the GGM should be conducted.

There are eight main expected results of assistance which will be rendered through this evaluation:

1. Assess the strength and weaknesses of the scoring in the current methodology and identify any proposed improvements, including consideration of whether the methodology gives appropriate relative weight to the individual checks and are the appropriate co-efficient used in determining scores;

2. Identify whether the methodology and scoring applied consistently across checks, researchers, analysts and institutions, and identify any proposed improvements of their application;

3. Identify whether additional data quality assessment(s) should be done for future editions of the GGM, methodology, and how the results of such assessments should be used;

4. Provide a detailed cost estimates for future editions of the GGM, and for all related data quality assessment(s);

5. Identify what procedures should be used, and by whom, to validate the methodology, data collection and analysis to ensure that future editions of the GGM are as consistent and accurate as possible;

6. Identify what additional tools, if any, would make GGM data collection, assembly and analysis more efficient, reliable and consistent; and

7. Assess whether GGM researchers are adequately trained and prepared for their interactions with the participating institutions and compiling data; whether training materials are complete, appropriate; and, identify what additional training and materials would be necessary. The estimated cost of any such training, materials, including tools, if any should be included in detailed cost estimate for future editions of the GGM.;

8. Advise on how the results of the GGM should be released to participating institutions and the public to encourage better performance by government institutions in Serbia.

VI. DELIVERABLES

The Contractor shall deliver the followings:

1. The evaluation team should conduct an out briefing to USAID/Serbia covering its findings and recommendations and present a draft of its evaluation report and cost estimate for future editions of the GGM at the conclusion of its field work in Serbia.

2. An evaluation report addressing all eight of the points specified in the above summary of Evaluation Activities. The final report format will be discussed upon the team's arrival in Belgrade.

As part of the final Evaluation Report the Contractor should present:

a. Cost estimates for future editions of the GGM, including data collection, assembly, analysis, preparation of reports, presentations of results, validation of methodology

for data quality assessment, and the training of researchers (including the preparation and distribution of training materials);

b. Cost estimates for publicizing the results of future editions of the GGM.

VII. PERIOD OF PERFORMANCE

This evaluation is expected to last a total of eight weeks beginning o/a October, 2010 and ending o/a November 2010. The period of performance for this Task Order is approximately eight weeks. However, the evaluation work itself may not exceed five weeks according to the timeline below.

5 week Timeline:

5 days of preparation time including travel
15 days of field work
5 days drafting the report
2 days revising and submitting the final report after USAID/Serbia's comments

Total: 27 days

USAID/Serbia will have 15 working days to review the draft report, cost estimates and to provide comments and suggested revisions. Once USAID feedback has been provided the contractor will have two business days to submit a final version of the report and cost estimate to Task Order COTR.

VIII. AUTHORIZED GEOGRPHIC CODE

The authorized geographic code for procurement of goods and services for the resulting Task Order awarded as a result of the RFTOP is 000 (US) and proposals should comply therewith. However, local procurement is authorized in accordance with and subject to the limitations contained in 22 CFR 228.

IX. AUTHORIZED WORKWEEK

Short term expatriate personnel are authorized to work six-day workweek.

ANNEX II. PERSONS INTERVIEWED

Person Interviewed	Organizational Affiliation or GGM Role
Ellen Kelly	Democracy Officer and Senior Rule of Law Adviser, USAID/Serbia
Milos Mojsilovic	Project Director, CeSID
Miodrag Bogdanovic	M&E Specialist, Program Strategy and Coordination Office, USAID/Serbia
Nenad Moslavac	Project Management Specialist, Economic Growth Office, USAID/Serbia
Sanja Nikolin	New Business and Consulting Director, Multi Servis; GGM designer and report writer
Danilo Pejovic	Analyst, Transparency Serbia
Rade Djuric	Independent consultant; GGM enumerator
Violeta Veselinov	GGM enumerator
Alexsandar Simoncic	GGM enumerator
Nemanja Nenadic	Analyst, Transparency Serbia; GGM analyst
Milos Stanojcic	Head of Administration, City of Sabac
Ljiljana Nikolic	Head of Local Economic Development, City of Loznica
MilojkaSmiljanic	Head of Administration, City of Loznica
Miroslav Mitrovic	Head of Finance, City of Loznica
Marija Kulic	Head of Public Relations, City of Vrsac
Stanojla Mandic	Commissioner for Information of Public Importance and Personal Data Protection, Republic of Serbia
Milica Markovic	Program Specialist, Unit for Organization, Analysis and Improvement of Local Self-Government System, Ministry of Public Administration and Local Self-Government, Republic of Serbia
Srdan Majstorovic	Deputy Director, Office of EU Integration, Republic of Serbia
Gordana Petkovic	Secretary General, National Bank of Serbia
Mr. Radovic	Manager, Department of Administration, National Bank of Serbia

ANNEX III. ENUMERATORS FOCUS GROUP: MODERATOR'S GUIDE

Welcome [put welcome greeting on flip chart if we have one; invite people to put first names on tents with marker]

Introductions (ask for names and roles in GGM; signup sheet)

Purpose of this discussion (Distribute session agenda and refer to purpose as shown there):

Suggested ground rules (basic principles for how we will have this discussion):

- We need to hear of your experiences and opinions. Please do not be shy about speaking up. I will try to ensure that everyone has a chance to share his or her point of view.
- All opinions are equal. If someone's opinion is different from yours please point this out. We expect that everyone will maintain a respectful attitude in the discussion.
- Your opinions are confidential within this group. Please do not pass on any substance of this discussion to anyone else. We will be summarizing the results of this discussion in our report to USAID, which in time will be made publicly available. Your name will be included among the list of many information sources for our evaluation.
- I will guide the discussion to have it focus on certain topics listed in the agenda. If you have a comment about some other related topic, just let me know and I'll make a note for us to come back to that at an appropriate time.
- We plan to finish by 530p.

Do you have any questions about our ground rules?

…If not then let's get started.

1. How did you come to be an enumerator for GGM? **[Probe a bit for relevant previous experience.]**

2. Could you describe in particular the things you did as enumerator? What steps did you take to get the job done? **[Look for key steps and differentiation among enumerators]**

3. How did you make the initial contact with the GGM participating institution? Was this actually the central point of contact for your GGM work with this institution? Why or why not?

4. How did you identify the person or persons you needed to interview?

5. How many interviews did you carry out? Was this the right number, too many, too few?

6. Were you able to actually complete all the interviews you needed?

7. Were the questions in the questionnaires clearly understood by your information sources in the institutions? **[Probe: Any pre-testing of questionnaires?]**

8. If an interviewee did not understand what information you needed, what did you do?

9. If an interviewee refused to provide information about something, what did you do?

10. If an institution answered with delays or with various logistical explanations for not responding, what did you do?

11. How easy was it to arrange the time and place for conducting the interviews you needed to carry out?

12. Did you find that you were adequately prepared to carry out the enumeration? **(Probe: interviewing; understanding the questionnaire content and institutional context; organizing and recording information)**

13. Would you have benefitted from more preparation time or access to more preparatory materials?

14. What process did you use to organize the data collection? Did you have separate paper folders for each dimension, for example? Did you work sequentially from one dimension to another, or were you collecting information across dimensions as you went?

15. Could you please describe the training you received from CeSID to prepare you for GGM enumeration? What happened in the training? How many hours was the training? What materials were used in the training? Was it completely in a group setting, or did it include any individualized training?

16. What suggestions do you have for improving the training process or training materials?

17. How many of you were enumerators for the 2006 GGM (first round)? **[If none, skip this question.]** In what ways was your experience as enumerator different in the second round as compared with the first? Why?

18. Overall, how would you describe the attitude of your GGM institution(s) about the GGM data collection? **(Cooperative, reluctant, refusing, apprehensive…)**

19. What has been the strongest or best aspect of the GGM? Why?

20. What has been the weakest or most limiting aspect? Why?

21. What are your suggestions for improving the GGM and its data collection?