

Collaborating to improve the measurement of results from support for Governance and Public Sector Management reforms

A Discussion Note
Draft: October 17, 2012



For further details, contact the Governance and Public Sector Management Department, Poverty Reduction and Economic Management, World Bank:

Measuring governance results:

Robert Beschel

rbeschel@worldbank.org

Francesca Recanatini

frecanatini@worldbank.org

Measuring public sector management results:

Vivek Srivastava

vsrivastava@worldbank.org

Jurgen Blum

jblum@worldbank.org

Overall work program:

Nick Manning

nmanning@worldbank.org

Acronyms

ACA	Anti-Corruption Authorities
AGIs	Actionable Governance Indicators
CSIs	Core Sector Indicators
DfID	UK Department for International Development
GPSM	Governance and Public Sector Management
ICR	Implementation Completion Report
IDA	International Development Association
IE	Impact Evaluation
ISPMS	Indicators of the strength of Public Management Systems
ISR	Implementation Status Report
PAM	Public Accountability Systems
PSM	Public Sector Management
SIEF	Strategic Impact Evaluation Fund

Contents

1. Introduction.....	4
2. Indicators of Country Governance and Public Management Systems	6
Harnessing existing cross country comparative data	6
Identifying new country-level metrics.....	8
Uses of ISPMS and AGIs.....	8
3. Extracting data from Bank projects	10
Harnessing Existing Project Evaluations	10
New Impact Evaluation research on GPSM interventions.....	10
Building research into projects.....	12
4. Qualitative Results Stories.....	13
Short but robust stories of achievement.....	14
Capturing and systematizing practitioner knowledge.....	15
5. Opportunities to Work with the Bank	15
Summarizing the main areas of potential collaboration	15
Testing the relevance of available ISPMS/AGI data	16
Development of New ISPMS/AGIs	16
Collaboration in better learning from projects	17
Testing the robustness of the Corporate Scorecard approach	17
Developing a systematic program of Impact Evaluation on GPSM interventions	18
Testing Core Sector Indicators (CSIs) in other settings	18
Capturing Tacit Practitioner Knowledge.....	18
Moving from short stories to robust case studies	18
Joint help desks and other approaches for capturing practitioner knowledge	18

Boxes

Box 1: Principles of good GPSM diagnostic work.....	5
Box 2: What are “public management systems”?	6
Box 3: What indicators might be “incubated”?	8
Box 4: Early GPSM candidates for formal Impact Evaluation studies within the Bank.....	11
Box 5: Early candidates for GPSM Core Sector Indicators in Bank projects.....	13

Tables

Table 1: Criteria for assessing the “utility” and “feasibility” of GPSM indicators	7
Table 2: Targets, Rankings and Intelligence Compared	9
Table 3: Selected achievement stories for the IDA 16 Mid-Term Review	14

Figures

Figure 1: The Public Sector Results Chain	5
Figure 2: Measurable performance improvements in public management systems supported by Bank projects under implementation during fiscal years 2010-12.....	10
Figure 3: Schematic overview of the Multi-Donor Program	17

1. Introduction

The World Bank is a major player in supporting public sector management reforms. The Bank has a large governance and public sector management (GPSM) lending portfolio, committing some US\$3.5 billion to this area of work each year. Governance and public sector management reforms are integral to all its work, whether in education, agriculture or transport. It has a large and highly qualified body of staff working on GPSM, who track major developments within the field. *Can it do better? Through targeting of feasible and effective GPSM reforms, can it make a bigger and faster difference to the lives of the poor and those who depend on public sector services?*

The challenge here is that there is relatively little hard evidence about what matters most in improving GPSM. The connection between institutions and accountability on the one hand, and social and economic development on the other, is evident to any experienced government official or practitioner - but hard to pin down precisely. The institutions that have been shown to matter for economic development are largely those that protect the returns to private investment, in particular property rights and the rule of law. But equivalent evidence is lacking for those institutions centrally involved for example in managing public sector financial and human resources and for the broader governance environment within which the public sector operates. Lacking robust empirical evidence, practitioners rely on a strong body of tacit or “craft” knowledge to develop practical reform strategies.

The Bank has extensive implicit knowledge on GPSM reform, but explicit research is lagging. Currently, GPSM specialists bring a significant and powerful body of tacit knowledge on reform to bear on complex, multi-dimensional problems. However, the explicit theory and evidence base for understanding what works and why in public sector reform remains strikingly limited compared with other policy areas.

While the Bank and other donors have made progress in collecting data on the strength of public financial management systems, such data are still lacking in other key GPSM areas. In the field of public financial management, the success story of the multi-donor [Public Expenditure and Financial Accountability](#) (PEFA) tool, developed in 2004 and today applied in over 110 countries, has highlighted the promise of such data for learning and debate in client countries, among donors and in research on GPSM reform. PEFA has demonstrated that it is possible to track behavioral changes resulting from PFM reforms and has fueled new research. To date, no equivalent data are available for other institutional areas, such as civil service or tax systems or for the non-executive accountability institutions. This note proposes a way forward for the development of Indicators of the Strength of Public Management Systems (ISPMS) for the former and Actionable Governance Indicators (AGI) for the latter.

The Bank has not fully reaped the potential for learning from its GPSM project portfolio. In principle, the Bank's projects potentially provide a rich source for learning “what works” in GPSM reform. However, recent portfolio reviews have shown that such learning (e.g. from the Bank's own project evaluations) is often restricted to a focus on a project's performance against its own (often idiosyncratic) indicators. These reviews generally do not take the extra step to systematically understand the underlying causes that explain why results were achieved or not. Thus, currently, much of the knowledge generated by such projects remains tacit and unshared - and many opportunities for systematic learning, such as building research into project design, remain unexploited.

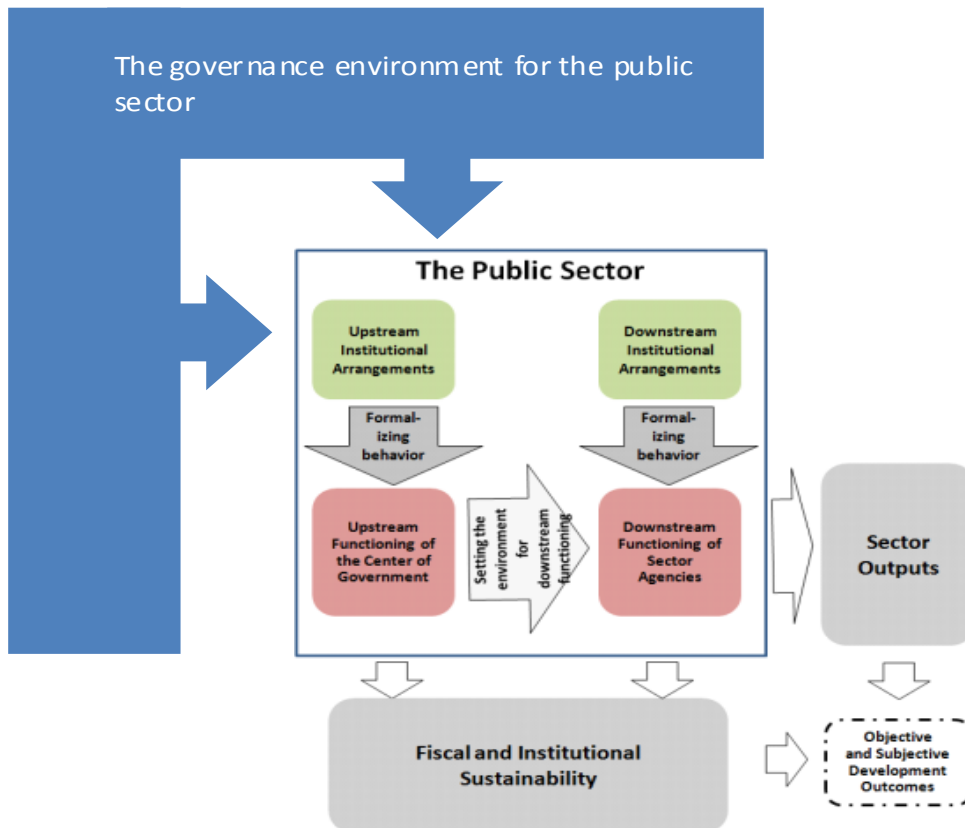
The Bank's [Public Sector Management Approach \(PSM\) for 2011-2020](#) (PSM Approach) has set a direction towards balancing the Bank's tacit understanding of GPSM with “scientific” knowledge. Investing in learning about what drives results in public sector reform is fundamental to enabling the Bank and its clients to make better-informed reform decisions in the future. Such learning investments include deepening and broadening the metrics available for measuring the strength of country institutions, eliciting more powerful learning from Bank projects, and showing leadership in developing a research agenda on GPSM including more rigorous qualitative and quantitative research on reform impacts. The PSM Approach locates knowledge generated by such learning investments in designing new projects, by encouraging Bank project managers to follow a “diagnostic protocol” (see Box 1) that carefully draws on available evidence.

Box 1: Principles of good GPSM diagnostic work

1. Focus on the functional problem, rather than the solution – the key question is “what is not working and why?”, not “what should this look like?”
2. Engage stakeholders to take advantage of local knowledge and “insider” information to identify functional problems – *set out the likely binding constraints and the potential mitigators*
3. Use Political economy analysis prospectively – develop scenarios that fuel debate about the likely outcomes, with and without Bank actions
4. Use available evidence and accepted theory to make the case that a given reform will fix the functional problem – *spell out the assumed theory of change*.

Source: PSM Approach

Figure 1: The Public Sector Results Chain



2. Indicators of Country Governance and Public Management Systems

The PSM Approach calls for a major push towards improving country-level tracking of public sector institutions and their functioning. World Bank economic and human development data fuel research and learning around the world, but currently the Bank plays only a limited role as a provider of data on public sector institutions. Looking into the “black box” of the results-chain is crucial for understanding where the causal links are broken (Figure 1).

ISPMS metrics, such as those provided by PEFA, can provide powerful clues about the origins of service delivery failures downstream, deepening understanding of sector-specific governance and public management failures. Better data on the strength of public management systems (Box 2) will be also be fundamental for maximizing the impact of the new [Program for Results](#) (PforR) financing instrument, which relies on disbursement-linked indicators of reform progress. Such data are also enabling more comprehensive reporting within the Bank’s [Corporate Scorecard](#).¹

Box 2: What are “public management systems”?

“Public management systems” are, in essence, the key management and oversight responsibilities of the core (“upstream”) ministries and agencies at the center of government² which have functions that cut across sectors and are broadly seen to matter for “downstream” public sector results and development outcomes. Most would agree that these functions include budgetary and financial management systems, procurement and revenue mobilization systems, and public administration, but whether this is exhaustive and the exact dimensions of these functions could be open to considerable debate.³ The idea behind defining these upstream management and oversight arrangements as “public management systems” is to create a comprehensive map of the center of government operations that are broadly seen to matter for public sector results and development outcomes. The term “public management system” is meant to provide a common frame of reference that governments, donors and researchers can use both in analytic work and country dialogue.

Similarly, metrics on the [governance environment for the public sector](#) (known as AGIs or actionable governance indicators) measure the quality of systems through which the executive is held to account by political and non-executive institutions and directly by the public.⁴

Harnessing existing cross country comparative data

GPSM reforms are often thought of as changes to the formal (*de jure*) responsibilities of oversight agencies (ombudsman, parliamentary public accounts committees, and anticorruption agencies etc.)

¹ The World Bank’s Corporate Scorecard is similar in structure to DfID’s results framework with the 4 “Tiers” in the Corporate Scorecard corresponding to the 4 “Levels” in DfID’s result framework.

² Such bodies include the Ministry of Finance and the offices that support the head of government.

³ For reasons of practicality and Bank’s traditional comparative advantage, the management and oversight responsibilities of non-executive state institutions such as judiciaries, legislatures and institutions such as Supreme Audit Institutions are not covered by this term. These are, however, included in the broader governance environment.

⁴ These would include judiciaries, legislatures and institutions such as ombudsmen and Supreme Audit Institutions.

and in the institutional and managerial arrangements in the center of government and in sector agencies, such as new civil service laws or budgetary procedures, revised funding arrangements for health care etc. Changes to formal arrangements are often critical, but ultimately GPSM reform is about changing the informal *de facto* behaviors of agents acting on and within the public sector. Changing these actual behaviors does not necessarily commence with legal or other formal reforms – changes in how agencies and departments function day-to-day can provide the springboard for more formal changes in the laws and procedures. Poor public sector performance can often be traced to the lack of implementation of formal rules. For example, many countries have meritocratic employment regimes *de jure* – but *de facto*, these often remain only partially enforced and provide insufficient protection against patronage or the sale of public posts.

The Bank is currently leading an effort to develop a set of Indicators of the Strength of Public Management Systems (ISPMS) and [Actionable Governance Indicators](#) (AGIs) from existing sources to be supplemented with new indicators where such indicators are not available. ISPMS measures the strength of public sector management systems by capturing the intermediate results in the middle of the public sector results chain in Figure 1. AGIs capture actual improvements in the strength of non-executive institutions of accountability (the “governance environment” in Figure 1).

The threshold for inclusion in ISPMS and AGI datasets is that the indicator captures changes in the performance of the system or in the behavior of the public agents working within it. Indicators are tested against “utility” and “feasibility” criteria (see Table 1 and below).

Table 1: Criteria for assessing the “utility” and “feasibility” of GPSM indicators

Utility criteria	Indicators of the strength of public management systems (ISPMS), measure features of public management which...	Actionable governance indicators (AGIs) measure features of the governance environment which...
Action-worthy	Make a difference for development outcomes, and/or inform donor decision-making	Make a difference for development outcomes <i>or</i> are themselves widely seen as an intrinsically valuable outcome and/or inform donor decision-making
Actionable	Are specific enough to point governments towards policy actions that they can take to achieve an improvement and monitoring reform progress	
Behavioral	Capture the functioning or performance of public institutions, avoiding the fashion trap of best practices which encourage mimicry of specific legal, organizational or institutional forms	
Replicable	Can be measured more or less objectively	
Feasibility criteria	ISPMS and AGI	
Sufficient scope	Available for a minimum of 20 countries	
Affordable	Collected repeatedly at regular intervals	

The utility criteria for ISPMS and AGIs differ significantly in relation to “action-worthiness”: the potential link with development outcomes. AGIs may identify important areas for work, transparency for

example, where there is a wide agreement that it has an intrinsic value and so its utility outweighs any limitations to the evidence available concerning its instrumental value. For ISPMS indicators, the assumed link to development outcome is always critical.⁵

In addition to the “utility” criteria, for inclusion of indicators as ISPMS/AGIs, two “feasibility” criteria must also be met. The indicators need to be available for a minimum of 20 countries and should be collected repeatedly at regular intervals. These criteria are to ensure sustainability given the cost of data collection. A number of datasets that could potentially contribute very usefully to the proposed work do not yet meet these pragmatic feasibility thresholds. An important consideration for new indicators would be the cost and feasibility of regular data collection.

Identifying new country-level metrics

There are likely to be areas where measuring progress is widely regarded as important, but where the exact nature of any metrics may not be immediately apparent. The Bank is establishing an “incubator function” to develop and test new dimensions of measurement for governance or public management systems. This incubator function will assess the practicality, feasibility and utility of developing new measures in emerging areas (Box 3).

Uses of ISPMS and AGIs

In line with the Busan Aid Effectiveness Agenda, ISPMS and AGIs, at the cross-country and country levels, can be used by governments, donors and researchers wishing to support a dialogue based around the question “what tends to work in general?” This is a powerful point of entry, although it does not answer the question “will it work here?” That question requires much more detailed diagnostics, with finer-grain measures, tailored to the needs of the respective country.

- *At the cross-country level*, ISPMS and AGIs can provide an entry point for dialogue with client governments and inform learning and research.⁶ As the example of PEFA shows, comparable measures of PSM systems can provide powerful levers for engaging governments in a constructive dialogue on their PSM reform agenda. While “*ranking*” countries on indicators can be helpful to point out where they stand compared to peers, it brings distinctive gaming risks with it (see Table 2). In addition, they can assist in building

Box 3: What indicators might be “incubated”?

The World Bank has been examining selected areas where the best approach to measurement is not yet evident, such as the implementation of public accountability systems (PAM) or the effectiveness of anti-corruption authorities. The World Bank has developed a new approach to measure the implementation of PAM systems and it is currently testing it in a few selected countries. Similarly, the World Bank has been exploring the role of anti-corruption authorities (ACAs) and how to measure their effectiveness (and the effectiveness of similar check and balance institutions). This has led to the development of a framework to evaluate ACAs and to the identification of selected countries where this framework could be tested.

⁵ The need for this linkage is at the root of the, now well-accepted, aversion to best practices in public management.

⁶ Building on existing ISPMS and other country level indicators, the World Bank is for example developing “Public Sector Performance at a Glance Scorecards” that can serve as a rapid diagnostic tool for countries in the ECA region, informing more in-depth diagnostics.

up a picture of how public management systems, on average, evolve over time, for example in countries of a certain income level or from a certain region. And they can provide the empirical basis for cross-country learning (“intelligence”), for example about which institutional arrangements result in better performance than others in the average country.

- *At the country level*, ISPMS and AGIs can assist governments and donors in tracking reform progress over time and against “targets” set for tracking reform impact. They can be an important vehicle for strengthening government ownership and for improving project effectiveness by providing a measure against which progress can be assessed and the need for strategy correction identified. They can to some extent inform *diagnostics* (or “intelligence”) in a particular country, by showing whether a country does well or poorly on an indicator, compared to other countries.

It is important to note that while using ISPMS and AGI for these ends can often significantly enhance public sector performance, using indicators can also weaken performance if they create unhelpful incentives (Table 2).

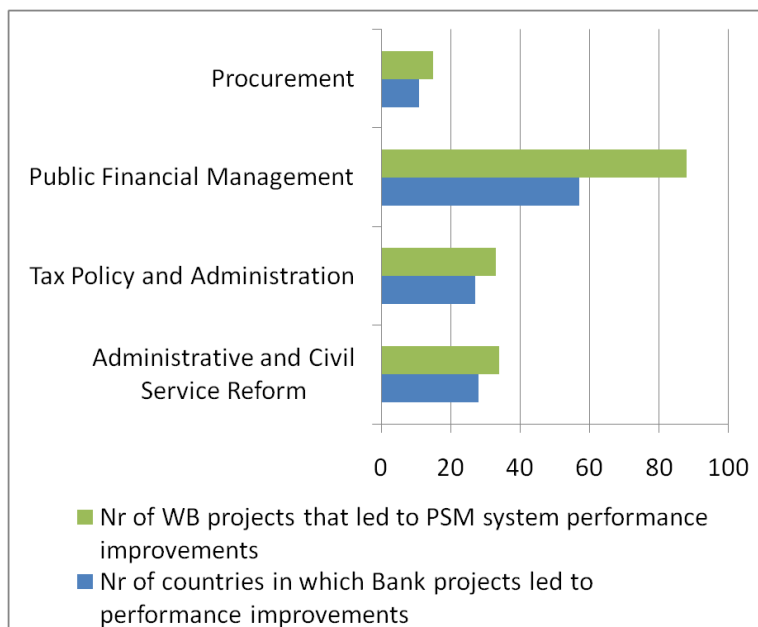
Table 2: Targets, Rankings and Intelligence Compared

Type of Application / Use	What it Involves	How it aims to enhance performance	How it can obstruct performance
Targets	Using numbers to set and monitor minimum thresholds of performance	Concentrate attention on improving performance in a limited number of priority areas	Can produce ratchet effects, threshold effects, output distortions
Rankings	Using numbers to compare performance of different units	Encourage “sweating and stretching” to raise overall performance, avoiding ratchet effects by focusing on relative performance among rivals	Can produce threshold effects (where ranking is categorical) and output distortions
Intelligence	Using numbers as background information for choice by users or for policy change or management intervention	Encourage informed choice or developing learning capacity and diagnostic power by adding knowledge about performance, avoiding ratchet effects, threshold effects, and output distortion from gaming behavior	Can produce ambiguity, complexity, and fragility and may be ignored by key players, especially service users

Source: Based on Hood, Christopher (2012), “Public Management by Numbers as a Performance-Enhancing Drug: Two Hypotheses”, *Public Administration Review*, forthcoming.⁷

⁷ Hood suggested that the magnitude of these gaming risks is significantly shaped by the culture of those being measured: “hierarchist”, “egalitarian”, “individualist”, or “fatalist”

Figure 2: Measurable performance improvements in public management systems supported by Bank projects under implementation during fiscal years 2010-12



Source: [World Bank Corporate Scorecard](#)

3. Extracting data from Bank projects

Harnessing Existing Project Evaluations

The Bank is now using the utility criteria to identify measures within its own project data which measure the strengthening of public management systems as a result of World Bank interventions. The annual results are included in Tier II Institutions and Governance indicators in the [World Bank's Corporate Scorecard](#) (Figure 2).⁸ Indicators are selected from Bank reviewed reports on progress at the project component level, applying the utility criteria set out in Table 1, with the institutional strengthening agenda defined by the [Country Policy and Institutional Assessment](#) (criteria 14-16) serving as a proxy for action-worthiness. Similar indicators are being developed for Transparency and Access to Information.

New Impact Evaluation research on GPSM interventions

The last decade has seen a tremendous growth of Impact Evaluation (IE) research on policy questions relating to specific frontline services, very particularly in the [education and health sectors](#), generating a body of robust evidence on the effects of particular policies. The results of IE studies can complement the ISPMS and AGI country-level metrics described above. While they, broadly, can help answer the question “what tends to work in general?”, IE can be more finely grained and adjusted to test for the salience of distinctive environmental conditions, and thus able to answer the parallel question: “what is likely to work here?”

⁸ Tier II of the Corporate Scorecard is similar to Level 2 of the DfID results framework (“DFID Results”).

Downstream PSM reform areas that have already received particular attention in this emerging literature include: changes in teacher incentives (which have been shown to have direct impacts on teacher attendance and students' test scores); the impact of school-based management reforms: the effectiveness of demand-side accountability mechanisms, such as citizen score cards; and the impact of performance-contracting in service provision on service delivery results.

However, IE research on the governance of other core public sector and justice agencies is still relatively thin. The Bank will help in filling this gap by creating a community of practice of Bank client country representatives and Bank specialists and facilitating the design of IEs that address relevant questions in public sector management. Such Impact Evaluations will also contribute towards the [IDA16 commitments](#), under which the Bank committed to rigorously evaluate 51 IDA projects in all sectors in the FY12-14 period.

IE methods are well-suited for measuring the results of reforms which affect multiple actors or entities, and where the results can be seen in a relatively short timescale. Thus it is a promising approach for considering which incentive schemes are effective for motivating public servants to perform, or for measuring the effect of new budget execution processes, such as internal audit or procurement – but a priori IE is not suited for evaluating the effects of a medium-term fiscal framework (Box 4).

Box 4: Early GPSM candidates for formal Impact Evaluation studies within the Bank

Early discussions concerning priorities for extending IE to GPSM areas include:

- How to improve the performance of civil servants through recruitment, pay and performance management systems? A significant body of IE research exists on shaping performance incentives for frontline staff, such as teachers, nurses and doctors that address problems of absenteeism or underperformance. But, with a few exceptions, no equivalent research has been conducted on how to attract, retain and motivate those civil servants who work in government's core ministries such as the Ministry of Finance and Education and bear the responsibility for designing a county's policies, collecting its taxes etc.
- To date, few impact evaluations exist on the effectiveness of different tax collection processes in increasing compliance rates? On other public financial management questions, does more money translate into actual learning if schools are provided with block or capitation grants? How to design effective internal audit and control systems that reduce the risk of abuse of public funds? How can procurement processes be improved to increase the value for money of public investments?
- How can e- and mobile government contribute to better public sector performance and accountability? In the Indian state of Kerala, and in some states in Pakistan, after visiting front-line administrations such as the office responsible for issuing driving licenses, citizens get a text message, asking for their feedback on the service quality they received, and whether they were exposed to any acts of corruption. Such citizen feedback cuts information asymmetries between service recipients and senior management – and can effectively reduce inefficiencies and corruption in service delivery. But how to design such systems effectively so that citizens like to use them and management can take action?

Building on the successful implementation of other Impact Evaluation programs within the Bank, an extended program of Impact Evaluation on GPSM interventions will disseminate available evidence to stimulate a process of critical thinking and provide training in IE methods to make policy makers informed consumers of IE. The short term intention is to create a community of like-minded policy makers and researchers working together to help staff design programs in a way that allows lessons to be drawn and incorporated into policy and program design, for example by testing different modalities

to enhance performance of public servants. It will secure quality research for each country case and place capacity on the ground to help manage the work, build capacity and collaborate on the analysis. Finally, it will organize periodic cross-country events where the community of practice discusses their experience, disseminates their results and finds opportunities to scale up solutions that work.

Building research into projects

The Bank is introducing Core Sector Indicators (CSIs) to improve the consistency and quality of development results monitoring and reporting at the level of the project. While projects would have idiosyncratic context-/project-specific indicators, all projects would use a sub-set of the relevant CSIs indicators in their results frameworks thereby ensuring that performance against these indicators would be regularly monitored. Where these CSIs are relevant to a project's development objective, project performance is assessed against improvements in these indicators. In other cases the project is used as a vehicle for collecting these data. These results are then aggregated and reported at Bank-wide level. By using the utility criteria set out in table 1, the indicators under development for GPSM capture the outcomes/performance improvements that matter, not just activities/inputs on institutional strengthening. Box 5 sets out some early candidates under evaluation.

Box 5: Early candidates for GPSM Core Sector Indicators in Bank projects

1. Budgetary and financial management system

- Planning and budgeting arrangements are sound and tightly linked
- Are expenditures linked to performance goals or objectives? (base a new question around this existing OECD question)
- Extent of unreported government operations (base a new question around PEFA PI-07)
- Financial management is effective
- Aggregate expenditure out-turn compared to original approved budget
- Composition of expenditure out-turn compared to original approved budget (base a new question around PEFA PI-02)
- Effectiveness of payroll controls (base a new question around PEFA PI-18)
- Stock and monitoring of expenditure payment arrears (base a new question around PEFA PI-04)

2. Revenue mobilization system

- Composition of revenue out-turn compared to original approved budget (base a new question around PEFA PI-02)
- Cost of revenue administration (%)
- Tax compliance (base a new question around PEFA PI-15)
- Average tax productivity (personal, income, and VAT)

3. Public administration system

- Administrative capacity / productivity - Effectiveness at attracting and retaining competent staff (needs a new indicator)
- Management of overall HRM - Year on year changes in general government wages as % of expenditures (needs new indicator)

4. Accountability and oversight systems

- Citizen access to information
- Key reports are available to citizens in a timely manner (base a new question around PEFA PI-10)
- Open budget index (score)
- Monitoring and evaluation
- Credible M&E system in place for sector ministries (needs a new question)

5. Procurement

- The average prices paid by the public sector are no higher than those paid by the private sector for a standard basket of goods (base a question around PEFA PI-19)
- Timeliness of the process (from procurement initiation to receipt of goods or services)
- Percentage of contracts where the quantity and quality of the good or service procured is below requirements
- Number of staff hours required to process a procurement (procuring agency and procurement staff)

4. Qualitative Results Stories

Not all results are quantifiable. While this discussion note began with an emphasis on balancing the Bank's tacit understanding of GPSM with more "scientific" knowledge, this is not to say that all knowledge comes in the form of statistically valid metrics. The Bank's work on improving the measurement of results from support for GPSM reforms includes some work to capture the more qualitative lessons emerging from practical experiences. The challenge always is to ensure rigor in such qualitative analyses and to provide safeguards against incentives for over-claiming.

Short but robust stories of achievement

For the IDA 16 Mid Term Review, the Bank has been reviewing evidence of institutional strengthening attributable to Bank interventions in IDA countries. Institutional strengthening and governance was measured for three thematic areas:

- Administrative and Civil Service
- Public Financial Management
- Tax Policy and Administration

Projects in these three thematic areas were identified using the theme codes that are applied to the projects by the World Bank. Two separate approaches were undertaken to extract relevant indicators in relation to projects close to their closing dates and so which would have progress (or not) to report. For closed projects the [Implementation Completion Reports](#) (ICRs) or the latest [Implementation Status Reports](#) (ISRs) were referred to in order to extract the outcome indicators. In the case of ongoing projects, only the ISRs were considered. In addition, the Bank’s databases were reviewed to extract PSM project level data for all countries and all projects that were approved from FY2000 – FY2012. The results (Table 3) were then sent to the Country Management Units for validation and verification, prior to republication at the 2012 World Bank Annual Meetings.

The key safeguard in this process is that by drawing from ICRs and ISRs, all claimed results have been published and made available for wider public scrutiny.

Table 3: Selected achievement stories for the IDA 16 Mid-Term Review

Country	Achievements
Bangladesh	Government tax revenue improved from 9% of GDP in FY10 to 10.5% of GDP in FY 12 driven by improved performance in the collection of VAT and income taxes. Between 2005 and 2011, the proportion of invitations for public procurement contracts published in the newspaper increased from 70 to 100 percent and the proportion of contract awards published increased from 20 to 80 percent. There is increased competitiveness in public procurement contracts – with the average number of bidders rising from 7 to more than 10.
Honduras	There are now operational Internal Audit Units under regulations established by the Supreme Audit Institution covering the entire Central Government. 100 percent of Central Government expenditure is now included in the Single Treasury Account. Access to the courts has increased with some 40 percent of the population having had contact with justices of peace courts in the “ poverty corridor ”.
Malawi	Control over the payroll was strengthened, reducing discrepancies in the wage bill from 5 percent to 0.2 percent (2004/05-2010/11) including the removal of 4,878 ghost workers in 2011 alone from the payroll, a number that could fund an increase of 10 percent in the number of primary school teachers employed.
Tajikistan	In 2010, 2514 vacancies in the civil service (below the political level positions) were filled through a competitive based recruitment process, representing 82 percent out of 3065 vacancies filled. There had been no competitive based recruitment system in place in 2006.

Country	Achievements
Yemen	<p>A Loans and Grants Management Information System was rolled out to 23 Project Management Units. This reduced the lead time taken by line ministries for disbursements from 40 days to 8.5 days between 2010 and 2012.</p> <p>The Government Budget is now prepared using GFS economic, administrative and functional classifications, and published on the government website to obtain public feedback. This led to a 15-point increase (5th largest in the world) in the Open Budget Index between 2008 and 2010.</p>

As is discussed below, this is essentially a precursor to a more systematic use of detailed case studies to unpack key reform processes.

Capturing and systematizing practitioner knowledge

The World Bank has had an advisory, help-desk service for many years. This is now being upgraded, primarily to ensure that questions, expert responses and summaries of discussions, on different technical themes, will be fully available to Bank staff and others in a searchable format – and linked to the Bank’s various thematic websites.

5. Opportunities to Work with the Bank

As noted at the outset, the results chain from upstream GPSM interventions to results that matter for citizens is so long, that it often hard to distinguish whether changes in results are attributable to changes in the intervening institutional arrangements – or to other intervening factors. There is an urgent need to understand better “what works” in GPSM reforms and how success and improvements can be measured.

The development of more “scientific” knowledge in the field is critical for the Bank and its clients to design better and more effective interventions, to measure success and to hold themselves accountable.

Other partners have also expressed a similar need. In view of the challenging nature of the task, and the importance of getting international consensus on what works and is worth doing, the Bank looks forward to working with like minded partners on this set of issues. Some areas for potential collaboration are summarized below.

Summarizing the main areas of potential collaboration

The areas where the bank would welcome collaboration with other donors and research agencies are broadly as follows:

- Development of ISPMS/AGIs:
 - Testing the relevance of available data
 - Developing new ISPMS/AGIs
- Better learning from projects:
 - Testing the robustness of the Corporate Scorecard approach
 - Developing a systematic program of Impact Evaluation on GPSM interventions

- Testing core sector indicators in other settings
- Capturing qualitative practitioner knowledge
 - Moving from short stories to robust case studies
 - Joint help desks and web portals for capturing and sharing practitioner knowledge.

Testing the relevance of available ISPMS/AGI data

Joint analytical work is under way with DfID to take stock of existing data collecting methodologies and their quality/costliness and to test the salience of existing ISPMS/AGIs for sector outputs and outcomes. This research aims to produce a set of research papers to advance learning on “what works and why” in GPSM reform and to build a multi-donor initiative for expanding the scope of available ISPMS/AGIs.

The research direction, at present, is indicative – but is likely to focus on the following questions:

1. Which political and other contextual variables explain ISPMS/AGIs results or trends?
2. Do ISPMS predict downstream functions / outputs as theory would predict?
3. Which theories best explain variation in ISPMS/AGIs?
4. Are there country level patterns in the data (e.g. country clusters - by economic classification, political context etc.) that call for theoretical explanations?

DfID has expressed particular interest in focusing on the second set of questions, i.e. on whether ISPMS predict downstream functions / outputs. Specifically, it is envisaged that a first DfID research paper could focus on one of the following GPSM systems and its impact on downstream results (depending on DfID priorities and data availability):

- Public Information Systems: Can disclosure of information - in particular transparent management of public resources - improve trust in government?
- Public Financial Management Systems: Does the quality of PFM systems influence spending efficiency of public resources in education / health (given a level of resource allocation to these sectors)?
- Revenue Mobilization: Do measures of the quality of tax administration predict government ability to raise taxes?

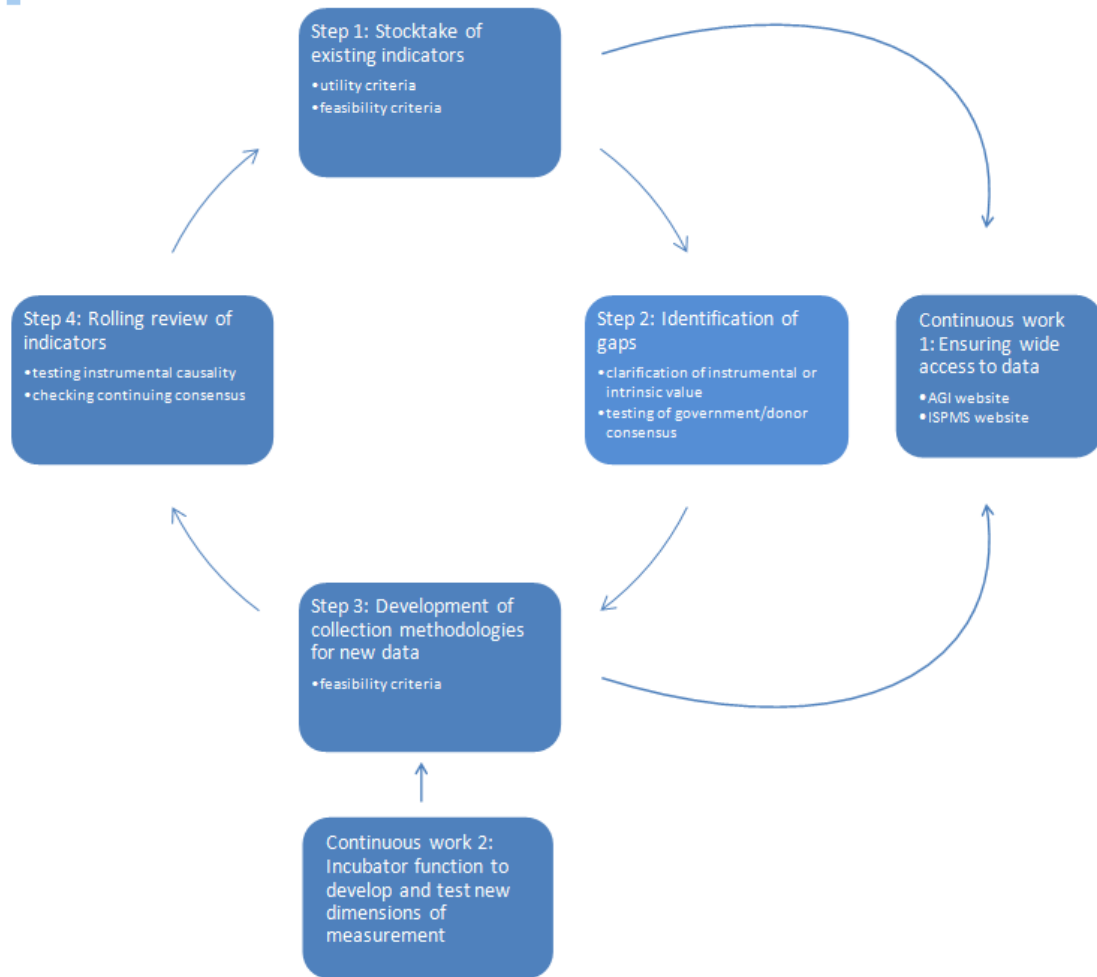
DfID is currently finalizing its decision as to which of these GPSM systems and which exact question the research will focus on.

Development of New ISPMS/AGIs

A core group of donors with a joint interest in working on ISPMS/AGIs has been identified. The German Development Co-operation, DfID and the World Bank have been discussing work possibilities since the launch of the ISPMS initiative in April 2012. Other donors, including IDB and ADB, have expressed interest in the work.

The consortium is proposing a pragmatic and collaborative approach which comprises a rolling program with three steps and two parallel functions, summarized in Figure 3.

Figure 3: Schematic overview of the Multi-Donor Program



Collaboration in better learning from projects

Testing the robustness of the Corporate Scorecard approach

The Bank’s work in developing defensible metrics to identify progress that it is making in institutional strengthening is a reasonable start in a difficult area. In the latest (2012) edition of the Corporate Scorecard the Bank has shifted from counting the number of projects directed at improving the quality of GPSM systems to counting the number of countries that with Bank support have strengthened public management systems in (i) Civil service and Public Administration; (ii) Public Financial Management; (iii) Tax Policy & Administration; and (iv) Procurement. Similar indicators for Transparency and Access to Information are being developed and will be added next year. While this is a big step forward, there is scope for further improvement and the use of more sophisticated measures of what constitutes strengthening.

Collaboration would be welcome in further refining this – and very particularly in testing different approaches for aggregating evidence of progress across a range of institutional dimensions.

Developing a systematic program of Impact Evaluation on GPSM interventions

DfID and other donors are funding major programs of Impact Evaluation in the Bank, very particularly including the [Strategic Impact Evaluation Fund](#) (SIEF). But these are largely in “downstream” service delivery areas. However, as noted earlier, recent research demonstrates that IE methods can successfully be applied to learn about the effects of upstream reform elements in case these reform elements can be applied selectively or successively. This offers the opportunity to stimulate IE in GPSM – a long neglected area. A set of at least 3 GPSM projects is included in the 51 IDA projects that the Bank has committed to rigorously evaluate under its IDA 16 commitments. Exceeding this commitment, it is currently planned to launch impact evaluations for about 6 GPSM projects in FY 2013.

Collaboration in building a community of practice amongst researchers developing IE approaches in core GPSM work would be of immense value as (i) the number of amenable projects in the portfolio of any single donor will be small; (ii) the work is time and resource intensive; and (iii) achieving some international consensus on the value of this approach for evaluating upstream reforms will be key.

Testing Core Sector Indicators (CSIs) in other settings

As in many of the other areas of this work, getting consensus within the donor community and clients on what is worth measuring, achieving consensus on the CSIs would enhance their value as public good. Thus the research dividend from extending the Bank’s work on core sector indicators for inclusion in all relevant Bank projects could be leveraged significantly if agreement could be found on a common suite of indicators to be included in project activities supported by other donors also.

Capturing Tacit Practitioner Knowledge

Moving from short stories to robust case studies

Arguably, the value of case studies in GPSM reforms has been undermined by a tendency for all donors to use cases selectively to highlight successes and by a lack of rigor in conducting qualitative research. Yet, in the field of public management, qualitative methods are and will remain a central research tool for gaining in-depth insight into complex reform processes. Achieving a high level of rigor in case analysis is particularly important in view of the limitations of other methods for understanding what works in GPSM reforms. Cross-country econometric analysis remains constrained by a lack of quality comparative data (ISPMS/AGIs) on public administrations and Impact Evaluation methods are particularly suited only for some types of GPSM reform. A multi-donor approach to developing a body of rigorous case studies on key GPSM reform questions could aim to increase the number of projects and countries subject to such analysis, avoiding a positive selection bias. This would help in reaping the tremendous learning potential from donor-projects, which currently often remains tacit knowledge. The result would be the development of a library of rigorous and peer reviewed cases spanning different public management systems and countries.

Joint help desks and other approaches for capturing practitioner knowledge

The PSG advisory service within the Bank (Askgov@worldbank.org) is building a knowledge management architecture by systematizing the common advisory services database and sharing the results with other donor partners via a password protected portal. There is significant scope for economies of scale in teaming up with others on this.