THE GEORGE WASHINGTON UNIVERSITY

WASHINGTON DC

The Institute of Brazilian Issues (IBI) / Center for Latin American Issues (CLAI)

School of Business

THE MINERVA PROGRAM

MONITORING AND EVALUATING GOVERNMENTAL PROGRAMS:

LESSONS FOR DESIGNING A GOVERNMENTAL NETWORK OF M&E

RAFAEL VENTURA ABREU

Rio de Janeiro State Secretariat of
Planning and Management (SEPLAG-RJ)

Advisor: Kathryn Newcomer

Washington, DC

Fall 2013

CONTENTS

1. INTRODUCTION

1.1 *The role of SEPLAG within monitoring and evaluating State programs*

The activities of monitoring and evaluation carried out by the Rio de Janeiro State Secretariat of Planning and Management (SEPLAG) are directly related to some of the main objectives and functions of this agency. According to its internal statute the agency must "coordinate and evaluate the formulation and operationalization of public policies, as well as propose corrective alternatives and resizing of governmental actions, when necessary" (RESOLUÇÃO SEPLAG Nº 894; 2013).

Another important function of SEPLAG under the responsibility of the Undersecretary of Planning is the elaboration of the State Pluriannual Plan, submitted by the Executive to the Legislative every four years. The Pluriannual Plan is a legal instrument defined by the Federal Constitution of 1988, art. 165, and by the State Constitution of Rio de Janeiro, art. 209, and defines the governmental programs for each quadrennium. The plan covers the intended actions of the administration within this period, organizing them in programs aligned with the strategic orientations of the government and the budget availability predictions. The plan should set out the mechanisms for evaluating and measuring government performance as well as the impacts of its actions on the population of the state of Rio de Janeiro. Moreover, it should be an instrument that helps decision-making by orienting budget allocation to fulfill the administration's goals.

SEPLAG, through its Undersecretary of Planning, is responsible for coordinating the processes of elaboration, revision, monitoring and evaluation of the Pluriannual Plan, in partnership with all other state agencies. Each one of the state's agencies is responsible for creating and executing their programs, dividing them in different goods and services that should be delivered to the population in order to reach the objectives of the programs. The Undersecretary of Planning is responsible for the methodological support to guarantee that the intended actions of each agency are aligned with its functions and expressed in a logical and measurable way.

To improve the process of monitoring and evaluation (M&E) throughout the different agencies of the State, SEPLAG has decided to create in 2013 a *Planning Network* focusing on M&E. The objective of this network is to qualify governmental

officials in monitoring and evaluation mechanisms in order to help the elaboration of a Pluriannual Plan better aligned with each agencies' actions and more measurable.

The literature on M&E has many lessons applicable to the design of the Planning Network. Two of them should be highlighted. First of all, the evaluation literature can be a useful resource to gather information about different methodologies and techniques of evaluation: the strengths, weaknesses and applicability of several quantitative and qualitative methods. The importance of this methodological discussion to improve the quality of the work performed by governmental officials is undisputable. But I believe, before that, evaluation literature can help in another crucial dimension that has to be faced prior to the definition of the *content* that should be taught to other agencies' officials.

Thus, reviewing the monitoring and evaluation literature provides many insights about structural factors that should be considered by a governmental network whose objective is to improve the state's M&E process. This literature review aims to answer the following questions: (a) how should the general principles of evaluation be implemented in a horizontal network composed of state agencies?; (b) historically, what were the most effective institutional incentives for conducting evaluation?; (c) what can we learn from the different existing approaches on requisites that constitute good evaluation practice?

In order to search for these answers one must recognize that Brazilian and Rio de Janeiro state agencies still have to resolve some essential questions in order to guarantee an environment that encourages evaluation. One must change the focus from the important discussion about evaluation methodologies and techniques to a discussion on what can and should be done before engaging in evaluations. Finally, one must escape the false statement of *any evaluation is better than no evaluation* to establish what factors a program and an agency should guarantee to support good evaluation practice, whatever the method chosen.

1.2 *Program Evaluation and Governmental Programs*

This paper will discuss some general features of program evaluation, but will focus on governmental program evaluation. We will focus on the aspects of the literature that are more relevant to governmental programs, and to the circumstances of the Rio de Janeiro State Government in particular. That said, it is important to clarify one point. The word *program* is used in a broad sense, meaning "a set of resources and activities directed

toward one or more common goals" which "may consist of a limited set of activities in one agency or a complex set of activities implemented at many sites by two or more levels of government (…)" (Newcomer et al, 2010, p.5).

This definition is required to mark the distinction from more formal understandings used in states manuals in Brazil, from Federal to States Governments. The definitions are not excludable, as we can see in the definition of *program* in the Manual of Elaboration of the State of Rio de Janeiro Pluriannual Plan:

> "A program is the instrument of organization of governmental action. It is an articulated set of actions (budgetary and non-budgetary) that contribute to the achievement of a common goal (…). Once the one problem is identified, one should seek their causes and consequences and develop actions to combat them, reduce them or eliminate them. The amount of resources and operational capacity of an agency define the goals and values associated to each action. A program is implemented through the execution of actions (projects, activities, special operations and non-budgetary actions) that should be sufficient to the achievement of the objective of the program (…)". (Rio de Janeiro, 2011, p. 33, translated by the author)

In this definition, we can see that programs can have a narrower connotation of a set of budgetary and non-budgetary actions, following the Ordinance n. 42 (Brasil, 1999) from the Federal Department of Planning, Budget and Management, used as a base for the structure of the Rio de Janeiro Pluriannual Plan[1].

We can see that it is possible to talk about evaluation of governmental programs that may not be the same programs formally inserted in the Pluriannual Plan. Frequently, an evaluator will emphasize a sub-set or a specific action within the program. The reasons may vary. Sometimes a program inserted in the Pluriannual Plan contains actions so diverse that it is difficult to recognize the main goal behind it. In reality it may include different initiatives loosely tied as a unique program. In other cases, even in a well-constructed program, an evaluator will focus on some particular aspect when defining the *evaluand* – the intervention that is intended to be evaluated. The following statement about federal programs in the United States seems to be perfectly valid to most of the state programs in Rio de Janeiro:

---

[1] The structure of the Pluriannual Plan is formed by *Programs* that contains one or more *Actions*, divided in different *Products*, that are defined as goods and services delivered directly to society. The *actions* are the smaller unit for budget purposes (the intended resources in the Pluriannual Plan and the allocated resources in the Annual Budget Law).

"Sometimes the term "program" is simply a funding stream, such as a line item designating a program in the federal budget that allocates money for a broad purpose. Those programs may encompass many different types of activities in different locations, rather than a specified set of activities or components. Using the term "program" for such a diverse set of local activities makes federal-level evaluation more difficult and likely to yield ambiguous outcomes, for there is not a well-developed and well-defined evaluand to be evaluated." (Scheirer, 2012, p. 265).

This means that when we think of evaluation we cannot think of the *program* as given: we have to think if the logic underlying the program is clear and explicit, and also if the program is being implemented as it is formally stated.

2. EVALUATION: REASONS, STRATEGIES AND INCENTIVES

2.1 *Making evaluation useful for governmental programs*

One of the first questions evaluators should ask is: why is it necessary to evaluate? Making clear the reasons for evaluation is one of the first steps to making well-informed decisions when the evaluation process is confronted with the limitations of everyday life. To say that is to recognize that it is impossible to evaluate everything in every manner, and that we have to deal with the constraints of resources – in terms of budget, personnel and expertise. This is particularly relevant for an emerging network of M&E, such as the one presently being created in Rio de Janeiro's state government, that must confront a lack of evaluation experience throughout the government.

Discussing motivations for evaluation we should recognize that this process usually involves interests of a wide range of groups. In a governmental program, citizens and elected officials may be interested in the outcomes of a program that receives a significant amount of public resources. Program staff are interested in improving program performance through evaluation. Finally, managers and leaders have to make decisions based on some evidence about the results of the program. Recognizing that, we see that a *useful* evaluation that is worth the cost and the work involved should not only answer accountability questions, but also be used to improve program performance (Newcomer et al, 2010, p.5-6).

It should be clear that the interests of different publics are related to different purposes of evaluation, which in our opinion should be viewed as non-excludable. These purposes are: (a) evaluation for accountability, to ensure funding is being used appropriately; (b) evaluation for causal knowledge, to evaluate if the program causes the expected results; and (c) evaluation for program improvement, focusing on using data about short-term outcomes to help managing the program (Chelimsky, 1997 apud Scheirer, 2012, p. 266). Considering the interests of multiple publics and different purposes of evaluation leads us to a first fundamental reminder to a governmental M&E network: *when evaluation is made only to answer external accountability without helping managers improve their programs, they tend to be less involved, and the results are not worth the cost*.

One fundamental approach that makes clear the link between evaluation and usefulness is the concept of *utilization-focused evaluation.* It is expected that every

evaluation will be used somehow, but the following definition makes explicit the importance of designing an evaluation in order to maximize its use to improve a program:

> "Program evaluation is the systematic collection of information about the activities, characteristics and outcomes of programs to make judgments about the program, improve program effectiveness and/or inform decisions about future programming. Utilization-focused program evaluation (as opposed to program evaluation in general) is evaluation done for and with specific, intended primary users, for specific, intended uses." (Patton, 1997 apud Preskil and Catsambas, 2006, p. 37)

This attention to the utilization-focused approach may sound strange as we discuss governmental programs, which are typically subject to evaluation for others purposes: external control/accountability, decisions about funding, etc. It has to be clear that we do not ignore these other fundamental reasons to evaluate a public program. Later on we will discuss the role that institutional incentives play in stimulating evaluation. What we argue instead is that without making the evaluation useful to program improvement we tend to jeopardize other motivations. Public administration in Brazil is full of *control mechanisms* that are not well justified to those who are subject to them. Therefore, through emphasizing the usefulness of evaluation to program staff we can probably produce better evaluations to fulfill accountability reasons.

This emphasis on the use of evaluation to improve a program proposes that evaluation should be viewed as an ongoing effort and part of the program life-cycle: from planning and execution to replication and dissemination. Evaluation should be viewed as a managerial function through which data are continuously collected to inform decision making, not only as an ex-post activity. Even when planning a new program, the work of evaluation can be helpful to understand the problem being addressed and to understand the evidence from prior literature and experiences.

During the beginning of implementation phase, if possible in a pilot experience, *evaluability assessment* can be used to explore the assumptions about the program, to see if program activities are logically related to one another and to the desired outcomes[2]. The collection of data during the implementation phase is also crucial to posterior impact evaluations. Records about a program's execution are important to associate the findings to the real implementation of the program: "was the program not implemented well (implementation failure), or was it not more effective than its comparison, even when well implemented (theory or program development failure)?" (Scheirer, 2012, p. 269).

---

[2] To see a further discussion on Evaluability Assessment, see section 3.2.

This continuous evaluation throughout program implementation is nothing more than the other part of the M&E Network: the *monitoring* activities. Monitoring, in this sense, is an ongoing evaluation that is intended to measure the performance of a program, directly related to the purpose of program improvement. It is typically an internal phase of the evaluation process conducted by the program manager and staff, aimed at short-term outcomes that can evaluate the implementation of the program and lead to corrective alternatives in the short-run.

2.2 *Setting basis for evaluation*

Certainly, a good evaluation focused on improving a program's performance relies on a rigorous definition of methodological approaches. But before that, we have to answer some fundamental questions and see if the program is eligible for evaluation:

"(i) Can the results of the evaluation influence decisions about the program?

(ii) Can the evaluation be done in time to be useful?

(iii) Is the program significant enough to merit evaluation?

(iv) Is program performance viewed as problematic?

(v) Where is the program in its development?" (Newcomer et al, 2010, p. 7)

As we will see ahead, these questions lead to the necessary involvement of stakeholders of the program, to certify that the results of the evaluation process will be useful. When dealing with governmental programs, we have to acknowledge that some legitimate political factors play an important role in answering these questions. But the challenge and objective of a good evaluator is to give as much evidence-based information to guide the decisions about a program and on the eligibility of evaluation itself.

The answers to the questions above will help in deciding on an evaluation strategy, which necessarily should be defined before choosing between different techniques of collecting and analyzing data. The strategies reflect the general approach behind evaluation. Newcomer (2010, p. 560) defines three usual strategies when evaluating public programs. The first is the *problem-based investigation* that normally answers questions about poor implementation, mismanagement or even fraud. The second one is the *performance assessments*, focused on collecting data about the performance of the program that goes beyond financial data. The third is *impact evaluation*, which deals with evidence about the real impact of the program and the extent to which it really causes the intended outcome. Considering the present stage of public administration in the State of

Rio de Janeiro (and maybe in Brazil), evaluation is dominated by the first strategy (problem-based), with some examples of the second (performance assessment) and few of the third (impact evaluation).

The first strategy of evaluation is centered in the actions of the States Courts of Audit (Tribunais de Contas Estaduais) and, at the federal level, by the Federal Court of Audit (Tribunal de Contas da União) and the Office of Controller General (Controlodoria Geral da União)[3]. The problem is that, despite having some investigation about implementation, the attention tends to be concentrated on financial audits instead of feedback to program managers. Actually, the audits are viewed by program managers as overly centered on procedural and bureaucratic controls.

The second strategy (performance assessments) is starting to become more disseminated through public agencies. The transition from a bureaucratic model, centered in external control and inputs, to a model of public management that is focused on results stimulated the creation of monitoring systems within agencies. The problem is good experiences are not shared with others within the government, and there is still a lot of confusion around what should be monitored. Normally, the way to assess performance is with records of service delivering, that sometimes have incorrect indicators to measure what is intended.[4] One of the objectives of the M&E Network being constructed is to disseminate choices of methodologies to the agencies.

Unfortunately, the third strategy is more the exception than the rule. Some programs measure the changes in social indicators after implementation, but studies measuring causalities while controlling for other intervening variables are still rare. As States lack the personnel able to conduct impact evaluations, most of the evaluations are conducted with external consultants.

To construct the basis of a network of M&E in Rio de Janeiro, a realistic approach would probably not ignore the important role played by audits. However, any change in this process involves a major shift in the equilibrium of powers within the government. What can be registered is that audits should focus more on implementation and outcome evaluation and, also, that is important to have a more proactive posture to evaluation, instead of reacting only when a problem is identified.

---

[3] The Courts of Audits are responsible for the external control and are indepnc endent of the Executives Branches, while the Office of the Controller General is responsible for the internal control of the Federal Government. In State's level, the internal control are usually located in the Finance Departments.

[4] Of course, this is not an exclusivity of Brazilian experience. For the US case, see, for example Behn (2009).

The role that the M&E Network can fulfill involves performance assessments and impact evaluations. Considering a pragmatic approach, what is defended here is that we still have a long way to go before internally handling impact evaluation – what is lacking most of all is the expertise to do it. It does not mean that the executive branch is incapable of doing any impact evaluations, but rather that focusing on improving performance assessments can be a strategy that more realistic can develop, in the mid-term, a *culture* of impact evaluation. A well-designed monitoring system, for example, can help in gathering data for an impact evaluation. And we should not forget that evaluation should be viewed as a part of the program life-cycle, and the two strategies shall not be viewed as two separate moments. That is why on this paper we focus on some important points that can be implemented in the short-term, setting the basis for stimulating impact evaluations in programs of the State of Rio de Janeiro.

Finally, we should consider nine crucial points as the foundations of good evaluation. These points should be the starting point for any evaluation process: 1) identify the audience for evaluation; 2) consider the timing of evaluation design and use; 3) make sure that the data collected is relevant to the intended purpose of the evaluation; 4) pre-define the format of the final report; 5) assure the staff has the competence and skills required to perform the evaluation; 6) make sure that the number and format of the final recommendations are appropriate for the final audience; 7) choose the adequate communications channels to report the results; 8) ensure that evaluation information is used; and 9) encourage the use by decision makers of evaluation data to improve programs (Newcomer, 2010, 569-71).

These points can be viewed as a kind of *golden rule* that guides the decisions in the beginning of the process. In particular, the consideration of point number 5 (the competence required) is a justification of the pragmatic approach followed here: what can be improved with the present resources and what aspects should be developed in the future. What is defended is that, although it lacks expertise, the State of Rio de Janeiro can improve significantly evaluation with its current resources if it takes in consideration some established principles about program evaluation.

2.3 *Lessons for a brief history of program evaluation: some institutional incentives and constraints to evaluation*

The literature is full of examples of how the development of program evaluation in the public sector is related to some institutional changes within the government and/or the legislation. Of course, we do not want to reconstruct the entire development of program evaluation, but rather highlight some of these aspects, focusing mainly on the case of the United States and France. Hopefully, these lessons will help us to think about the necessary incentives to be built in the M&E Network of the State of Rio de Janeiro.

The literature is replete with rich examples of different initiatives that led to the development of program evaluation. Despite isolated cases that can be tracked to the beginning of the 19th century, the late 1960s and 1970s are usually viewed as the moment when some actions converged to give prominence to program evaluation.

In the United States, the Elementary and Secondary Education Act (ESEA) of 1965 was a landmark. This legislation mandated the government to access student performance and teacher's quality standards and to provide resources to develop these activities. It also set aside part of the budget to this and thereby institutionalized evaluation (Imas and Richt, 2009, p. 22). Since the 1970s, there was a confluence of incentives that raised the demand for systematic data about social services delivery. Actions by several institutions – federal and local governments, nonprofits, foundations, think tanks and the academy – converged to the development of program evaluation (Newcomer, 2008).

The 1980s are viewed as a moment of backlash of evaluation within the public sector. Growing government deficits led to spending cuts, and program evaluation was considered a low priority. Budget constraints and the political orientation of the Reagan administration led the type of evaluation undertaken by federal agencies towards a punitive approach. This punitive approach brought some important changes within the federal bureaucracy. With the budget limitations, the offices responsible for evaluation focused on rapid feedback and played a secondary role. With this scenario, Congress relied on the General Accounting Office (GAO) for evaluations of programs' performance. Until the 1970s, the GAO was mainly focused on financial audits, but during the 1980s program evaluation became an important part of GAO tasks. Another important change involved a reorientation of the Inspectors Generals Offices (IGs) from a reactive role, investigating fraud, to a more active role in assessing performance,

ensuring accountability and reflecting the missions of federal agencies (Newcomer, 2010).

One fundamental turning point addressed by the literature is the Government Performance and Results Act (GPRA), approved by the U.S. Congress in 1993. From this point on, all federal agencies had to submit performance plans and reports on all federal programs to Congress. Also, the law required that all federally funded programs should develop performance measures. The GPRA played a major role in spreading the requirement and practice of evaluation within the government.

In France, a first step in introducing evaluation was made in the beginning of the 1980s. It was introduced by the Finance ministry as a way of rationalizing public expenditures. It was a brief experience, and for that time on, evaluation has never been directly related to the budgetary process. In 1990, a National Scientific Council of Evaluation (CSE; Conseil Scientifique de l'Evaluation) and an Inter-ministerial Committee of Evaluation (CIME; Comité Interministériel de l'Evaluation) were created. The CIME was responsible for deciding what evaluations were eligible for special funding. According to Barbier (2003), the centralized structure made this experience less relevant in terms of the capacity of evaluations than it was able to assess than the role it played in developing standards that contributed to the institutionalization of evaluation practice in France. A structural feature of the French tradition made the State a quasi-monopoly in matters of policy analysis. This centralization, combined with the absence of a professional association of evaluation with its owns standards, made the evaluation field in France less defined in terms of the differences and frontiers between evaluation, research, auditing and control. With this picture, the evaluation market in France was driven by its demand side, mainly stimulated by the influence of the European Union, because EU level programs all include the explicit implementation of evaluation regulations.

The first evident conclusion of the exposition of the cases of U.S and France is that legislation and institutional changes play a big role in stimulating the culture and practice of evaluation. But what is also evident is that different institutional structures lead to diverse results in institutionalizing evaluation, especially if confronted with different traditions. Making evaluation a stated goal of a government through legislation is not enough: we have to think about the way that it is implemented.

The experience in the United States shows that pieces of legislation made a significant impact on the dissemination of evaluation within the government. Putting

together the pieces of the puzzle, you have a budget reserved for evaluation, the obligation of measuring performance of programs with federal funds and agencies dedicated to fulfilling this role, either inside each department, as the IGs, or as an independent agency with an advisory role, as the GAO.[5] We are not affirming that everything is perfect within this institutional background.[6] But it is indisputable that the incentives that were given to the practice of evaluation were considerable, especially in comparison with other experiences. On the other side, the French case showed a different experience, with a centralized initiative less effective in stimulating the evaluation practice within the public sector. The stimulus for evaluation in France seems to have been driven by the country's insertion into the European Union, not by its own institutional incentives.

The examples of institutional incentives given to program evaluation within the public sector in Brazil are not many. We can find several laws that express the auditing approach, centered in very strict regulations, but is hard to find examples of institutional incentives given to program evaluation. At least in the federal level and in Rio de Janeiro's government, there is not any general legislation requiring evidence gathered through some sort of evaluation which decides the continuity of a program, or ties the funds of a program to a performance measure. Of course there are experiences of evaluation in Rio de Janeiro and Brazil, but they are more discretionary than the result of some sort of rule. And, what is more relevant, there is little evidence that the conclusions of evaluations are used in a systematic way.

This picture leads us to a common discussion in public policy in Brazil: do we need more legislation? For me the answer is: if by legislation it is understood that creating rules without the conditions and incentives to make the rules applicable, no. Creating obligations (for example, that you have to evaluate your program) without giving the ways to accomplish that (the expertise, the personnel) or the incentives do it (the program with which I am involved can benefit from it?) is certainly not enough.

The second aspect to be observed is that besides the institutionalization of evaluation within public sector, we can see differences of these experiences on the degree of professionalization of the role of the evaluator as a profession on its own. One

---

[5] For reason of focus, we focused briefly only on the federal public sector, without considering experiences in the local level and within non-profits. Certainly, the full picture of evaluation practice in U.S has to considerate these other relevant actors.
[6] The GAO/GGD- report "Evaluations Help Measure or Explain Performance" (2000), for example, gives examples of evaluations initiated in response to legislative provisions and, at the same time, points out limitations of governmental agencies capabilities to gather and use performance information and make GPRA fully effective.

important aspect of the French experience was the underdevelopment of professional associations of evaluation. The French National Society of Evaluation (Société française de l'évaluation - SFE) was created only in 1999 (Barbier, 2003). As said by Barbier, the frontiers between evaluation, auditing, research and control remained blurred. In the U.S. the professionalization of the evaluator can be traced back to the 1980s. The American Evaluation Association (AEA) was created in 1986 as a merge of two existing evaluation-focused organizations, the Evaluation Network (ENet) and the Evaluation Research Soceity (ERS) (IOCE, 2012).

The closest there is to a professional association of evaluators in Brazil is The Brazilian Monitoring and Evaluation Network. The BMEN is even more recent: it was created in 2009. The late date indicates the degree of professionalization of the evaluation work in Brazil. It was created with support of the World Bank, the Inter-American Development Bank and by a State foundation (João Pinheiro Foudation). In 2012, 54% of its members were from the government.[7] We can see that members of the government are the biggest part of the network, making it clear the distinction to a professional association of evaluators. But this does not mean there is strong involvement with evaluation within the public sector. A study about the practice of evaluation in Brazilian Federal Government programs concluded that is not known what is being evaluated and how (quality) and what results the evaluation studies are generating for the improvement of public policies and if they are being properly disseminated (transparency).[8]

The M&E Network of Rio de Janeiro must be aware of these two points: being a government network based in qualification given by an agency (the Planning and Management Secretary) to others agencies of the State it can, at its best, ensure the quality of the information transmitted, but not assure that the knowledge will be applied. This is not an insurmountable barrier, but rather draws attention to three factors. First, that it is highly recommended fostering contacts with the BMEN and the academy in order to exchange experiences. Second, that the M&E Network should advocate for some kind of regulation that gives real incentives to institutionalize program evaluation. And, finally, that a significant part of the qualification given to the agencies should focus on the concrete benefits – the motivations – that program evaluation can have to themselves. Or to put in another way, that they do not have to respond for an external obligation or some

---

[7] Members of the BMEN in 2012 (IOCE, 2012): Government: 1913 (54%); Private sector/NGOs: 489 (14%); Academics: 344 (10%); Consultants: 548 (16%); Other: 225 (6%).
[8] http://www.ioce.net/en/PDFs/national/2012/Brazill_BMEN_Presentation.pdf

abstract principle, as transparency and accountability. If the Networks fails in doing that, probably the utilization of the knowledge transmitted will be jeopardized.

3. LESSONS FOR STIMULATING THE PRACTICE OF EVALUATION WITHIN RIO DE JANEIRO'S GOVERNMENTAL PROGRAMS

The literature on program evaluation contains several theoretical and practical insights about the benefits and limits of different methodologies and techniques used in evaluation work. This section of the paper is not dedicated to summarize the variety of procedures, neither to defend any one in particular. First of all, because doing so would be hugely naïve in a context as plural as the programs implemented within a government. Each program, with their specificity, and depending on the questions you want to answer through evaluation, will require different approaches or – more probably – a combination of different approaches. What this section intends is to explore some aspects of the literature that are more relevant to programs in the context of Rio de Janeiro's government. Doing so, we consider the current stage of evaluation's institutionalization within the government: what are the aspects that should be dealt first to improve evaluation capacity in order to stimulate the practice of it throughout government.

It is important to keep in mind that different evaluation approaches should be viewed more as continua than as excludable choices. Newcomer et al (2010) distinguishes six continua on which evaluation approach differ: 1) formative – summative; 2) ongoing – one-shot; 3) objective observers – participatory; 4) goal-based – goal-free; 5) quantitative – qualitative; 6) problem-orientation – non-problem. We will explore certain aspects related to some of these continua that we believe are more relevant for a network that is in its building stage.

3. 1 *Formative and summative evaluation: do we have to focus on one of them?*

We can begin this section anticipating the answer for the question posed above: probably, we should not focus specifically on one approach instead of other, but it seems reasonable to defend that a network that intends to stimulate the practice of evaluation through governmental programs has a lot to gain by focusing on its beginning in the formative approach. However, as we emphasize, to have a focus on formative approach is a way to prepare the field for future summative evaluations.

Let's begin with a brief definition of the formative and summative approaches:

"Formative evaluations are evaluations intended to improve performance, and are most often conducted during the implementation phase of projects or programs. (…)

Summative evaluations, by contrast, are studies conducted at the end of an intervention (or a phase of that intervention) to determine the extent to which anticipated outcomes were produced." (OECD 2002, apud Imas and Rist, 2009, p. 9)

It is obvious that there is a difference in *timing* between the two approaches, but what is more relevant are some aspects that follow this timing difference. Thinking in a context of a network that has to deal with programs in various phases of its life cycles, it is evident that both approaches will be used. But what we want to make clear is that a good summative evaluation has a lot to gain if some formative questions were addressed previously during designing and implementation stages. If we are dealing with a program that was not designed and implemented with some evaluation questions in mind, addressing the impact of the program can be compromised if we do not take a few steps back to reflect on some formatives aspects of the program.

Formative evaluation focus on the way a program is implemented. It evaluates if the program is being implemented according to the intended operational logic and the immediate outputs of the implementation stages. A formative evaluation can be a process evaluation, focusing on the operations of a program. It can also be viewed as a midterm evaluation, addressing in a middle stage of a program what is working well and what should be improved. It is clear that formative evaluation is very close to monitoring activities, with indicators designed to assess if a program is being implemented properly.

Summative evaluation is typically what comes in people's mind when they hear *evaluation*: it is an ex-post activity worried about the impact or the outcome of a program. Summative evaluation measures if the intended results were accomplished, and aims to establish causal relationships between the program implemented and their results. Summative evaluation is of vital importance to make decisions about continuity, replication, scaling up or ending a program.

It is important to stress again that the two approaches are complementary. We can highlight the specificities of each, but in practice they tend to overlap. A formative approach is not limited to the planning stage. Methods often viewed as formative are fundamental for evaluation in later-phases of the program, especially if we want to address the causal linkages between the program and its results (Scheirer, 2012).

Therefore, it is evident that both of these approaches are needed, and related to different phases of the cycle of a program. However, focusing on some aspects brought by the formative approach can be important for three reasons. First of all, this kind of implementation evaluation requires information that – at least part of it – is already

available with program staff, even if it is not organized or used as data for monitoring or evaluation activities. In this sense, it is easier to start with this approach, and try to engage program staff showing that part of the work they already do is part of an evaluation process, even though they do not think about it. Second, designing a program with this is mind, or reorganizing relevant data already existent to assess the implementation of a program will be of vital importance to decide the adequate methodology for an impact evaluation. Finally, formative evaluation plays a crucial role in defining what you are evaluating, *before* addressing its impacts. It is a way of analyzing the implementation of the program, to see if it is being implemented as planned. After that, you can question if the observed impacts can be related to the policy actually implemented.

3.2 *The case of Evaluability Assessment and the reconstruction of Logic Models*

Evaluability assessment can be viewed as a first step to further evaluation, closer to the formative approach or, to be more specific, as a type of a prospective approach (Imas and Rist, 2009) or exploratory evaluation (Wholey, 2010). Being a first stage for future impact evaluations, it is not a surprise that we argue that the Network of M&E under-construction should start by paying attention to it. In a way, every good evaluation should do so. However, some specificities of governmental programs in Brazil make this strategy not so evident. The hurry to evaluate the immediate effects of a policy can lead to poor evaluation if you do not have the expertise or the information required. Evaluability assessment is, in this sense, an approach directed to answer two fundamental questions that were brought through the initial parts of this paper: (i) what should be evaluated; and (ii) is it useful evaluating it?

Evaluability assessment (EA) "assesses the extent to which programs are ready for future evaluation and helps key stakeholders come to agreement on realistic program goals, evaluation criteria, and intended uses of evaluation information" (Wholey, 2010, p. 81). It is important to stress that an EA addresses if the program is *ready* for evaluation, not if it *can* be evaluated, because in a broader sense every program can be evaluated. However, the evaluability assessment approach gives us some criteria to know if the evaluation is likely to be useful: (a) there is an agreement between different stakeholders upon the programs goals and they are realistic; (b) there is an agreement upon the process, inputs, outputs and outcomes that will be evaluated; (c) the data requires is obtainable; and (d) intended users are willing and able to use the information produced (idem, p. 83).

One of the crucial steps of an evaluability assessment is the engagement of intended users of evaluation and different stakeholders[9] in order to clarify program's goals and design. It is necessary to make explicit a goal of the program that can be implicit or not consensual upon different stakeholders. The assumptions underlying the program design have to be clarified: the inputs, the activities, the intended outputs and outcomes and the causal linkages between these aspects. Although it seems like an obvious step to do, it sheds light on the fact that different perspectives may occur among program staff, states officials and so on. To know if the program is being implemented as expected, the stakeholders should agree in what is predicted. After these initial steps, the evaluator should evaluate if the program is being implemented as intended. Doing so, we start reflecting about some performance measures and factors that are inhibiting the expected performance of the program. Doing so, the evaluator can have, from the start, a notion about the feasibility of accomplishing the intended results, or an estimate of it. Some important recommendation about the program design and implementation can then be made.

Evaluability assessments can give some important feedback about a program even if we are not evaluating the final impacts of it. To put in another way, we do not have much to say about impacts in this moment, but we may have indications that the expected outcome is not likely to happen. Also, as the final product of an evaluability assessment, it can be defined the options and focus for further evaluations addressing the impact of the program. Alternatively, the conclusion of an evaluability assessment can be that the program is not ready for further evaluation. It can be the case that there is no formal program design or model in place, or that the program's goals are unrealistic or unattainable (JJEC, 2003). In this case, program staff still has some work to do. To help in this necessary revision, logic models can be a useful tool.

Logic models are a useful tool to clarify some factors and terms that have been brought up in different parts of this paper. A logic model is "a plausible and sensible model of how a program will work under certain environmental conditions to solve identified problems" (McLaughlin and Jordan, 2010, p. 56). A program can be viewed as a hypothesis: if implemented as planned, we expect to reach determined results. Logic models help to make explicit the underlying assumptions that are made when a program

---

[9] Bryson and Patton (2010) defines stakeholders as "as individuals, groups, or organizations that can affect or are affected by an evaluation process or its findings" (p. 31).

is designed: the relations between the resources (inputs) used, the activities (processes) executed and the intended outcomes (results).

One important lesson for logic models is the distinction between outputs and outcomes. Outputs are the products, goods, and services provided to the public. It is what a program delivers to achieve intended outcomes. Together with the inputs and the activities they compose the *program structure*. On the other hand, outcomes are the concrete changes or benefits that are expected to result from the implementation of the program and the outputs delivered. Programs can have multiple and sequential outcomes – short-term, intermediate and long-term outcomes – composing the *outcomes structure*.[10] Pervading these *internal* factors that ought to be incorporated into the logic of the program, there are the *contextual* factors, those aspects that can affect program's outcomes positively or negatively but are not under control of program staff (McLaughlin and Jordan, 2010).

Having a logic model in place is important to develop program performance measures for the different components of the model and then evaluate the causal relationships among them. To make clear the distinction and the relationship between performance measures and evaluation it is important to stress that "performance measurement describes levels of performance in relation to some standard and is typically a univariate measure, whereas program evaluation enables the explanation of why certain levels of performance were observed and is thus multivariate, using a number of performance measures to support the explanation" (idem, p. 72). Monitoring systems, by themselves, do not tell much about cause-effect relationships nor identify observed results as the impacts of a particular program. However, when the underlying program logic is clear and reasonable, the data generated by monitoring systems can be interpreted in terms of results of the program (Poister, 2010).

In an evaluation process if you do not have a clear logic model, the evaluation team and program staff should consider building it. In many occasions, it is a matter of making explicit some underlying assumptions. But even if you have a logic model already designed, it should be considered if the model corresponds to the program actually

---

[10] The distinction between outputs and outcomes is made in the Brazilian literature about program indicators, between *product indicators* (measure the outputs) and *result indicators* (measures the outcomes. The intermediate and long-term outcomes are classified as impact and because of that we have *impact indicators*. However, the distinction is more about the terminology used than of actual content. What is important is the reasoning behind program's logic model: the distinction between outputs and outcomes and the existence of multilayered outcomes.

implemented, or if there is some difference between design and implementation that can compromise the evaluation.

These two possibilities are frequently found in public policy. The hurry to implement polices reduces the time for planning. Sometimes the logic model can be clear to program managers, but may not be explicit or shared among program staff. The duration of the program and changes in program staff may lead to unregistered changes in program implementation that deviate from the original model. All of these factors have to be considered. If we only evaluate the improvement of a determined indicator that is supposedly affected by the program, we really do not have a lot to know if or how much the implementation of this specific program contributed to it. Unfortunately, this is the main practice in program evaluation in the context of Rio de Janeiro.

3.3 *Internal and External evaluation and Independent evaluation criteria: how governmental agencies can evaluate each other?*

This section address one institutional feature that affects evaluation: is the evaluation conducted internally or externally? Considering the characteristics of the Network of M&E being implemented by the government of Rio de Janeiro, it is important to be aware of the advantages and handicaps of a horizontal network, without a defined hierarchy between evaluating and evaluated agencies. To speak properly, it is imperative to be aware of the possibilities and limits of the strategy of requiring evaluation throughout the government.

Thus, it is not the case to defend a particular approach, but rather seeing how we can maximize the advantages of an internal evaluation and minimize its disadvantages. Internal evaluation is typically conducted by a team reporting to the donor or the implementing agency, while external evaluation is conducted by entities or individuals outside the donor or implementing organization (Imas and Rich, 2012).

The closeness of an internal evaluator is its advantage and disadvantage. Being part of the implementing team, an internal evaluator knows more about the program, its problems and experiences of success, and can probably address more pertinent questions while conducting an evaluation. On the other side, they can be too close to see clearly and may not recognize some solutions that can be viewed from outside the program. They can also be more subjected to pressures from the inside, and the conclusions of the evaluation may have less credibility. External evaluators probably have to make an extra-effort to

learn about the program, but they usually are viewed as more objective and more credible, as they are independent from the final decisions about the program. Yet, an external evaluation in not the guarantee of credibility, since evaluating and evaluated teams can have prior ties (idem, 2012).

What differs between internal and external evaluations is mainly the response to the independent evaluation criteria. An independent evaluation is defined as "an evaluation carried out by entities and persons free of the control of those responsible for the design and implementation" (OECD 2002, apud Imas and Rich, 2012, p. 32). By this criterion, internal evaluations are not independent, and serve more of a learning purpose than an accountability function, typically performed by external evaluations.

In the case of a governmental structure, it can be difficult to classify as internal or external an evaluation conducted by an agency over another when they belong to the same branch of government. For example, take the case when the Planning Secretariat of Rio de Janeiro evaluates a program of the Education Agency. Both of them are part of the executive branch of the government. There is no hierarchy between them, or a delegated authority to evaluate one other. In this case, implementing agency and evaluating agency are not the same (an external evaluation aspect), but both respond ultimately to the same donor (an internal evaluation feature). However, we think that this structure is likely to produce evaluations more external than internals. First of all, although they respond to the same donor this figure is typically present in private organizations. In the public sector, even if they share the origin of the resources, one agency is not reporting to the donor of the other to give feedback for a specific donation. Second, in the case of Rio de Janeiro, agencies tend to be very apart. They can cooperate between themselves, but the specialization and power disputes within the government tend to build a well-defined separation between the agencies, and even a distrust.

That is the present scenario that the Network of M&E of Rio de Janeiro has to face: we have to deal with a structure that is likely to produce an external evaluation, sharing the disadvantages of it (less knowledge of the program implementation) but without some benefits. When made by an agency that belongs to the same branch of the government, an evaluation could not have the credibility that it potentially would have if it was made by an external agency formally delegated to do it, i.e. does not meet the independent evaluation criteria. This is the result of the lack of defined hierarchy between agencies. In this picture, the evaluation is subjected to the existence of cooperation between agencies, without rules or institutional incentives stimulating or regulating it.

That is why the network is aimed at establishing cooperation between agencies in order to enhance the capacity and attractiveness of internal evaluations.

A network environment has to deal with an additional challenge: implementing monitoring activities within a network is much harder than in a single organization. Different members (or agencies) can disagree on the definition of goals and the ways to measure performance and evaluate outcomes. Also, differences of commitment to monitoring activities and expertise in evaluation can vary (Poister, 2010). This is typically the case in governmental programs, that usually involve different agencies in the implementation stage, and there is a sharp distinction between evaluator and evaluated.

Considering that, probably it is the case we should consider an *alternative approach to internal and external evaluations: the participatory evaluation approach.* The participatory approach differs from the internal and external evaluations as it tries to deemphasize the distinctions between evaluator and evaluated. Evaluators and members of the implementing agencies work together to design, conduct and implement evaluations.

As the Planning Secretariat, by itself, has some barriers (lack of personnel, expertise, authority) to play the role as an external evaluator, what can be made in a short/medium term – and that is the objective of the M&E Network – is to enhance the capacities of different agencies to perform their internal evaluations, acting as a partner in the process. Here, we came back to the importance of making the evaluation useful to program managers and staff: internal evaluations under this participatory approach tend to answer questions more about learning than accountability.

One important resource to internal evaluations that agencies typically already have is some data collected to assess the performance of a program. Most programs have some kind of an administrative records. If well designed, these data can be sufficient for further evaluation. But most of the times, additional information will be needed to measure the outcomes of the program. The evaluator has to see if the data are incomplete, inaccurate, overly aggregate or badly defined. Also, the data required may be found in different administrative records owned by different agencies, which will require some coordination effort (Hatry, 2010).

Many times, the collection of the data is viewed as a bureaucratic task to address some formal obligation. This perception tends to produce more and more useless data to evaluate. To make data useful we should try to convince mamagers that its collection is important not only because of some *obligation*, but rather that the program itself has a lot

to gain with it. Therefore, it is important that we use a task already done – collection of data – while enhancing the quality of it, discussing what should be measured and the way to do it.

Defending a learning approach does not mean to ignore the importance of accountability, but rather highlighting the specificities of this network. Besides that, accountability can certainly be enhanced at the end of the process, by accumulation of rich data that can be subjected to further assessments. And the structure within government has its accountability agencies that are somehow overly centered in an auditing approach. What can be pursued is the reduction of the gap between evaluation and auditing, discussed ahead.

3.4 *Auditing and evaluation: a broken continuum?*

Auditing and evaluation have their specificities and commonalities that suggests that the best arrangement between them is a balanced combination of both, without highlighting one as opposed to the other. Unfortunately, this is not the case in Brazil, where audits are hugely weighted over evaluations. As we have been talking about the *evaluation tradition*, we should say a few words about the *auditing tradition*.

Audits have an investigative, financial and accounting orientation, and address whether one program did what was supposed to do it and whether the money was spent according to rules and regulations. Auditing and evaluation can be viewed as giving different but complementary information about accountability and results. Especially in the case of performance audits – audits concerned with relevance, economy, efficiency and effectiveness – there are some overlaps between auditing and evaluation, with evaluation focusing on policy analysis and outcomes and auditing focusing on internal controls and management systems. Audits are centered on normative questions, on compliance with requirements, while evaluation focuses on descriptive and cause-and-effect questions. Both of them can provide relevant information to the design, implementation and remodeling of programs (Imas e Rist, 2012, p. 23-25).

In many countries, there is a widespread image of public auditors as *watchdogs*. This is certainly the case in Brazil. Audits are viewed as excessively worried with the detection of bad experiences, following a *problem-oriented* approach. The results of audits are more often a collection of examples of what not to do than a guide to success (Bourn, 2007). Of course, when this is true, auditors have to work to develop approaches

capable of changing this bad image, without abandoning the important function of detecting problems.

Yet, this is a valuable lesson for those focusing on evaluation. In a broader sense, evaluation also detects problems and positive results. Being pragmatic, the danger is that every effort to stimulate evaluation can be viewed as a way to reinforce controls and punishments, as auditors supposedly do. In a tradition of distrust of auditing, a network of M&E that aims to encourage evaluation has to constantly explain and emphasize other advantages of evaluation besides this *negative accountability* side. Of course, we are not saying that evaluation should not focus on accountability, nor that auditing is indeed an activity tied to the detection of problems. What we are defending instead is a strategy to enhance evaluation in a context dominated by a bad image of audits. And as we defend that the two activities should be balanced and are not mutually exclusive, evaluations can helps auditing: auditors could see if programs that are not having the intended results are also having problems in terms of fulfilling to regulations.

In the US, at the federal level, auditing and evaluation are centered in the Government Accountability Office (GAO). As we discussed, after the Government Performance and Results Act (GPRA) of 1993, evaluation activities gained more space within GAO. The change of the name in 2004 from General Accounting Office to Government Accountability Office is a symbolic expression of this reorientation. In Brazil, at the federal level, the Federal Court of Audits (Tribunal de Contas da União - TCU) plays a similar role. It is in charge of financial and operational audits. The last ones are focused on activities closer to evaluation: assessments of efficiency, efficacy and effectiveness. However, there is a widespread perception among public administration that this is not the dominant activity of TCU. Navigating the main pages of TCU's website, we can find evidence that reinforce those views. Words like *control* and *supervision* are hugely more present than *evaluation*. And being the agency responsible for the external control of federal government (as GAO) there are various complaints made about the lack of integration between TCU's auditors and the staff of programs that they are evaluating.

At the state level of Rio de Janeiro, the State Court of Audits (Tribunal de Contas do Estado – TCE-RJ) has similar functions to those of TCU. However, TCE is typically responsible for financial audits, and program evaluation is not really one of its preoccupations. Maybe even more than TCU at the federal level, TCE is not traditionally viewed with good eyes by other governmental agencies. So, the pendulum between

evaluation and auditing in the state of Rio de Janeiro is on the side of auditing, at least at the level of the external control. As we have been discussing, evaluation is made in a non-systematic way depending on each program manager.

Considering this picture, the network of M&E has a considerable challenge: how to reestablish the balance between evaluation and auditing without a delegated authority and dealing with reaction predisposed to reject any external interference. We hope that throughout this paper we have pointed out some useful insights from the literature in order to help in fulfilling this task.

4. CONCLUSION

This paper was approached as a *working-paper*, questioning how lessons from the literature about program evaluation can inform the design of a governmental network of monitoring and evaluation. We have focused here on some factors that could be called *infra-structural* that address some institutional aspects that are related to evaluation activities. Basing that choice there is a strong belief that a public policy's success cannot be grounded only in terms of good intentions and/or duties. Addressing the importance or obligation to perform program evaluation is not enough if it is not accompanied by incentives to actors to comply with the targets stipulated by government. That is why we discussed the importance of institutional incentives/obligations, while focusing also on making evaluations useful to a state's agencies.

To follow this utilization-focused approach does not mean to ignore accountability purposes of evaluation. For many readers, it can be a little bit strange that the main argument for evaluation was not *addressing results to make the use of public funds more efficient and evaluate their allocation*. At this point, we must stress that we completely agree that evaluation should, indeed, be defended as a way for accountability: showing to citizens the results of governments' actions. However, accountability arguments, especially in the case of Brazil, can acquire an overly negative/punitive connotation. It is unfortunate to see that, for many programs' managers and staff, this kind of requirements are viewed as getting in the way of their daily jobs. That is why we discussed some aspects that are important to making evaluation useful, to show that evaluation can have gains for the programs and the people involved with them.

One of the main ideas underlying this paper is the full integration of evaluation in the life cycle of public policies. Evaluation is not an ex-post activity and should be integrated since the beginning of the design of a program. Most of the topics discussed here – formative evaluation, evaluability assessment, logic models – reflect this point. This focus was not only a matter of theoretical choice, but rather a consequence of the initial stage of the M&E Network. Being in its beginning we should start from the basis of evaluation process, paving the way to improve future evaluation skills.

We also addressed some additional institutional features of the network. Considering its present resources we defend that a participatory approach aimed at improving internal evaluation of government agencies should be pursued, with SEPLAG acting as a partner enhancing evaluation capacities. Also, we should pursue bringing the

evaluation-auditing pendulum more to the center. This means to stimulate regular evaluation activities that are not directly related to the audits already performed by the agencies. We strongly believe that accountability of governments' actions has a lot to gain if we address programs' performance and results, not only compliance to rules and regulations.

Some argue that one of the challenges to develop an evaluation culture within public service is the lack of analytical capacities: public servants enters public service and are trained to do it, not to analyze it. Also, the nature of the bureaucracy within public service tends to stimulate compliance with rules and procedures and avoid innovations. (Behn, 2009). We do not completely agree with the choice of words: it is probably more accurate to say that considering the present routines within governments, most public servants are not faced with the incentives to develop evaluation/analytical skills. Nevertheless, this points to an important topic: building an evaluation network implies a major shift in the public service culture. It is a long term goal that we should seek by setting the basis for evaluation in the short/medium term.

Of course, the path to achieve this long term goal is not an easy one. We have discussed some historical backgrounds and saw that despite improvements, we always have to face challenges. A recent article of the American Evaluation Association (AEA) argues that, despite some experiences, evaluations still have been sporadic and inconsistent (AEA, 2009). The conclusions of AEA are consistent with the approach followed here: that evaluation should not be viewed as an ex-post activity, but rather as an integral part of program management at all stages.[11]

Since we have discussed the foundations of a Monitoring and Evaluation Network in the State of Rio de Janeiro, we have now to look forward to the next steps. One first move is to turn the focus to the dissemination of evaluation techniques and methodologies among governmental agencies. This will be a major effort, because it requires some skills that are scarcely found among agencies' officials.

Second, another crucial point is that even though we followed a pragmatist approach here – what can be done with the current contextual factors and resources – this should not imply a renunciation of the effort to change these factors. To put in another way, the network should try to gain support and strengthen the institutional background behind it, fighting for establishing some kind of formal evaluation units in a state's

---

[11] The article has some relevant concrete recommendations for improving evaluation within government. We highlighted some of them in an extract in the Appendix.

agencies. Another move can be establishing connections between the government's resources and evaluation, creating a way to improve allocation of funds. Some kind of regulation is needed, but an important lesson from this paper is to avoid a punitive/mandatory approach, and always have in mind ways to make evaluation useful and truly effective.

APPENDIX

Extract of AEA (2009), p. 3, 9-10.

Recommendations
We recommend that each Federal agency adopt the following framework to guide the development and implementation of its evaluation programs:

Scope and Coverage
- Invest in a body of evaluative work that covers public programs and policies throughout their life cycle and utilizes evaluation as a tool for improving programs and assessing their effectiveness
- Evaluate Federal programs and policies in a manner that is appropriate for program stewardship and useful for decision-making
- Build into each new program and major policy initiative an appropriate framework to guide the conduct of evaluations throughout the life of the program or initiative
- For existing programs, assess what is already known and develop evaluation plans to support future decision making

Management
- Assign senior, experienced officials to administer evaluation centers or coordinate evaluation functions at appropriately high levels of government agencies
- Prepare annual and long term evaluation plans to guide decision-making about programs
- Provide sufficient and stable sources of funds to support professional evaluation activities
- Coordinate and communicate about evaluation efforts across agencies with overlapping or complementary missions
- Develop written evaluation policies across and within Federal agencies that can guide evaluation efforts and help assure quality
- Assure that evaluation units and staff receive high-level, public and consistent support for their evaluation efforts

Quality and Independence
- Develop and adopt quality standards to guide the evaluation functions
- Promote the use and further development of a robust set of appropriate methods for designing programs and policies, monitoring program performance, improving program operations, and assessing program effectiveness and cost
- Safeguard the independence of the evaluation function with respect to the design, conduct, and results of evaluations
- Preserve and promote objectivity in examining program operations and impact
- Transparency
- Consult closely with Congress and non-Federal stakeholders in defining program and policy objectives, operations to be assessed, and definitions of success
- Disseminate evaluation findings relating to public accountability to policy makers, program managers, and the public at large
- Create clearinghouses for sharing information about effective and ineffective program practices

Based on the general principles discussed in the previous section, for the **Executive Branch** we propose two organizational frameworks as desirable for supporting the practice of evaluation within Federal agencies:

Centers of Evaluation. Agencies could establish one or more Centers of Evaluation to provide stable organizational frameworks for planning and conducting evaluations. The organization of such centers should be such that every program of the agency falls within the purview of one of the centers with respect to program evaluation. The heads of these evaluation centers should report directly to the senior executive of the organizational component in which they reside. Each of these centers would:
- have stable budgets sufficient to plan and carry out an appropriate level of program evaluation over several years
- strategically plan a body of evaluation work for the agency, each operational component, or program group for which it has evaluation responsibility
- hire professional evaluators or engage consultants or contractors with the diversity of skills necessary to plan and execute (or procure) independent evaluation studies
- publish the results of evaluations related to public accountability of the programs within their jurisdictions
- consult with agency program and budget offices, and, in concert with the agency's legislative liaison office, with the Congress in the development of evaluation plans
- issue policies and procedures to guide the conduct of evaluation work within its purview, including guidance on appropriate methods for conducting both formative and summative evaluations
- share information about effective programs and evaluation methods across government agencies
- promote and facilitate the ongoing training and professional development of evaluators within the center

Evaluation Coordinators. Agencies that choose to scatter their evaluation offices, associating small evaluation offices with individual programs or small collections of programs, should appoint one or more senior officials who will be responsible for coordinating the strategic planning of evaluation activities across the agency or across larger agency components or groups of programs. Such officials need not necessarily supervise the production of evaluations. Their role would be one of coordination, advice, and facilitation. Evaluation coordinators would:
- advise the agency head or senior officials on matters pertaining to evaluation
- ensure that each program operating division or major program group of the agency has a current annual evaluation plan
- promote, facilitate, and coordinate the development of evaluation plans for programmatic issues that cut across agency lines
- facilitate the preparation of evaluation budgets
- establish appropriate standards, frameworks, and procedures for evaluation activities within the agency
- facilitate the development and efficient and effective production of evaluation products
- facilitate the dissemination of evaluation reports related to public accountability
- share information about effective programs and evaluation methods
- promote and facilitate the ongoing training and professional development of evaluators within the agency

BIBLIOGRAPHY

Books, chapters and articles:

RESOLUÇÃO SEPLAG Nº 894/2013. http://download.rj.gov.br/documentos/10112/1587843/DLFE-61786.pdf/RES.SEPLAG894REGIMENTOINTERNOVERSAOFINAL.pdf

BRASIL, PORTARIA Nº 42/1999.
http://download.rj.gov.br/documentos/10112/779362/DLFE-46884.pdf/portaria_42MPOG.pdf

RIO DE JANEIRO, Secretaria de Estado de Planejamento e Gestão, Subsecretaria de Planejamento – SUBPL. *Manual de Elaboração do Plano Plurianual 2012 – 2015 – Rio de Janeiro*, 2011.
http://download.rj.gov.br/documentos/10112/780695/DLFE-46913.pdf/ManualdeElaboracaoPPA20122015.pdf

NEWCOMER, K. E., HATRY, H. P., and WHOLEY, J. S. "Planning and Designing Useful Evaluations" in Wholey, J. S., Hatry, H. P., Newcomer, K. E., *Handbook of Practical Program Evaluation*, Jossey-Bass Publishers, 3rd ed., 2010, p. 5-29.

SCHEIRER, M. A. *Expanding Evaluative Thinking: Evaluation through the Program Life Cycle*, American Journal of Evaluation June 2012, 33, p. 263-277.

PRESKIL, H. S., and CATSAMBAS, T. T., *Reframing Evaluation through Appreciative Inquiry.* Sage, 2006.

NEWCOMER, K. E. "Evaluating Public Programs" in Wholey, J. S., Hatry, H. P., Newcomer, K. E., *Handbook of Practical Program Evaluation*, Jossey-Bass Publishers, 3rd ed., 2010, p. 555-573.

BEHN, R. D. *PerformanceStat as a Search for Strategic Evidence.* APPAM, 2009.

IMAS, L. G. M. and RIST, R. C. *The Road to Results: Designing and Conducting Effective Development Evaluations.* The World Bank, 2009.

NEWCOMER, K. E. "Assessing Program Performance in Nonprofit Agencies" in *The International Handbook of Practice-Based Performance Management*, Sage, 2008.

BARBIER, J. C. *Devising and Using Evaluation Standards: the French Paradox.* Centre D'Étude de L'Emploi, Document de Travail, 23, 2003.

GOVERNMENT ACCOUNTABILITY OFFICE (GAO), *Evaluations Help Measure or Explain Performance*, 2010.

WHOLEY, J. S. "Exploratory Evaluation" in Wholey, J. S., Hatry, H. P., Newcomer, K. E., Handbook of Practical Program Evaluation, Jossey-Bass Publishers, 3rd ed., 2010, p. 81-99.

JUVENILE JUSTICE EVALUATION CENTRE (JJEC). *Evaluability Assessment: Examining the Readiness of a Program for Evaluation.* Program Evaluation Briefing Series, 6, 2003.

MCLAUGHLIN, J. A., and JORDAN, G. B. "Using Logic Models" in Wholey, J. S., Hatry, H. P., Newcomer, K. E., Handbook of Practical Program Evaluation, Jossey-Bass Publishers, 3rd ed., 2010, p. 55-80.

BRYSON, J. M, and PATTON, M. Q., "Analyzing and Engaging Stakeholder" in Wholey, J. S., Hatry, H. P., Newcomer, K. E., Handbook of Practical Program Evaluation, Jossey-Bass Publishers, 3rd ed., 2010, p. 30-54.

POISTER, T. H. "Performance Measurement" in Wholey, J. S., Hatry, H. P., Newcomer, K. E., Handbook of Practical Program Evaluation, Jossey-Bass Publishers, 3rd ed., 2010, p. 100-124.

HATRY, H. P. "Using Agency Records" in Wholey, J. S., Hatry, H. P., Newcomer, K. E., Handbook of Practical Program Evaluation, Jossey-Bass Publishers, 3rd ed., 2010, p. 243-261.

BOURN, J. *Public Sector Auditing: Is it Value for Money?* John Wiley & Sons, 2007.

AMERICAN EVALUATION ASSOCIATION (AEA). *An Evaluation Roadmap for a More Effective Government.* Prepared by AEA Evaluation Policy Task Force, February 2009.

Websites:

http://www.ioce.net/en/PDFs/national/2012/USA_AEA_Survey.pdf
http://www.ioce.net/en/PDFs/national/2012/Brazill_BMEN_Presentation.pdf