

**Understanding and Applying Innovation in Evaluation in the Canadian Federal
Government Context**

ADMN 598: Master's Project
August 11, 2014

Tyler Toso
School of Public Administration
University of Victoria

1. EXECUTIVE SUMMARY

The purpose of this paper is to summarize current knowledge related to innovation in program evaluation in order to inform future thinking within the Canadian Federal Government so that federal government evaluators in Natural Resources Canada can continually improve their practice through exposure to these ideas. This project achieves this purpose by identifying what innovations are occurring in program evaluation literature and practice while discussing some of the benefits and drawbacks to these innovations. This paper examines some of the barriers and limitations to implementing innovations. Lastly it includes lessons learned regarding the implementation of innovations within federal government program evaluation.

Evaluation divisions in the federal government such as Natural Resources Canada are responsible for producing program evaluations that provide information on the relevance, effectiveness, efficiency, and economy of departmental policies, programs, and initiatives (Brown, 2011, p.3). This is mandated through Treasury Board Policy. The performance of these evaluation divisions is assessed yearly by the Treasury Board of Canada Secretariat in regards to how well they achieve criteria set out in the guidance for the Management Accountability Framework (MAF). In the fiscal year 2012-13 one of the criteria for these performance measures asked the evaluation divisions to demonstrate their use of innovation in evaluations (internal government document: Capital Assessment Survey, 2012). In order to help address this requirement this project was commissioned by Environment Canada and was later supported by Natural Resources Canada when the Client professional relocated to another government department.

Information for this project was gathered through several methods including a literature review, key informant interviews and a document review. A literature review was conducted to establish a broad understanding of program evaluation and current innovations occurring in this field. The research focused on identifying recent trends in evaluation methodology as well as tools that were being used in different stages of evaluations, i.e., in the planning, conducting, or presenting of evaluation findings. This research helped inform the interview question design. After presenting the initial findings of the literature review to the client, the client requested that the researcher focus the interview design towards discrete innovative tools rather than emphasizing innovative methodologies.

The literature review provides a basic understanding of program evaluation including a definition for innovation as it applies to program evaluation. The literature review covers recent trends that have been identified, as well as any novel tools or approaches that are being explored, particularly over the last 5 years (since 2009). It investigates the established body of basic methods for conducting evaluations and was used to help provide focus for interviews, as well as for later use in providing some comparison with interview findings.

The purpose of the interviews was to identify successful innovations, limitations and the barriers to innovations, as well as to learn more about how innovations can be practically incorporated into evaluations. The interviewees were chosen for this study based on their experience and familiarity with hands-on practices, methodologies and tools that are used in evaluations within the federal government context to sufficiently answer the questions being asked of them. These interviews were used to help inform the research as to what is occurring in evaluation more

broadly, as well as to help identify current tools and trends and to learn from their experience with innovative tools. Furthermore, the interviews were used to identify successful innovations, limitations and the barriers to innovations, as well as for advice on how to incorporate innovations into evaluation.

The literature review and findings from the interviews demonstrate that there are methods being developed and used that offer program evaluators options that can help to better address stakeholder needs. These options are not without drawbacks and there are barriers and limitations that would need to be taken into consideration when attempting to utilize these findings. Successful innovation occurs within a context and it is the evaluation manager who is closest to this setting and therefore well positioned to make this goodness of fit assessment. Included are four lessons learned for directors of evaluation to help facilitate innovation within their evaluation division:

1. The federal government context of innovation in evaluation presents several challenges to being innovative including an exacting policy, limited capacity and limited resources.
2. Geographic Information Systems (GIS) is an innovative tool that has potential to add value to some evaluations for example through proximity analysis or the geographical visual representation of outcomes. Innovations in data display and presentations of evaluation results, social networking and social media and Big data analysis offer limited value for most evaluations within the federal government context although there may be potential benefits of these tools in certain instances. Furthermore, this research does not support a strong need for innovation in evaluation generally. An innovation should be invoked when there is clear value that can be added and it is likely to succeed in providing this value.
3. Director level support of innovation and professional development for evaluators could contribute to innovation as evaluators who have exposure to innovative tools and methods may gain the knowledge and skills required for implementing innovations that could add value to an evaluation.
4. Communication with Treasury Board Secretariat Center of Excellence in Evaluation (TBS CEE) to understand the requirements related to innovation can help directors to properly meet the criteria used in the Management Accountability Framework (MAF) assessments. Through accurately striving towards the MAF assessment criteria directors of evaluation can ensure they have done diligence to receive acceptable MAF ratings.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	3
TABLE OF CONTENTS.....	5
2. INTRODUCTION.....	8
3. BACKGROUND, METHODOLOGY AND CONCEPTUAL FRAMEWORK.....	9
3.1 Background.....	9
3.2 Conceptual and Theoretical Framework.....	9
3.3 Methodology.....	10
4. LITERATURE REVIEW	13
4.1 Basics of Program Evaluation.....	13
4.1.1 Definition of Program Evaluation	13
4.1.2 Purpose and Benefits of Program Evaluation.....	13
4.1.3 The Logic Model	13
4.1.4 Key concepts: Evidence, Triangulation, Causality & Validity	14
4.2 Innovation	15
4.2.1 Innovation in General	15
4.2.2 Definition of Innovation in Evaluation.....	15
4.2.3 Need for innovation	16
4.3 Evaluation Trends in Literature	17
4.3.1 Theory-Based, Realist Evaluations, Contribution Analysis or ‘Causal Evaluations’ ..	17
4.3.2 Evaluation Capacity Building.....	19
4.3.3 Evaluation Use and Utilization-Focused Approach	19
4.3.4 Value for Money and Program Resource Utilization	20
4.4 Specific Evaluation-Related Tools in the Literature.....	22
4.4.1 Innovations in Presentations of Evaluation Results and Data Visualization.....	22
4.4.2 Photovoice	24
4.4.3 Geographic Information System (GIS).....	24
4.4.4 Big Data	26
4.4.5 Social Networking and Social Media	26
4.4.6 E-Learning and Webinars	27
5. FINDINGS FROM INTERVIEWS	28
5.1 Interviewee Background Information	28
5.1.1 Position Title and length in position.....	28
5.1.2 Length with organization.....	28
5.1.3 Position Responsibilities.....	28
5.2 Innovative approaches to plan, design or carry out an evaluation.....	29
5.2.1 Cluster evaluations	29
5.2.2 Project Management Tools.....	30
5.2.3 Theory Based Evaluation (TBE)	30
5.2.4 Video	30
5.2.5 Photovoice	31
5.2.6 Qualitative Data Analysis Software	31
5.2.7 Collaboration with Internal Corporate Services	31
5.2.8 Online Software.....	32
5.2.8 Tool Templates.....	32
5.2.9 Evaluability Assessments (EA) and Risk-Based Planning Approach (RBPA).....	33

5.3 Innovative Tools	33
5.3.1 Data Display and Presentation of Evaluation Results	33
5.3.2 Geographic Information System (GIS) in Evaluation	34
5.3.3 Social Networking and Social Media in Evaluation.....	35
5.3.4 Big Data in Evaluation	35
5.4 Sources of Innovative Ideas in the Evaluation Division.....	36
5.4.1 Monthly Team Meetings or learning circles.....	36
5.4.2 Other Sources of Division Innovation Generation	37
5.5 Unimplemented Innovations.....	38
5.5.1 Unimplemented Contribution Analysis.....	38
5.5.2 Unimplemented (Failed) Qualitative Data Analysis	38
5.5.3 Unimplemented Geographical Information System (GIS).....	38
5.5.4 Unimplemented Theory Based Evaluation (TBE).....	38
5.5.5 Unimplemented Innovations in the Private Sector.....	39
5.6 Barriers or Limitations to Innovation	39
5.6.1 Limited Resources	39
5.6.2 The Effort Required to Innovate.....	39
5.6.3 Risk Averse Culture.....	40
5.6.4 Resistance to change.....	40
5.6.5 Failure and the Management Accountability Framework	40
5.6.6 TBS Policy on Evaluation (2009).....	41
5.6.7 Policies on data display (Web Standards for the Government of Canada).....	41
5.6.8 Contracting Policy	41
5.6.9 Limited Knowledge and Exposure to Innovations	42
5.6.10 A Director or Director General who is an auditor	42
5.7 Requirements for Better Innovation Implementation	42
5.7.1 Internal Capacity:	42
5.7.2 Culture and Support.....	43
5.7.3 More Freedom and Flexibility.....	43
5.7.4 Improved Collaboration.....	44
5.8 The Ongoing Need for Innovation.....	45
5.8.1 Innovation to Address Resource Constraints.....	45
5.8.2 Innovation to be More Useful for Clients and Improve Evaluation Quality	45
5.8.3 Innovation to Stay Relevant with Change and to Overcome Challenges.....	45
5.9 Additional Thoughts and Suggestions about Innovation in Evaluation.....	46
5.9.1 How Evaluators Defined Innovation	46
5.9.2 Innovation for the Sake of Innovation.....	46
5.9.3 Innovation as Ongoing Continuous Improvement.....	46
6. DISCUSSION	47
7. LESSONS LEARNED.....	51
7.1 The Context of Innovation in Evaluation.....	51
7.2 Innovative Tools and the Need for Innovation	51
7.2.1 Geographic Information Systems (GIS).....	51
7.2.2 Data display and Presentations of Evaluation Results.....	52
7.2.3 Social Networking and Social Media	52
7.2.4 Big data.....	52

7.2.5 Need for innovation	52
7.3 Director Support and Professional development	53
7.4 Communication with TBS CEE regarding MAF	53
8. CONCLUSION.....	54
9. REFERENCES CITED.....	55
10. APPENDICES	59
10.1 Appendix 1 – Validity.....	59
10.2 Appendix 2 – Overview of the Three Generations of Contribution Analysis	60
10.3 Appendix 3 – Contribution Analysis Process	61
10.4 Appendix 4 – Interview Instruments.....	62
10.5 Appendix 5 – Operational-Efficiency Analysis.....	68
10.6 Appendix 6 – Dashboard	69
10.7 Appendix 7 – Wordle (Word Cloud)	70
10.8 Appendix 8 – Interactive Phrase Tree.....	71
10.9 Appendix 9 – List of Data Visualizations.....	72
10.10 Appendix 10 – Data Visualizations by Complexity	73
10.11 Appendix 11 – Online learning Modules.....	74
10.12 Appendix 12 – Hype Cycle.....	75
10.13 Appendix 13 – Example of GIS.....	76

2. INTRODUCTION

The purpose of this paper is to summarize current knowledge related to innovation in program evaluation in order to inform future thinking within the Canadian Federal Government so that federal government evaluators in Natural Resources Canada can improve their practice through exposure to these ideas. This project will achieve this purpose by identifying what innovations are occurring in program evaluation literature and practice while discussing some of the benefits and drawbacks to these innovations. This paper examines some of the barriers and limitations to implementing innovations. Lastly it includes recommendations regarding the implementation of innovations within federal government program evaluation.

This project draws on three sources to provide information for the above purposes. First this project investigates program evaluation literature, second it uses information provided by program evaluators working within the federal government evaluation context and third it looks at internal documents provided by program evaluation divisions. By using multiple lines of evidence this research project aims to connect theory and experience from practice to provide a richer understanding of innovation in evaluation when addressing this project's purpose.

This project will be formatted into ten sections. The first two sections are the executive summary and introduction. The third section contains three sub-sections: the background sub-section which will provide the historical genesis surrounding this project, the conceptual and theoretical framework and the methodology sub-section which will describe how this project was conducted. The fourth section is the literature review which contains four sub-sections: The first sub-section examines the basics of program evaluation, the second sub-section looks at innovation and the need for innovation in program evaluation. The third sub-section examines innovative methodologies in program evaluation and the fourth sub-section identifies innovative program evaluation tools. The findings from interviews are the fifth section of this project which identifies the background of the interviewees, innovative approaches and tools identified in the interviews and the sources and activities that generate innovative ideas. This section also discusses unimplemented innovations, barriers or limitations to innovations, requirements for better innovations, the ongoing need for innovation and additional thoughts and suggestions about innovation in evaluation. The sixth section is the discussion—which identifies some of the key aspects identified in the literature review and interview findings sections. This helps to inform the four lessons learned regarding the facilitation of implementing innovations within federal government program evaluation which is the seventh section. The eighth section contains the conclusion and is followed by the ninth and tenth sections which are the references cited and the appendices respectively.

3. BACKGROUND, METHODOLOGY AND CONCEPTUAL FRAMEWORK

3.1 Background

Federal Government evaluation divisions such as Natural Resources Canada are responsible for producing program evaluations that provide information on the relevance, effectiveness, efficiency, and economy of departmental policies, programs, and initiatives (Brown, 2011, p.3). This is mandated through Treasury Board Policy. The performance of evaluation divisions is assessed yearly by the Treasury Board of Canada Secretariat (TBS) in regards to how well they achieve certain criteria set out in the Management Accountability Framework (MAF). In the fiscal year 2012-13 one of the criteria for these performance measures asked the evaluation divisions to demonstrate their use of innovation in evaluations (internal government document: Capital Assessment Survey, 2012). In order to help address this requirement this project was commissioned by the evaluation division of Environment Canada and was later supported by Natural Resources Canada. The purpose of this project is to summarize current knowledge related to innovation in program evaluation in order to inform future thinking within the Canadian Federal Government so that federal government evaluators in Natural Resources Canada can improve their practice through exposure to these ideas.

3.2 Conceptual and Theoretical Framework

TBS has established standards for conducting federal government program evaluations. Most significantly this is set through the 2009 TBS policy on evaluation and the accompanying guidelines. Room for innovation is limited due to the standards, policies and directives surrounding federal government evaluations. A key requirement in the 2009 Policy is that all departmental direct program expenditures must be evaluated on a 5 year cycle (Policy on Evaluation, 2012). As promulgated by TBS the following five questions must be address in these evaluations:

Relevance

1. Continued Need for the Program
2. Alignment with Government Priorities
3. Alignment with Federal Roles and Responsibilities

Performance – Effectiveness Efficiency and Economy

4. Achievement of Expected Outcomes
5. Demonstration of Efficiency and Economy (Policy on Evaluation, 2012)

TBS has recently suggested that there can also be room for innovation in program evaluations (Theory-Based Approaches to Evaluation, 2012, p. 17).¹ The definition of innovation in evaluation used to assess tools and methodologies for this project comes from the Treasury

¹ “Theory-based approaches can be particularly useful in addressing issues one, four and five” (Theory-Based Approaches to Evaluation, 2012, p.17).

Board Secretariat's Centre of Excellence in Evaluation (TBS CEE).² They defined innovation in evaluation in this way:

“Innovation generally occurs when a new and novel design, approach or method is adopted (perhaps as a pilot project) and if successful has the potential to add value by addressing a user's needs. Examples of innovation might include the adoption of a newly developed project design or approach that reduces the overall time or cost to conduct an evaluation, the use of a method that improved the quality of evaluation results, etc.” (Capital Assessment Survey, 2012).

TBS CEE provides further clarification for innovation in evaluation by suggesting that it can occur in how information is “utilized by different users who may derive value from the innovation” (Capital Assessment Survey, 2012). Potential users include the Departmental Evaluation Committees (DEC), program managers, policy makers, evaluators, heads of evaluation, the public and other stakeholders. The understanding taken by this paper is that value added can occur by incorporating innovation into the planning, conducting and reporting of evaluations. With this in mind, this paper aims to identify examples of innovation that could be used within the federal government context to improve evaluation quality, improve efficient use of resources, and to create a more useful evaluation for the end users.

3.3 Methodology

Information for this project was gathered through several methods including a literature review, key informant interviews and a document review. A literature review was conducted in order to establish a broad understanding of program evaluation and current theoretical and methodological innovations occurring in this field. The research also focused on tools that were being used in different stages of evaluations, i.e., in the planning, conducting, or presenting of evaluation findings. This research helped inform the interview question design. After presenting the initial findings of the literature review to the client, the client requested that the researcher focus the interview design towards innovative tools rather than emphasize innovative methodologies.

All of the interviewees for this project were asked questions regarding known innovative practices in evaluation. In total there were 17 interviews conducted with interviewees working in evaluation mostly within the federal government context. The interviewees were mostly composed of evaluation managers although some were senior evaluators and a few with director level and higher management experience. All interviewees had close to 2 or more years in their current position. Most had at least 5 years of experience or more in this position with some being in their position for 10 years more. 13 of interviewees worked for Canadian federal government departments directly with 4 others working in evaluation for private sector evaluation consultant organizations, and or had affiliations with academic institutions. The interviews were conducted in May and June of 2014 and were mostly done in person with one done over the phone. Identifiers in the interview findings have been removed to keep the identity of each interviewee confidential.

² The Treasury Board TBS CEE has the role, “To oversee the 2009 policy. To provide advice, oversight, and guidance” (Shannon Townsend and Michael Paquet Presentation at a 2013 CES conference).

The interviewees were chosen for this study based on their experience and familiarity with hands-on practices, methodologies and tools that are used in evaluations within the federal government context to sufficiently answer the questions being asked of them. These interviews were used to help inform the research as to what is occurring in evaluation more broadly in order to help identify current tools and trends as well as to learn from their experiences with innovative tools. Furthermore, the interviews were used to identify successful innovations, limitations and the barriers to innovations, as well to learn how innovations in evaluation are being incorporated into practice. The Advanced Notification Email, The Invitation to Participate Email, the Interview Consent Form and Interview Guideline can be found in Appendix 4.

In addition to interviews, some internal documents from program evaluators working in the federal government context were shared with the researcher in order to provide examples of tools and methodologies being employed within the federal government context. Analysis of the available documentation was used to provide additional information related to innovative practices. From all of these sources qualitative content analysis was done to identify examples of innovation in evaluation and to discuss some of their successes and limitations. The following is a list of tasks that were completed in order to meet the objectives of the project:

- Researched trends, definitions, examples of innovation and or new practices, methods and tools that can be used to enhance the effectiveness and efficiency of planning, conducting and or reporting of evaluations
- Identified potential candidates for interviews
- Invited interviewees to participate and conducted the interviews
- Collected internal documents from evaluators within the federal government context
- Compiled and coded the research data
- Analyzed, discussed and compiled lessons learned regarding the content of the research data
- Presented project findings to the clients.

There were barriers and limitations to completing the aforementioned tasks. The first was the limited information on, ambiguous nature and interpretation of what constitutes evaluation practices as being ‘innovative’. There is a substantial amount of literature written on program evaluation standard practices, models, and methodologies. Identifying and distinguishing between what is standard and what can be considered ‘innovative’ may rest on professional judgment where and when no clear definition is delineated. To address this the literature review includes well established and referenced works in evaluation in order to establish what are considered standard practices in evaluations in order to contrast these standard approaches to what can be considered innovative. As well, this paper relies on the definition for innovation in evaluation as elucidated by TBS CEE and introduced in the conceptual and theoretical framework section.

Another barrier was the unwillingness or inability of interviewees to participate in interviews due to absences or scheduling conflicts. This was addressed by giving the opportunity for evaluators to meet within a two month window at a time and day of their choosing.

Interview data can present some limitations such as the possibility of memory gaps on the part of interviewees. As well the selection of interviewees may not be fully representative of federal government evaluation given that unfeasibility to interview every relevant evaluator. These limitations were managed by: carefully selecting a wide sample of interviewees to ensure that relevant and diverse perspectives are adequately covered by knowledgeable respondents, asking interviewees to provide evidence of examples to support the views they expressed and comparing interviewee responses and corroborating the interview findings with evidence from provided documentation and the relevant literature.

4. LITERATURE REVIEW

This literature review intends to provide a basic understanding of program evaluation and also includes a definition for innovation as it applies to program evaluation. This literature review will cover recent trends that have been identified, as well as any novel tools or approaches that are being explored, particularly over the last 5 years (since 2009). This literature review was used to help provide focus for interviews, as well as for later use in providing some comparison with interview findings.

4.1 Basics of Program Evaluation

This section of the literature review will present an understanding of program evaluation involving its most basic principles. The purpose of this description is to have a benchmark for what is considered standard practice in evaluations. This will also help the reader to understand the most basic evaluation components and allow the innovative trends to be put into context.

4.1.1 Definition of Program Evaluation

A commonly referred to definition for **program evaluation** comes from Patton, “Program evaluation is the systematic collection and analysis of information about program activities, characteristics, and outcomes to make judgments about the program, improve program effectiveness and/or inform decisions about future programming” (Patton, 1997, p. 6). Similarly, TBS gives this definition: “The application of systematic methods to periodically and objectively assess effectiveness of programs in achieving expected results, their impacts, both intended and unintended, continued relevance and alternative or more cost-effective ways of achieving expected results (Results-Based Management Lexicon, 2012).

4.1.2 Purpose and Benefits of Program Evaluation

Program evaluations can serve many purposes, but first and foremost “The purpose of an evaluation is to produce valid conclusions and recommendations based on research methods that conform to accepted professional standards” (Bamberger et al., 2006, p. 3). These conclusions are intended to help inform decision makers as well as stakeholders. How this information will be used can vary depending on the evaluation and its goals but common types are evaluations that either seek to improve the program, or identify if the program has achieved its goals, or both (McDavid, 2006, p. 22).

Perhaps the purpose of evaluation is best understood by listing its benefits. Program Evaluations can provide information about:

- Relevance to need
- Program operations
- Program strengths, weaknesses and issues
- Attributable impact
- Efficiency and cost effectiveness (McDavid, 2013).

4.1.3 The Logic Model

The logic model is an important tool for evaluators to identify, understand and evaluate a program. For a program, the **logic model** serves as its road map:

“It outlines the intended results (i.e. outcomes) of the program, the activities the program will undertake and the outputs it intends to produce in achieving the expected outcomes. The purpose of the logic model is to:

- help program managers verify that the program theory is sound and that outcomes are realistic and reasonable;
- ensure that the Performance Measurement (PM) Strategy Framework and the Evaluation Strategy are clearly linked to the logic of the program and will serve to produce information that is meaningful for program monitoring, evaluation and, ultimately, decision making;
- help program managers interpret the monitoring data collected on the program and identify implications for program design and/or operations on an ongoing basis;
- serve as a key reference point for evaluators in upcoming evaluations; and
- facilitate communication about the program to program staff and other program stakeholders” (Supporting Effective Evaluations, 2009).

Table 1. Basic Logic Model Components



(Supporting Effective Evaluations, 2009).

4.1.4 Key concepts: Evidence, Triangulation, Causality & Validity

This section briefly introduces three key concepts in evaluation—evidence, triangulation and causality—all of which contribute to strengthening an evaluation by supporting the results.

“**Evidence** is the essential core around which any program evaluation is built” (McDavid, 2013, p. 91). The rigorous design of an evaluation is intended to provide the justification for its recommendations. Although some of an evaluation’s recommendations are left in the hands of the evaluator it is through evidence that an evaluator makes tenable recommendations.

Triangulation is the confirmation of using three independent measures and through this, “uncertainty is greatly reduced” (McDavid, 2013, p. 109). This is a practice popular in social sciences research including program evaluation, this concept is also referred to as using ‘**multiple lines of evidence.**’

A program evaluation can seek to determine if a program intervention *causes* an observable change in outcomes. “The three conditions for establishing **causality**—(1) temporal asymmetry, (2) covariation between the causal variable and the effect variable, and (3) no rival hypotheses—are at the core of all experimental designs and, implicitly at least, are embedded in all evaluations that try to focus on program effectiveness” (McDavid, 2013, p. 135).

4.1.5 Validity

Validity is about designing research processes in such a way that will seek to eliminate rival hypotheses that could explain the outcomes as well as those that could corrupt the results of the evaluation. These validity concerns include those related to statistical conclusions, internal validity through assessing causal linkages, construct validity through ensuring what is intended to be measured is actually being measured and through external validity—ensuring that the

results can be generalized elsewhere (McDavid, 2006). For more information on validity consult Appendix 1.

4.2 Innovation

The previous section of the literature review laid out some of the most basic but fundamental concepts behind program evaluations. This section of the literature will explore innovation in general through examining several definitions of innovation as well as how it applies to evaluation. Also this sub-section will discuss the need for innovation in evaluation and its driving forces.

4.2.1 Innovation in General

The literature referred to innovation, “As a planned social change process, innovation is an idea, practice, or object perceived as new by an individual or any other unit of adoption” (Earl, 2002 as cited in Rey et al., 2012, p. 71). Furthermore, the literature offered a diffusion model of how innovation travels from inception to its realization, “...the innovation diffusion process passes from (a) first *knowledge* of an innovation, to (b) forming an *attitude* toward the innovation, to (c) taking a *decision* to adopt or reject, then to (d) *implementation* of the new idea, and finally to (e) *confirmation* of this decision” (Rey et al., 2012, p. 71). The literature also suggests that innovation is transfused through an outside or inside agent who dedicates his or herself to encouraging the change. This change also occurs within the context of the organization and is affected by organizational, sociocultural and political factors” (Orlandi, Landers, Weston, & Haley, 1990 and E. M. Rogers, 1995 as cited in Rey et al., 2012, p. 72).

4.2.2 Definition of Innovation in Evaluation

At a Canadian Evaluation Society Conference (CES), Bradley Cousins, a University of Ottawa professor described innovation as it pertains to evaluation in saying, “Innovation is the development of new values through solutions that meet new needs, inarticulate needs, or old customer and market needs in value adding new ways” (Bradley Cousins presentation at a 2013 CES conference). He based this definition off of a Wikipedia search and cautioned that in terms of defining innovation, “there is no solid set of answers” (Bradley Cousins presentation at 2013 CES conference). In addition, innovation can be found in “Alternative, new ways of conducting evaluations, occurring in: methods, analysis, presentation and governance” (Simon Roy and Francois Dumaine presentation at a 2013 CES conference).

Two representatives from TBS also spoke at this conference providing this definition for innovation as it related to evaluation: “A new or novel design, approach or method that...adds value by addressing a user’s need” (Michael Paquet and Shannon Townsend presentation at 2013 CES conference). It has been explicated by TBS that “Innovation occurs when a new and novel design, approach or method is adopted and therefore adds value by addressing a user’s needs. Examples of innovation might include the adoption of a newly developed design or approach that reduces the overall time or cost to conduct an evaluation, the use of a method that improved the quality of evaluation results, etc.” (Capital Assessment Survey, 2012, p. 13). Elaborating further, TBS goes on to state that “the term ‘user’ here is not restricted only to users of the evaluation results (i.e., departmental evaluation committees, program managers and or policy makers). ‘Users’ also includes evaluators and Heads of Evaluation who may also derive value from the innovation” (Capital Assessment Survey, 2012, p. 13).

In summary, some of the criteria for identifying **innovation** include that it:

- Improves quality
- Saves time
- Saves resources
- Uniquely addresses a user's needs.

4.2.3 Need for innovation

As elucidated in the sub-section on program evaluation basics there are already established program evaluation tools and theories in place. These standard methodologies can be sufficient to conduct a successful program evaluation. This section will look at some of the drivers of need for innovation in evaluation found within federal government program evaluation context.

Changes in TBS policy on innovation may contribute to the need for innovation, “In April 2009, Treasury Board of Canada unveiled a new Evaluation Policy that broadens the existing mandate to evaluate programs. A 5-year cycle is envisioned to evaluate all programs (or clusters of programs). Although no new resources have been budgeted to meet these requirements, the expectation is that deputy heads will allocate the resources within their budgets” (McDavid as cited in Gauthier, B. et al., 2010, p. 6). The full coverage of evaluation was to be implemented by 2013.³ This new and more demanding framework in which federal government evaluators now operate requires the need for a change to meet the expanded coverage of its evaluations.

As demonstrated in the methodology section, innovation has appeared as a requirement in TBS Management Accountability Framework assessments of evaluation divisions. Innovation in evaluation has also permeated itself into TBS discourse and has been presented as a way to do business in evaluation better. In a talk given at a Canadian Evaluation Society (CES) conference representatives from TBS quoted the great inventor Thomas Edison who once said “There’s a way to do it better, find it.” The representatives provided several reasons to be innovative; to:

- Reduce time/increase speed of turn around
- Reduce project costs
- Improve value/usefulness
- Strengthen analysis
- Improve ability to use results (Shannon Townsend and Michael Paquet presentation at a CES conference in 2013).

Two regarded contributors to the field program evaluation in the Canadian and federal government context, Francois Dumaine and Simon Roy, gave a joint presentation on innovation in evaluation. They explained that the drivers of innovation include the opportunity to use new technologies and incorporate knowledge from other disciplines and professions such as from Audit in economy and efficiency issues (Francois Dumaine and Simon Roy at CES conference 2013). Dumaine and Roy also pointed out some potential challenges and barriers to innovation:

- Lack of resources and time

³ For the actual policy see (Policy on Evaluation, 2012).

- Difficulties in the contracting Process (One example is that contractors cannot communicate with clients until the Statement of Work is signed)
- The risk-averse culture within the federal government
- Obsession for quantitative, evidence-based approach
- Over prescriptive policy framework (e.g., TBS) (Dumaine and Roy at a 2013 CES conference).

4.3 Evaluation Trends in Literature

This sub-section discusses several innovative trends in evaluation methodology.

4.3.1 Theory-Based, Realist Evaluations, Contribution Analysis or ‘Causal Evaluations’

Theory-Based Evaluation (TBE) methodology is not something new to evaluation theory (Coryn et al., 2011). However, the topic has reappeared recently in 2012 when a paper was published by TBS which recognized the Theory-Based method as a viable option in federal government evaluations. There are numerous articles published in the American and Canadian Journals of Evaluation, as well as the New Directions in Evaluation journal regarding this topic.

TBE can be referred to using several different terms depending on the commentator but still share a similar methodology.⁴ TBEs are a methodology where specific causal links within programs are tested for effectiveness. In principle, “A theory of change can be used to test—with evidence—the assumed causal chain of results with what is observed to have happened, checking each link and assumption in the process to verify the expected theory (Theory-Based Approaches to Evaluation, 2012, p. 2). Establishing causation is the unique facet of employing this methodology. One method of how TBE accomplishes this task is summarized below:

“Causality is inferred from the following evidence:

- The intervention is based on a reasoned theory of change: the results chain and the underlying assumptions of why the intervention is expected to work are sound, plausible, and agreed to by key players.
- The activities of the intervention were implemented.
- The theory of change is verified by evidence: The chain of expected results occurred, the assumptions held, and the (final) outcomes were observed.
- External factors (context) influencing the intervention were assessed and shown not to have made a significant contribution, or if they did, their relative contribution was recognized” (Coryn et al., 2011 as cited by Dybdal et al., 2011, p. 37).

The TBS article points out that causation itself does not need to be established through mechanistic causations as this is not something that can always be identified, “in these situations, seeking a clear “one-to-one” causation that can be wholly attributed to one mechanism (finding *the cause*) is not possible. Rather, the relevant evaluation question is: In light of the multiple

⁴ Petrosino refers to TBE as the “causal-model evaluation” for two reasons: “First, it drops theory from the lexicon and sidesteps some of the confusion associated with using that word...Second, it accurately describes what most evaluators and theorists mean: *the evaluation is testing the causal model of how the program hopes to achieve its effects*” (Petrosino, 2000, p. 5).

factors influencing a result, has the intervention made a noticeable contribution to an observed outcome and in what way? Understanding *contribution*, rather than ascribing attribution, becomes the goal” (Theory-Based Approaches to Evaluation, 2012, p. 4). Nor does causation need to be established through a counterfactual, “In theory-based approaches...the specific causal mechanisms, are tested. If these can be validated by empirical evidence, then there is a basis for making a causal inference” (Theory-Based Approaches to Evaluation, 2012, p. 4). It should be recognized, however, that approaches using causal inference in this way could be subject to confirmation bias (As suggested by Jim McDavid in 2014).

John Mayne’s 2011 article on Contribution Analysis further developed this branch of methodology by introducing a practical guide for applying theory-based evaluations in order to ascribe attribution. Mayne recognized some of the shortcomings of TBEs—particularly that “establishing causal links between interventions and discernible outcomes using textbook prescriptions for optimal research design is not always possible or even appropriate” (Cook, Scriven, Coryn, & Evergreen, 2010 as cited by Dybdal et al., 2011, p. 29). Mayne “took note of this conundrum, and in a series of seminal papers and articles proposed a novel approach for addressing the question of attribution that he termed ‘Contribution Analysis’ (Mayne, 1999, p. 3, 2001, 2008, 2011 as cited by Dybdal et al., 2011, p. 30).

Contribution Analysis (CA) is defined as “[a] specific analysis undertaken to provide information on the contribution of a program to the outcomes it is trying to influence” (Mayne, 1999, p. 3, 2001, 2008, 2011 as cited by Dybdal et al., 2011, p. 30). According to Mayne, CA is useful where it is impractical, inappropriate, or impossible to address the attribution question through an experimental or even a quasi-experimental evaluation design. Thus, Mayne’s objective is to provide an alternative and non-counterfactual way to address the attribution challenge in the context of evaluation” which “builds a case for reasonably inferring causality” (Mayne, 2011, p. 6 as cited by Dybdal et al., 2011, p. 31-32). More recently Mayne has added different levels of causation based on the relative strength of the evidence “three basic kinds of contribution story can be told, depending on the relative strength of the evidence are...a *minimalist contribution analysis*, a *contribution analysis of direct influence*, a *contribution analysis of indirect influence*” (Mayne, 2011 as referenced in Dybdal et al. 2011, p. 34). See Appendix 2 for the development of CA and Appendix 3 for the CA process.

Although “Theory-based approaches present a number of positive features...they are not a panacea for attributing results to programs” (Theory-Based Approaches to Evaluation, 2012, p. 26). Furthermore, one comprehensive study of TBE indicates that this approach may not be particularly useful for most evaluations (Coryn et al., 2011, p. 216).⁵ Among limitations that theory-based evaluations face are the unpractical costs and ethical implications when testing control and causation (Theory-Based Approaches to Evaluation, 2012, p. 8).

Petrosino lists some implications for program evaluations that use a TBE or Casual Models:

⁵ “In many of the cases reviewed, the explication of a program theory unmistakably was unnecessary, or almost an afterthought in some instances, and was not visibly used in any meaningful way for formulating or prioritizing evaluation questions nor for conceptualizing, designing, conducting, interpreting, or applying the evaluation reported” (Coryn et al., 2011, p. 216).

- Casual models more rigorous than traditional evaluation methods, therefore are likely to require more resources to conduct.
- Causal models should be used to answer the why question, if this is not part of the evaluation it may not be necessary to conduct one.
- Causal models should be used when the links between steps of the logic model are not black and white, as causal studies are not needed when the implications of actions are more direct effects of intervention.
- Rigorous, causal-model evaluations can offer more information on programs and their effects (Petrosino, 2000).

4.3.2 Evaluation Capacity Building

Evaluation Capacity Building (ECB) has been covered frequently in both CES and AEA published articles in the last 5 years. ECB was also discussed as a potential avenue of innovation at the 2013 CES conference (As presented by Bradley Cousins at a CES conference in 2013). ECB can be simply defined as “... the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine” (Stockdill et al., 2002 as cited in Dreolin et al., 2008, p. 39). There are more detailed definitions in the literature.⁶

ECB can help evaluation divisions in the long run, “The literature suggests that there is an important connection between evaluation utilization and evaluation capacity building within an organizational context, such that the more an organization uses their evaluation results, the more likely ECB efforts will continue over time” (Dreolin et al., 2008, p. 41). For this reason, ECB may be of interest to federal government departments who work with programs on a cyclical basis and have a vested interest in building this capacity to help make evaluations more useful and better utilized by clients.

4.3.3 Evaluation Use and Utilization-Focused Approach

It is no small wonder that, “The issue of evaluation utilization continues to be a primary concern of the field. Over the past 40 years, many scholars and practitioners have proposed steps that evaluators should take to maximize use of their work” (Vanlandingham, 2011, p. 85). That evaluation is both useful and used is key aspect of program evaluation; as put by Bamberger “The universal concern of evaluators is that their findings and recommendations are not used” (Bamberger et al., 2006, p. 3) Increasing evaluation use is certainly a reoccurring theme in evaluation literature and innovative methods have been discussed in recent articles. Michael Quinn Patton has been writing about this for several decades, most recently his 2008 publication that explains the term **Utilization-focused approach to evaluation** (UFE). This evaluation approach emphasizes focusing the evaluation towards intended users—the people whom are using

⁶ “Second is a more recent and comprehensive definition put forth by Preskill and Boyle (2008):

Evaluation capacity building involves the design and implementation of strategies to help individuals, groups, and organizations learn about what constitutes effective, useful, and professional evaluation practice. The ultimate goal of evaluation capacity building is sustainable evaluation practice—where evaluation members continuously ask questions that matter, collect, analyze, and interpret data, and use evaluation findings for decision-making and action. For evaluation practice to be sustained, organization members must be provided leadership support, incentives, resources, and opportunities to transfer their learning about evaluation to their everyday work. Sustainable evaluation practice also requires the development of systems, processes, policies, and plans that help embed evaluation work into the way the organization accomplishes its strategic goals and mission” (as cited in Dreolin et al., 2008, p. 39).

the evaluation—and the intended user’s specific information needs. Doing so—he argues—can “increase an organization’s capacity for evaluation utilization” (Dreolin et al., 2008, p. 42). Patton asserts that utilization-focused evaluation is a process for “helping primary intended users select the most appropriate content, model, methods, theory and uses for this particular situation” rather than a theory itself (Patton, 2008, p. 592). There can be advantages to working closely with users, “a study of state legislative evaluators finds that those that regularly meet with stakeholders and provide readily actionable products were considered by senior legislative staff to have more impact (Vanlandingham, 2011, p. 85).

Some of the literature on evaluation utilization was cross-fertilized with similar concepts that come out of ECB (Patton, 2008, p.592).⁷ One study suggests “that there is an important connection between evaluation utilization and evaluation capacity building within an organizational context, such that the more an organization uses their evaluation results, the more likely ECB efforts will continue over time” (Dreolin et al., 2008, p. 41).

4.3.4 Value for Money and Program Resource Utilization

Value for money is not a new concept in public administration or even in reference to a question in an evaluation “All policies on evaluation, since the first in 1977, have required evaluators to consider some aspect of resource utilization as part of their evaluative assessment” (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p. 2). It is mentioned here as it has developed substantially in the literature in the last decade. Included in this is the 2012 TBS CEE article where it explains that “Value for money appears throughout the evaluation and audit literature. Treasury Board of Canada has recently defined a value-for-money “tool” as addressing two general questions: (a) Program relevance —Are we doing the right thing? and (b) Program performance — Are we achieving value?” (Treasury Board of Canada Secretariat, 2006 as cited by Mason et al., 2007, p. 3).

There are a number of different approaches to measuring and determining value for money. “**Cost-Benefit analysis** (CBA) refers to analytical approaches that seek to monetize all the costs and benefits related to a program and compare their net present values (used to compare benefits of real or potential programs)” (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p. 12-13). As a tool CBA is used to identify, “If the discounted present value of benefit exceeds the discounted present value of costs, the program or project should proceed” (Mason et al., 2007, p. 7). A limitation for federal government evaluators attempting to use is the challenge of “quantifying and monetizing the costs and benefits” (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p. 12-13).

On the other hand, **Cost Effectiveness** (CEA) refers to the comparative assessment of costs per ‘unit’ of outcome” (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p. 12). CEA “calculates the cost of producing a unit of *net outcome*. The term “net” indicates that the evaluator has controlled the external influences on outcomes and estimated the exact relationship” between the program and the changes towards the intended” (Mason et al., 2007, p. 11). Mason argues that “CEA will continue to have advantages over CBA

⁷ “From a collective perspective, utilization-focused evaluation repeatedly addresses organizational culture and climate in building an organization’s capacity to think evaluative. In this vein, Patton notes that evaluation itself constitutes a culture, making all evaluation practice ‘cross-cultural.’” (Patton, 2008, p. 592).

because of costs of execution and conceptual simplicity” (Mason et al., 2007, p. 23). Mason goes on to conclude that “If government is serious about results based management, then CEA needs to be a forethought, not an afterthought” (Mason et al., 2007, p. 23).

Cost utility analysis (CUA) compares the utility of a program in light of costs in order to contrast the usefulness of an intervention with the costs. **Operational-efficiency analysis** (OEA) Assessing the cost of specific outputs in relation to alternatives. Examples of these analytical approaches are included in Annex 5. Other Approaches to Value for Money includes “newer and innovative approaches (such as qualitative cost-utility analysis and testing implementation theories).” These emerging tools are believed “to have the potential to provide alternatives where traditional approaches may not be suitable or feasible (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p. 13-20).

The advantage of using the Value for Money approach is its good fit for addressing Core Issue 5 of the TBS Policy.⁸ However, “Evaluations that focus solely on the outcome achievement of programs, without taking into account the utilization of program resources, provide incomplete performance stories” (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p. 5).

⁸ “Core Issue 5, Demonstration of Efficiency and Economy, requires that evaluations include an assessment of program resource utilization in relation to the production of outputs and progress toward expected outcomes” (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p. 2).

4.4 Specific Evaluation-Related Tools in the Literature

Moving away from general approaches and methodologies, the scope of this literature review section is narrowed down to the specific tools of evaluation that have appeared in evaluation literature focusing on the last five years.

4.4.1 Innovations in Presentations of Evaluation Results and Data Visualization

This sub-section emerged as a result of changes in how end users look at, comprehend and want displayed information.⁹ Along with technological developments there is an increase in the variety of ways to display information to stakeholders. This sub-section of the literature review will explain the importance of data information display, provide a definition of data visualization, explain some benefits of emphasized effort in data visualization and provide several examples of these innovations that have been referenced in evaluation literature.

Conducting thorough research, effective analysis and drawing valid conclusions are all critical aspects to an evaluation. Reporting this information in a manner that is clear to all stakeholders may also be critical component in effective program evaluations.¹⁰ Take for example the anecdotal information below:

“In the field of survey design Christian and Dillman (2004)...found that one response option was selected more often, not because it reflected respondent opinions, but because unequal spacing between response options made it stand out from the others. Additionally, they found that large text boxes for open-ended responses led to longer answers and generated more themes during analysis. Thus considerations from graphic design like position, white space, symmetry, and emphasis influence survey data collection for evaluators”(as cited in Azzam et al., 2013, p. 17).

This example—although it comes from the field of survey design—demonstrates that a mundane and simply overlooked visual aspect can affect how a user interprets what is presented to them. Reflecting on the importance of the visual presentation of information the value of using new and innovative tools to accurately present data identified in evaluations well merits discussion.

The literature provides a definition of **data visualization** that includes three criteria, “Data visualization is a process that (a) is based on qualitative or quantitative data and (b) results in an

⁹ “We foresee that stakeholders will become more accustomed to interpreting, creating, and interacting with data visualizations. They may come to expect visualizations containing multiple perspectives on their programs and may require evaluators to produce such products as part of the evaluation. Advances in visualization software applications have become available for public consumption and use, often at no financial expense. For example, programs like ManyEyes, Tableau Public, and Gapminder now provide Web-based services that let users upload data for custom visualizations, allowing them to analyze and interact with the information at their own pace and with their own focus. This trend, when adopted by stakeholders, may require us to be more transparent about the data we collect, as demand for raw data increases.” (Azzam et al., 2013, p. 26).

¹⁰ “Evaluators have an obligation to present clearly the results of their evaluative efforts. Traditionally, such presentations showcase formal written and oral reports, with dispassionate language and graphs, tables, quotes, and vignettes. These traditional forms do not reach all audiences nor are they likely to include the most powerful presentation possibilities. In this article, we share our use of alternative presentation formats, undertaken to increase the utility, appeal, and salience of our work. “(Johnson et al., 2013, p. 486).

image that is representative of the raw data, which is (c) readable by viewers and supports exploration, examination, and communication of the data” (adapted from Kosara, 2007 as cited in Azzam et al., 2013, p.13). The benefits of data visualizations include that they may be able to “(a) to increase our understanding of a program, its context, and history; (b) to aid in the collection of data; (c) to conduct analyses of different forms of data; and (d) to communicate to a wide range of stakeholder groups” (Adapted from Kosara, 2007 in Azzam et al., 2013, p. 17).

When contemplating innovative data display tools considerations such as the purpose of the displays, audience, the available technology and the evaluator’s capabilities to display this information should be included (Lysy, 2013 p. 34).¹¹ It is said that good data visualization techniques will mirror good communication and “The best chart types for displaying quantitative data are often the simplest. Bar charts, line graphs, and scatterplots are as effective today as they were 15 years ago” (Lysy, 2013, p. 34). In looking at information in new ways both evaluators and stakeholders can create the opportunity for interaction with the data that leads to understanding and different ways of thinking and sharing insights (Henderson et al., 2013, p. 69). Ideally, these innovations will add richness and value to an evaluation project (Johnson et al., 2013, p. 501). By improving communication across stakeholder groups it is estimated that the evaluations will appear more relevant, salient and be more likely to be used (Johnson et al., 2013, p. 501-502).

A good example of providing users with key information is quantitative dashboards. These “can be used to create a visual representation of how a program is performing on multiple indicators at once. These dashboards can be viewed as a way to track program performance by centralizing critical performance measures into a single visual structure” (Azzam et al., 2013, p.19). Although the above example is used in a performance measurement context in an evaluation context such dashboards could be applicable to tracking progress towards a programs adoption of evaluation recommendations when reporting on evaluations. An example of a dashboard can be found in Appendix 6.

Other graphics can be created to communicate with external stakeholders and this use of **static imagery** is able to tell a story about an organization (Azzam et al., 2013, p. 24). There are multitudes of data display tools available, including “sparkcharts, heat maps, bubble charts, tree maps, and stack graphs” to name a few (Lysy, 2013, p. 34). One innovative infographic to display qualitative information found in the literature is a **Word Cloud** or **Wordle**—a computer generated image that graphs words used in a text, displaying them in a font size representative to its frequency of being used in a particular context; capturing discourse visually. An example of a Wordle can be found in Appendix 7.

Moving beyond the static presentations of data are the **interactive data displays** which are defined as: “any visualization that can be manipulated directly and simply by the user in a free-flowing manner, including such actions as filtering the data and drilling down into details” (Few, 2005, p. 8 as cited by Lysy, 2013, p. 42). Internet and wide spread technology can enable evaluators to “take advantage of new features such as interactivity, animation, and automation to make large complex data sets clutter free and approachable for lay audiences” (Lysy, 2013, p.

¹¹ “As is true in all reporting, evaluators should consider their audiences and their comfort with data when determining the best way to present” (Henderson et al., 2013, p. 69).

34-43). Through web applications evaluators have the potential to give stakeholders or wider audiences the ability to filter and manipulate data that has traditionally been accessible to analysts (Lysy, 2013, p. 42). The use of interactive visualization is not without its concerns as “interactivity gives the designer less control over the story being told” (Lysy, 2013, p. 34).

There are multiple applications identified in the literature that are available for exploration, including **Tableau Public**, IBM’s **ManyEyes**, **GeoCommons**, and **Google Fusion Tables** which could enable evaluators to create charts and maps that let the reader fluidly select, zoom, and filter this data (Lysy, 2013, p. 34). Other software referred to in the literature include GapMinder, Spotfire, and SAS’s J.M.P.”, which could “provide the evaluator with the power to create multiple interactive visualizations that can be used to highlight specific variables, drill down into subgroups, change the timeline, embed maps, and a host of other features” (Azzam et al., 2013, p. 19-26). One example of an interactive data visualization is an **Interactive Phrase Tree** which is similar to a Wordle in its function, however, it allows the viewer to interact with words and phrases in order to see what context the phrases were used in through zooming in to see where they lead. An example of an Interactive Phrase Tree can be found in Appendix 8. A list of additional data visualization suites can be found in Appendix 9.

4.4.2 Photovoice

A tool that has recently been contributing to program evaluations is Photovoice. Photovoice is a method that encourages participants to use their own pictures to share stories. The steps of Photovoice are detailed as follows, “Clients of a program are asked to take photographs about their experience in the program or about how the program has helped or changed things for them. Most often, the evaluation team will provide the cameras. The clients take the pictures, discuss them with the evaluators and or other clients, and write accompanying descriptions. Evaluators then analyze the photos and descriptions to understand how the program operates and affects its clients” (Bakker, n.d). This has been identified for its success with empowering the voice of marginalized groups and to capture and represent feelings from a group or individual that lacks a strong voice. One negative aspect of utilizing this approach is that it can be too powerful and results may dominate over other evidence (As presented by Simon Roy and Francois Dumaine at a CES conference in 2013).

4.4.3 Geographic Information System (GIS)

Geographic Information System (GIS) is both a data visualization tool and a potential evaluation tool that carries considerable promise for its use in evaluation, “Through this data visualization an evaluator can determine the level of community needs, available resources, and the potential contribution that a program can have within a specific community. This level of understanding can be gained early in the evaluation process and can significantly contribute to the development of appropriate designs and measures to inform future evaluative conclusions” (Azzam et al., 2013, p.16). GIS is a tool that “allows users to combine geographic information (e.g., streets, addresses, and school locations) with other types of data (e.g., demographic data, program satisfaction results, and outcome measures) to create multilayered visual maps that enable users to identify patterns or relationships between a program’s environment and its performance” (Azzam et al., 2012, p. 207-208). The advantage of using GIS is that it “offers a visual way of detecting patterns in data that may have remained unnoticed through other traditional methods of analysis” (Azzam et al., 2012, p. 208). Furthermore, “GIS has the ability to show many of the

social, economic, educational, and political structures that are embedded in communities. These structures are often needed to better understand the impact that programs have on society and the factors that help or hinder their success. (Hopson, Greene, Bledsoe, Villegas, & Brown, 2008 as cited by Azzam et al., 2012, p. 208). There are multitudes of potential applications of GIS in evaluation that are identified in the literature:

“Given the availability of data, an evaluator can create maps that pinpoint the distribution of job opportunities, educational services, health resources, the political affiliation of community members, and a host of other indicators to better understand the strengths and challenges in each area. These structural factors can also be mapped across time to show how the context has changed or evolved, and how that temporal change has affected other variables. For many programs GIS can reveal shifts in community demographics and resources across time that can help them anticipate and respond to change in a proactive manner” (Azzam et al., 2012, p. 208).

As well, “GIS could be used to select optimal program sites that are near populations who would benefit the most for the services” (Azzam et al., 2012, p. 208). Although there are some examples of GIS being used in evaluations,^{12,13} despite this potential the literature also indicates that “there are currently few real-world examples of GIS application in evaluation” (Azzam et al., 2012, p. 210). For an example of a GIS visual display please refer to Appendix 13.

There are some limitations to using GIS. Technical expertise is required for the effective use of this software. These skills can be grasped through an introductory course aimed at teach social scientists on how to use GIS as well as its capabilities (Azzam et al., 2012, p. 220). Furthermore, there is the potential for privacy issues when the data can be geo-coded to specific locations that could be used to identify individuals.¹⁴ However there are methods such as aggregating this data

¹² “Application and utility of geomatics to evaluation: Mapping vulnerable populations and community based services in Canada *Mrs. Kate-Lynn Duplessis, PHAC - Mrs. Kara Hayne, PHAC*”...“By mapping and analyzing program data through a spatial lens, geographic information systems (GIS) can be a powerful tool to support evaluation of program relevance and reach. This presentation will demonstrate the application and utility of GIS as an emerging evaluation methodology. A GIS case study was completed on two national children’s programs funded by the Public Health Agency of Canada, namely the Community Action Program (CAPC) for Children and the Canada Prenatal Nutrition Program (CPNP). The purpose of this project was to determine the location of at-risk populations in Canada, and to assess through spatial analysis, whether CAPC and CPNP projects are reaching these populations. This project is expected to help inform program evaluation and future program directions. Considerations and lessons learned for building spatial analysis into other areas of evaluation design will be discussed” (Canadian Evaluation Society 2011 Conference Program, 2011, p. 50).

¹³ One example from the literature demonstrates how GIS helped to identify tobacco billboards in relation to low socioeconomic neighborhoods and within view of schools. “The process of mapping the data revealed that over 80% of billboards were located in low socioeconomic neighborhoods, and more surprisingly the mapped tobacco billboards were within view of 87% of existing elementary, junior, and high schools. These patterns and subsequent findings would not be as easy to detect or as striking without the use of visualizations that GIS offers” (Azzam et al., 2012, p. 8).

¹⁴ “...if the actual spatial data are accessed, the evaluator may inadvertently open the possibility of additional information on the household individual being gathered by linking the original georeferenced data with data from other sources. With the rapidly growing availability of fine-grained demographic, social and economic data with which to link, this is an escalating concern. Moreover, the ever expanding access to GIS maps and data on the web makes such information mining easier and easier” (Azzam et al., 2012, p. 222).

into groups there by displaying information in larger groups such as tracts or neighborhoods (Azzam et al., 2012, p. 223).¹⁵

4.4.4 Big Data

Although **big data** has no set definition it can be loosely defined to describe such data sets that are “so large and complex that they become awkward to work with using standard statistical software” (Snijders et al., 2012, p. 12). It has been pointed out in the literature that “...the rise of digital and mobile communication has made the world become more connected, networked, and traceable and has typically led to the availability of such large scale data sets” (Rainie & Wellman, 2012 as cited by Snijders et al., 2012, p. 1). Big data has been gaining traction in the evaluation community, “Big Data has become the key core competency of the information age, what distinguishes the current era,” (O’Reilly Media Research Director Roger Magoulas, 2009, minute 0:47 as cited by Kistler, 2011 as cited in Smith et al. 2011, p. 568). The advantages to using big data, are that it taps into already existing data, and therefore could save resources in primary research (as cited in a PowerPoint presentation from Goss Gilroy Inc.). Accessing these complex data sets to identify useful information and lines of evidence for an evaluation requires skills and software beyond traditional statistical approaches. Sufficient to say, analyzing the plethora of information in datasets would require a skilled technician and the complexity of intricately describing these processes is outside of the scope of this paper. Big Data is also being explored in the context of social media and will be re-visited in the next section.

4.4.5 Social Networking and Social Media

Incorporating social networking features into evaluations can occur in several ways. Social Media can be used passively through observation of content for research up to and including the use of it for big data analysis. Social networking can also involve the direct engagement of evaluands through social media such as promoting surveys online through social networks or as a way for evaluators to form their own online social networks (As cited from Canadian Immigration Canada evaluators at a CES Conference in 2013). The available literature suggests that both of these methods are being experimented with by the evaluation community (Kistler et al., 2013). There are social networking analysis tools that were identified in the literature; amongst others are Netdraw and NodeXL (Kistler et al., 2013). Analyzing social media in the context of big data is touted for its potential to collect data in sensitive, remote locations, or obtaining views from marginalized groups and or others who may benefit from the anonymity (As cited in a PowerPoint presentation from Goss Gilroy Inc.) Government use of social media may be a sensitive topic due to privacy concerns and has been scrutinized in media coverage (Janus 2014, Bernier 2014, Kwong 2014).

¹⁵ “In order to both display and share GIS project information, but protect participant privacy, two techniques can be employed: edited geographic display and data aggregation. The first technique is the simplest; retain the data in its original form, but alter its display to obscure the ability to pinpoint exact location. This usually involves “coarsening” the units of display. For example, if participant addresses are geocoded as points, display the location with a corresponding, but more generalized unit, like census tracts or neighborhood...Aggregation is another form of protecting privacy and can usually be accomplished through areal aggregation or point aggregation. A real aggregation is accomplished when an a real unit is defined and then the grouped data of all or some of the cases located within that area are provided. This is the aggregation method used by the U.S. Census when reporting demographic variables by block groups or census tracts (Azzam et al., 2012, p. 223).

Social networking is also being used horizontally between evaluators as a way to share evaluation related information, “New tools have expanded the avenues for peer-to-peer exchange and discussion of ideas and expertise and provide a glimpse of what it is like for a practicing evaluator. AEA catalogs over 40 blogs by and for evaluators” (Smith et al. 2011, p. 569). EVALTalk, is a social networking function provided through the American Evaluation Association and reportedly “boasts over 3,500 subscribers” (Kistler as cited in Smith et al., 2011, p. 568). Federal government employees also have access to internal social network through GCConnex and GCPedia. These sites can only be accessed through Government of Canada networks and are a way for government employees to share information. Within GCConnex the Centre for Excellence in Evaluation has established its own group known as EvalConnex which is made for evaluators specifically to share ideas and evaluation related information through this forum.

4.4.6 E-Learning and Webinars

The literature discusses how e-learning and webinar tools can be applied to innovative evaluation purposes, “...interactive Web-based technologies have been used to complement the supports described...in order to build and enhance evaluation capacity in these organizations” (Sundar, 2011, p. 95). The literature describes one organization which used these web based learning tools to help teach and develop staff and relevant stakeholder’s evaluation skills (Tse, Choi, & Leung, 2008 as cited in Sundar, 2011, p. 97). Web-based learning tools have also gained attention for contributing to culture of learning (Preskill & Boyle, 2008 as cited in Sundar, 2011, p. 106). The organization designed an introductory learning module that can be completed in three hours that will help users learn about the steps of an evaluation and how evaluation findings can be used to improve the effectiveness and efficiencies of programs. This method is also intended to help programs have the necessary tools to self-evaluate. When surveyed the users of this module indicated an increased knowledge of evaluation.¹⁶ These web based modules can be used separately or in conjunction with more interactive forms of web based learning including webinars. Webinars offer the benefit of interacting with peers to share and learn about others experiences relating to evaluations as well as to get connected with other colleagues in the same field (Sundar, 2011, p. 104). Depictions and explanation of this Module’s purpose can be accessed in Appendix 10.

There were some identified limitations of using web-based technology for the purpose of facilitating learning in evaluation specifically in regards to a lack of comfort with the web-based platform being used and general technical challenges that occurred when trying to use such resources (Sundar, 2011, p. 105). Web-based technology has been cited as impersonal and there can be difficulty in fostering discussion when compared to face-to-face interactions (Sundar, 2011, p. 106). For these and other reasons, the literature recommended that the use of web-based technology and resources “...should not be used as stand-alone tools, but rather should be seen as one important element in a range of tools and strategies for building evaluation capacity” (Sundar, 2011, p. 105).

¹⁶ “It is clear that these innovative techniques were effective in communicating evaluation-related information, given the 87% average increase in level of knowledge reported by users, after having completed only the first module” (Sundar, 2011, p. 104).

5. FINDINGS FROM INTERVIEWS

This section gives an account of the information learned from the interviews with 17 interviewees. The purpose of these interviews was to identify successful innovations, limitations and the barriers to innovations, as well as to learn more about how innovations can be practically incorporated into evaluations. The interviewees were chosen for this study based on their experience and familiarity with hands-on practices, methodologies and tools that are used in evaluations within the federal government context to sufficiently answer the questions being asked of them. These interviews were used to help inform the research as to what is occurring in evaluation more broadly and help to identify current tools and trends and learn from the experiences of interviewees with innovative tools. Furthermore the interviews were used to identify successful innovations, limitations and the barriers to innovations, as well as for advice on how to incorporate innovations into evaluation. Identifiers in the interview findings have been removed to keep the identity of each interviewee confidential.

5.1 Interviewee Background Information

This sub-section gives a description of interviewees' background in evaluation.

5.1.1 Position Title and length in position

The 17 interviewees were mostly composed of evaluation managers, some were senior evaluators and a few with director level and higher management experience. All interviewees had close to 2 or more years in their current position. Most had at least 5 years of experience or more in this position with some being in their position for 10 years more. The greater majority of the interviewees worked for Canadian federal government departments with others working in evaluation for private sector evaluation consultant organizations, or had affiliations with academic institutions. The interviews were conducted in May and June of 2014. 16 of the interviews were conducted in person and 1 was done over the phone.

5.1.2 Length with organization

Almost all of the interviewees had experience in evaluation with multiple organizations but typically had more than 2 years with their current or most recent evaluation department or organization. Several interviewees also had prior experience with their organization as a junior or senior evaluator within the organization.

5.1.3 Position Responsibilities

All of the interviewees had the responsibility to manage evaluations from start to finish. Typical responses for most of the interviewees included:

- Planning, conducting, and reporting evaluations (Selecting methodologies, instrument design, data collection and interviews and evaluation reporting)
- Internal management functions including responsibility for a small team as well as their training and development
- Responsible for working with private sector or public sector counterparts in either hybrid teams or in other roles including developing or responding to the Terms of Reference for evaluation contracts.

- Responsible for presenting evaluation results to senior managers as well as to the Departmental Evaluation Committees or equivalents.

A few of the interviewees had stated that they:

- Contributed to the division’s overall evaluation plan.
- Provided performance measurement services and or advice to other programs.
- Developed evaluation templates and contributed to federal government evaluation policy, guidelines and training.

5.2 Innovative approaches to plan, design or carry out an evaluation

This sub-section will examine several of the innovative approaches identified in the interviews.

5.2.1 Cluster evaluations

Cluster evaluations were mentioned by several evaluators. **Cluster evaluations** refer to grouping programs with similar outcomes into the same evaluation (contrasted with evaluating each program separately). This type of evaluation usually looks at programs by how they fall under the Program Alignment Architecture (PAA)—a TBS management policy¹⁷—to aim at department specific strategic goals. The described advantages to this approach were that it was able to save time and costs to maximize use of resources and to help the division to reach their 100% coverage as prescribed by TBS policy.¹⁸ Also by looking at these programs as components in a larger whole the evaluators were able to help identify opportunities for senior management to efficiently allocate resources across these programs through comparative analysis in order to identify how to better reach the overall strategic goals. In line with this are the efficiencies from presenting only once to senior management and meeting with the same groups and stakeholders instead of multiple times for separate evaluations.

The drawbacks to using this method have been that programs—while similar at surface level objectives—may have considerably varying program design and delivery methods such that the same evaluation tools cannot be applied across the programs. The result of this drawback is a final report that excessively “uses many caveats” for the exceptions when statements cannot be used across the programs in the context of them being a whole. Difficulties also arise when interviewees are only able to speak to their specific component. These problems led one evaluator to state that cluster evaluations, “can be hit or miss in terms of success.” An identified prerequisite for cluster evaluations was that the individual programs within the cluster had to be functioning well. One evaluator pointed out that if there are large issues within the individual programs than it makes more sense to do individual program evaluations (to rectify these issues) instead of grouping it into a cluster. Some evaluators spoke critically of cluster evaluations by suggesting that it was sometimes being misused out of necessity to reach policy requirements regarding the 100% coverage and not for the purpose of conducting higher level analysis.

¹⁷ (Policy on Management, Resources and Results Structures, 2012)

¹⁸ One evaluator mentioned that TBS 100% coverage may have been a catalyst to re-examine how they do their evaluations and helped lead into the use of cluster evaluations.

5.2.2 Project Management Tools

Project management tools were used in order to be more effective and efficient in reaching timelines and using internal capacity. Teammate is a project management software tool that one evaluation division had borrowed from their counterparts in Audit. It required training to use but its implementation has helped the division to understand more about the resources that are going into evaluation projects so that they can plan accordingly. Good project management may also be attained without the assistance of specific software as one interviewee described the generation of generic timeline templates using Microsoft Word and Excel to be applied to typical evaluations. These tables help capture all of the tasks, processes and approvals involved using known timelines from past experience in order to “prevent the evaluation project from derailing.” By indicating the level of effort required for the evaluation (and the full time equivalent employee hours required) the division could apply this to the overall division’s evaluation plans to help identify and more accurately reach deadlines for all of their evaluations.

5.2.3 Theory Based Evaluation (TBE)

Theory Based Evaluation (TBE) was used in several evaluation shops. One evaluator explained how traditional approaches offered no consensus among program stakeholders (senior management and program managers) as to what the evaluation plan should look like for a particular evaluation that was being planned. One outstanding issue was that the program had no clearly articulated program theory. TBE offered the opportunity to translate the program theory and capture the intended outcomes of the program—which was a first for the program. TBE allowed the evaluators to indicate when certain program functions worked by highlighting the circumstances around this success. This evaluation was considered very successful due to the program’s appreciation and buy-in once the program theory had been depicted. Additional value added from this project was that the evaluator gained experience in this methodology and was able to apply this same method successfully to another program.

The identified drawback of this method was that the evaluation design did not strongly align with TBS policies regarding the 5 core questions. Despite this, the evaluators were able to map and align the evaluation to the TBS policy (after the fact) and it was accepted by TBS. As well, the expertise for this type of evaluation did not initially exist within the division and the advice of a consultant was required to help educate the evaluation team on how to properly conduct this method. Employing a new method also required considerable effort by the evaluator to get the initial buy-in for a new type of evaluation.

5.2.4 Video

Videos were used while conducting case studies in the field. One evaluator indicated that the evaluation shop did record interviews to get the recipients of the program to talk into the camera during the interview. The program was having difficulty in securing key funding for over 20 years. The idea behind using the camera was to capture a firsthand account of the program results to show the program impact to senior management and give them a new perspective. There were several benefits to this method. It clearly conveyed the value of the program to senior management and provided strong evidence which offered itself to better understanding of the program function and to better decision making. The result of the evaluation contributed to the support of key funding within a year.

5.2.5 Photovoice

Photovoice as was described in the literature review has been used in multiple departments as a case study collection method and presentation tool. Using Photovoice in an evaluation often required the use of consultants who are experts in these methods as well as at employing sensitive interpersonal data collection methods. It has been used to get recipients to open up about their experiences related to a program when dealing with client-sensitive issues. This method was also useful for extending evaluation reach as it provided access and participation of children and families; encouraging them to speak about their experiences and ultimately gave them a voice in the evaluation. One division hired a company to make a DVD using the results from Photovoice. The photos were dubbed over and music was included—later this DVD was shared by the program. The evaluator recounted how “through the use of photo voice, we actually were able to see how they lived and it was very good for the whole program area... The picture was worth 1000 words and you can see where the money went.”

Some limitations of Photovoice were noted in the interviews—that its use may be inapplicable or inappropriate for certain evaluations when compared to more traditional methods. The recruitment of participants in Photovoice was difficult and labor intensive for one division and it required travel to fly out to communities as well as compensation for the participants to recoup their costs while participating in the project. As an inexpensive alternative one division also uses photos taken by the evaluators during site visits. This offers the benefit of telling a strong story and can be done easily by the evaluator.

5.2.6 Qualitative Data Analysis Software

Some divisions had indicated that they had successfully used data analysis software (NVivo & MAXQDA) to help conduct their qualitative analysis and mentioned no drawbacks to this tool. On the other hand, one division had several setbacks when employing qualitative data analysis software. What this division found was that the results of the interviews may have not been homogenous enough to use this software. There were only 15-20 interviews done in the example and this was believed that a small sample may have contributed into the poor results generated from the software. To complete the evaluation the analysis was done manually.

5.2.7 Collaboration with Internal Corporate Services

Collaboration with other groups within the department has been used by several evaluation divisions. For some this novel approach is a reflection of improved central data systems within government.

By working with the Financial Division to extract financial data evaluators can access program information while reducing the burden on the program. This information can also be used for comparison against program’s own financial information which may or may not align and this can be looked into further. This internal services collaboration has been used to overcome the adversity of delays in information gathering and also allows the evaluator to have more information prior to the start of an evaluation which focuses the questions of the evaluation. One challenge to using the departmental financial data was that it required an initial investment to train evaluators on how to navigate the departmental financial systems.

Some evaluation divisions have been working with the centralized research divisions within their federal government department or with other divisions conducting research to extract data and statistics from studies and surveys that have been conducted previously. This can eliminate a duplication of effort and can reduce the survey burden of recipients who are subject to multiple studies. Besides raw data, other divisions conducting research may also have undiscovered insight and expertise related to the area being evaluated that can prove valuable to evaluators. In turn, some evaluation divisions have also been able to share their evaluation reports with research divisions to facilitate an information sharing network.

Through familiarity with each other's work evaluation and other internal services groups can make plans to work together on future projects and prevent future replication. An example given was to tap into the research and statistics from other divisions in order to answer TBS questions regarding program resource utilization. Another division in the department conducting research had already developed an allocative efficiency model. By using this model and mapping onto other evaluations the evaluation division has gained this corporate knowledge and can do it inside of their division.

5.2.8 Online Software

Using **online survey software**, such as Fluid Surveys has been used as an alternative to mail out surveys which were costly, labor intensive and for one division required the support of consultants. The user-friendly software has been simple to use and does not necessitate the support of a consultant.

One evaluator reported having issues when shifting to online surveys; they had faced challenges and technical issues including security and firewalls and other communication challenges that can occur when communicating across the internet. From the account given, they have gone back to using more traditional methods i.e., mail out surveys because they have found them to be more reliable in certain instances.

5.2.8 Tool Templates

Some examples of time saving or value adding templates mentioned were:

- a. For questionnaires
- b. For logic models
- c. For evaluation reports
- d. Post-evaluation lessons learned
- e. Client feedback surveys.

Lessons learned and information from client feedback surveys¹⁹ were gathered at the end of an evaluation and these are now being incorporated into the planning of other evaluations as well as being used at the renewal of the five year evaluation cycle to inform the next round of evaluations since the inception of 2009 TBS Policy.

¹⁹ These surveys asked questions such as whether the evaluation results were useful and timely etc.

5.2.9 Evaluability Assessments (EA) and Risk-Based Planning Approach (RBPA)

Evaluability Assessments (EA) and a **Risk-Based Planning Approach (RBPA)** were used by some of the evaluation divisions to determine program risks and the level of engagement required. The EAs involved meeting with some of the program staff well before the evaluation is scheduled to start—for example 6-12 months. Information gathered during these meetings can help to design the evaluation tools a head of time as well it can help the program to prepare for the evaluation. One example given was how a program began collecting performance data for the evaluation after the EA occurred.

Information ascertained from these meetings also feeds into the RBPA. Some factors that are considered for the RBPA include past program performance (as determined by previous audits or evaluations), whether or not the program's operations or strategic goals had changed sufficiently since the last evaluation and the program's visibility (public, media and or senior management attention). It is an innovative approach because it allows for flexibility to save time and resources.

5.3 Innovative Tools

5.3.1 Data Display and Presentation of Evaluation Results

Most interviewees spoke about reduced size of their final evaluation report from about 50-100 pages down to between 20-30 pages in length. One way this reduction is achieved has been by annexing the technical reports that support each line of evidence “The focus has shifted to using 2-3 few key points, instead of listing all of the information; it has taken a culture change to understand that we have to pare it (the evaluation report) down.” The advantage of the shorter reporting is that it is easier to digest for senior management. A shorter report saves costs such as the “html” costs to publish the report to the internet as well as translation costs.

One novel method for reporting evaluation results was the use of a very condensed matrix that includes specifically the findings, recommendations and conclusions. This method “leaves out a lot of the details of how the evaluation was conducted, which although interesting, are not necessarily required. It is very palatable for the reader.” The drawback to this method was that it is very linear and does not allow for a lot of analysis. For example, analysis across how some of the TBS core 5 issues being explored can affect each other cannot be captured in this type of a report. Matrixes were also used also used in other ways across divisions to condense information for the final evaluation report.

During a cluster evaluation an evaluator described using a process map of the program that they had created, “It was not strictly speaking a logic model, in terms of its rigidity.” The idea was not to use this to capture the program's logic but as a conversation piece to situate the interviewee in terms of what their role was in the processes within the Program Alignment Architecture (PAA). It was a tool that was designed to help learn and teach more about the program and offered a transformative process for the program. By virtue of asking these questions while using this tool—helped to bring the evaluand to a greater understanding and to think of themselves in the context of the larger whole and the PAA box that was being evaluated. Another evaluator who had used a similar approach with a process map had said that this can be a preferable method to logic models which are too rigid and can oversimplify internal processes.

One evaluator described some challenges to using innovative data displays. From this evaluator's experiences complex data displays could be confusing to senior management. As a best practice data display is kept very simple. As one evaluator put it, "Some of the design options actually remove meaning and are just wiz-bang graphs that look like innovation but are pure junk that detract from meetings." Including a lot of data display and visualizations also posed challenges for getting the final evaluation report published online. Due to government policies regarding display of information such as web standards for the government of Canada and display for the visually impaired all of the data displays require a detailed description of the graphic. One evaluator recounted a time where the department was sued for using a graph without sufficient description for the visually impaired. For one division that relied heavily on data displays writing these descriptions and having them translated was more costly than the division could sustain. As a result this division regretfully eliminates most graphs and charts from the final evaluation product and places these graphs in an annex to be made available upon request. The same challenges occurred in another division when required to include the long text description required for explanations. One example given was for an equation used in the evaluation, "it took a page of text just to explain it." As well, evaluators indicate that the standard practice for Departmental Evaluation Committee (DEC) is to involve little to no overhead data displays and just speaking notes to the evaluation report itself and follows a standard template.

5.3.2 Geographic Information System (GIS) in Evaluation

Several interviewees had used GIS in evaluations. It was spoken of highly for its ability to pictorially represent and illustrate data and allowing stakeholders to really see this data for the first time, "Mapped data is able to show a phenomenal image of how far or close things are." One example given was of how GIS was used was in the plotting of hazardous material sites across Canada to identify its proximity to roads and communities. Another example was a proximity analysis to look at family resource centers in communities across Canada to determine a 'remoteness factor' of some communities. These were mapped out to find out how far someone would need to travel to access these services. This information was used to help investigate what would be acceptable e.g. "2-5 hours to drive? 24 hours of travel? Or did they have to fly?"

One evaluator who readily used this technology said that it required specific knowledge and experience that may not be accessible within evaluation shops. It was by coincidence that this evaluator had experience working with GIS in a prior position and was able to apply the capabilities of this technology. This is supported by the anecdotal information provided by another division who was trying to incorporate GIS into an evaluation. Since the shop lacked the skills in house they have been trying to contract it out (which is a process that can lead to administrative delays). Another evaluator pointed out that the integrity of the information is very important when using GIS—to ensure for example—that transportation routes actually exist on land and not just in the maps. Another limitation to using GIS is that it requires all of the data to be geocoded to location which does not apply to all data sets. One interviewee also pointed out that in remote communities the postal codes—which are often how data is geocoded—can be extremely large making the observations broad. Other drawbacks mentioned were the costs associated with training and owning the software.

5.3.3 Social Networking and Social Media in Evaluation

Social networking was not used widely across the evaluators that were interviewed. When it was being used it was done rarely as an engagement tool and more commonly as a passive listener. Less direct than being used for actual evaluations—social networking was used amongst interviewees to share information with other evaluators.

The most direct way social media was used in an evaluation came from a second hand account of an evaluation shop that used social media to promote surveys in the context of an evaluation. The interviewee pointed out the limitations of promoting open surveys online, saying that that it has issues with representation and survey bias as well it can be difficult to control who is being surveyed.

Social networking is being used for general research related to evaluation as a way to look at opinions regarding a topic and was noted as a potential option for conducting media scans. One program explained how conducting a media scan on a particular aspect of a program led to new insights from a blogging site which contradicted some of the programs claims. This led the evaluator to probe into the issue further. A potential limitation is that social media may not be accessible within the government department's networks as was the case in one department when asked about this medium. Another division pointed out that only their communications branch was allowed to access social media. Several interviewees had pointed out that there are some privacy concerns with taking this approach "People can get worried and this is a hot topic. If anything goes wrong it's a big deal." Some evaluations that had been planning to use this approach had been put on hold until discussion takes place regarding guidelines or frameworks for how the government can use this data."

Several interviewees mentioned that as part of evaluations had created social networking maps for the programs being evaluated. These maps were used to indicate the strengths of relationships by using indicators such as contracts or other interactions with other departments or external groups. Through the internal federal government networks evaluators can also access GConnex and GCPedia for the sharing of internal practices and ideas as well as to open up discussions in the forums to help solve evaluation issues. Some interviewees had experience with this internal social networking and had used it as a way to find solutions to some issues regarding an evaluation. Some interviewees also pointed to EVALTalks which is hosted by the American Evaluation Association and is another access point for evaluation related social media and social networking.

Social networking was also being used in conjunction with big data which will be discussed in the next sub-section.

5.3.4 Big Data in Evaluation

Big data was not used by most of the interviewees. A lot of them admitted that either their department does not use large data sets or that the division typically does not work with large data sets. One evaluator who used big data in evaluations stated that it was done only occasionally and involved the use of contractors as this skill set was not available in house. An example given was a longitudinal study over a 30 year period to determine program need. Although the results were successful at demonstrating that there was no need for the program the

evaluation did not result in any major changes because of the political contexts surrounding the program.

An interviewee explained how their shop is combining social media and big data to feed into lines of evidence in evaluation. The advantage to using this data as a line of evidence is that this qualitative and quantitative data can already exist and be mined, “We (evaluators) spend a lot of time trying to figure out what people think and feel or to understand how they have been impacted by programs and government agencies; this is awfully expensive, big data taps into data that already exists.” Another identified need for using big data is when an evaluation requires very large outreach to access segments of Canadians. In some cases where you need to find out something broad or high level about Canadians—large data volumes may be required. Because the information already exists—big data was also cited as a way around survey burn-out. Big data and social media analysis was said to have the potential to access hard to reach groups, vulnerable subsets of populations who may benefit from anonymity. Again, some concerns over potential privacy issues were raised as a potential limitation to getting support for this innovation.

There were several limitations to using big data in evaluations. First, it was stated that for most standard evaluations this level of analysis may not be necessary. As one evaluator put it, “we recognize that there is a time and a place to use that type of method but that would not be useful all of the time.” Other challenges to using big data had to do with constant changes in policy and programs that occur typically every 5 years (i.e. funding is only allowed for 3-5 years). Major program and policy changes can distort frames of reference and the attributions that would be made are exposed to validity threats from the constant flux of other external factors. In support of this, one evaluator stated that, “Unless these programs were designed long term we can’t do this type of an evaluation. And there is no appetite in the federal government for this long term research.” One evaluator pointed out a potential methodological limitation of using big data, “In an evaluation you want the behaviors to be attached to an individual, in big data you are just tracking the number of times something was done and I am not sure how in evaluation this data would be used.” Another identified setback to using big data in evaluation that was mentioned was the current political status in regards to federally government supported research “a lot of the huge data sets from Statistics Canada—longitudinal surveys—are being scrapped; these could be useful for big data types of analysis...it is a calamity.”

5.4 Sources of Innovative Ideas in the Evaluation Division

This sub-section introduces sources and activities that interviewees identified that contributed to the generation of innovative ideas within their division. The sources and activities that are discussed are not linked to specific innovations instead they refer to the general sources for innovation generation.

5.4.1 Monthly Team Meetings or learning circles

Having re-occurring meetings for the purpose of sharing new ideas was fairly common amongst the interviewees. Loosely this took the form of a staff member volunteering to present a paper of the month as a way to keep in touch with new literature. It was pointed out by one evaluator that this does not necessarily result in innovative ideas specifically but it contributes to the culture of

sharing ideas. The benefits of this type of forum were captured well by an interviewee's comments:

“We had monthly evaluation practice meetings, these staff meetings were specifically devoted to a topic of advancing practices or knowledge of the division as a whole. Sometimes we looked at process, sometimes methodological, sometimes team members presented and sometimes we had guest speakers. Because of the group nature of the events it generates synergy and the setting is conducive to ideas...if we didn't do this some people's ideas would never be shared.”

As a limitation some evaluators pointed out that other priorities can put this activity on hold during peak business and that ideas discussed here do not necessarily get used.

5.4.2 Other Sources of Division Innovation Generation

Other methods of divisional innovation generation were also mentioned during the interviews and are comprehensively listed below. For additional information regarding these methods consult the associated footnotes:

- Deputy Minister and equivalents²⁰
- Directors²¹
- Innovation days²²
- The culture of the evaluation division²³
- Attending conferences such as CES annual events & Lunch and Learns (and debriefing the team on what was learned)²⁴
- Client Feedback Surveys²⁵
- Holding forums with stakeholders (not project specific but about direction)
- Evaluation retreats²⁶
- Learning from the private sector²⁷
- American Evaluation Association & CES articles and presentations

²⁰ In one division, the Risk Based Planning Approach was suggested by the Deputy Minister of the department.

²¹ “The director is pushing us to new ways of reporting.”

²² One contractor explained that, “We have 3 innovation days we can use each year where we are able to set aside time and search for innovative things to do.” As well the division has an innovation committee that meets every month and will approve innovative ideas and help figure out what the next steps should be. However, “the formal process does not always work and you have to use backdoors to get the innovation used.”

²³ “Any sort of innovations that we have attempted are for three purposes: to gain efficiency in regards to the full coverage, to make our evaluations more useful to senior management and to reduce the burden on the programs.”

²⁴ “When someone attends a conference, they will debrief the rest of the group so that the learning is shared with the group and it can spark ideas.”

²⁵ “We have client surveys and contractor surveys for evaluation quality assurance—these are used as one stream for decision making. We want to get feedback from them on how to do things better, it makes everyone's lives easier and hopefully (makes) better evaluations... We ask them whether our information was timely and relevant and whether it helped to make good decisions. It's a new process from 5-6 years ago and people are generally constructive with what they provide. The information goes directly do our director so we don't see the results directly but we get the gist of it.”

²⁶ The interviewee described a yearly retreat where the evaluation division talked about lessons learned from each evaluation.

²⁷ “We are risk averse (in government) so typically we will copy the private sector. They take the risk and we reap the benefits once it's tried and true.”

- Training (CEE, Canadian School of Public Service, University courses and Certificates in evaluation)
- Bi-annual Head of Evaluation Meetings
- GCConnex and EvalConnex
- EVALTalks

Some evaluators expressed the feeling that little to nothing was being done to encourage innovative ideas. This was captured well by one evaluator's comments, "Departmentally there is nothing being done to generate innovative ideas. Public servants are happy with that status quo. Government is highly risk averse. We (evaluators) are highly risk averse."

5.5 Unimplemented Innovations

This sub-section speaks to some of the unimplemented innovations that evaluators mentioned during interviews.

5.5.1 Unimplemented Contribution Analysis

The first example comes from one division that had attempted to include contribution analysis in an evaluation. The interviewee cited poor communication as playing a large role in the failed attempt to implement an innovation. The evaluator failed to get the programs' support for using this methodology as they were unable to explain how the theory behind contribution analysis would be used practically in the program evaluation. This was also a new approach for the evaluator and the evaluands did not want to be the pilot project. The end result was that the program evaluation was put on hold.

5.5.2 Unimplemented (Failed) Qualitative Data Analysis

One evaluation division has had challenges utilizing qualitative data analysis software due to the administrative barriers of getting the tool approved and licensed.

5.5.3 Unimplemented Geographical Information System (GIS)

One evaluator explained that for a particular evaluation the division had wanted to use GIS in an evaluation but lacked the in-house skill sets required to incorporate this tool. The administrative burdens associated with contracting polices has also contributed to delays in employing this method as of yet.

5.5.4 Unimplemented Theory Based Evaluation (TBE)

One evaluator tried to incorporate TBE into an evaluation as it seemed highly applicable but was unable to get the project off the ground:

"We were excited and did a lot of extra research and developed the preliminary project plan and explained why TBE would work well for this evaluation and for the program. However, it was shot down by the director. It was too complicated for the Director General and we (the evaluation team) were still learning and did not feel confident enough."

5.5.5 Unimplemented Innovations in the Private Sector

Lastly, from the perspective of evaluation consultants working in the private sector in support of government operations there are some innovations discussed but never implemented due to there being no need or demand for them written into Requests for Proposals (RFPs). Consultants expressed frustration at the desire to use innovative tools but had their hands tied by evaluation contracts and contracting policies from using them.

5.6 Barriers or Limitations to Innovation

This sub-section examines the reported barriers and limitations to incorporating innovation in evaluation as expressed during the interviews.

5.6.1 Limited Resources

Below is a list of the limits on resources that were identified as a barrier to innovations:

- Financial resources²⁸²⁹
- Time³⁰
- Internal capacity³¹
- Travel Budget³²
- Expertise³³
- Software and Licenses³⁴

5.6.2 The Effort Required to Innovate

An evaluator who successfully employed an innovative approach described the immense effort required to innovate in their division:

“It took a considerable initial investment to bring the evaluation team on board and to communicate the plan to stakeholders. It required first the support of my boss, and then my boss’s boss and then the program managers. It takes a lot of effort to do this investment. And you have to factor in the risk that it may not work. The consequences of failure—and this is just speculation—is that you might have to start the evaluation over; your credibility might be affected and if you want to try something new in the future they might say no.”

As well there are a lot of potential steps to getting an innovation off of the ground which becomes more complicated in relation to the complexity of the innovation, “To do something innovative you may have to find an expert, have to contract, you’re going to need to go through the investment plan to get approval.” The effort required to push innovations may be unappealing to evaluators “It takes a unique individual because it’s easier to go with the flow.

²⁸ “We do not do Cadillac evaluations, we just do what needs to be done and move onto the next.”

²⁹ “We only have enough to do the bare minimum”

³⁰ “Innovative approaches take a lot of time, a lot more time than you initially think.” And “Innovation requires an initial investment of time.”

³¹ Not enough staff to do more and the methodological, technical or soft skills may not be available within the division.

³² “There is not a lot of information in any one area, in Ottawa there is a lot, but you really need to get out there and there are limits to how much online is useful.”

³³ “Its hard to innovate straight from a text book.”

³⁴ (If required for an innovation tool).

Doing innovation is a challenge; it's a challenge to the established principals and the higher executives...it is a struggle." Since, much of the processes in federal government evaluation have already been established and internalized, innovations can be hindered by roadblocks and interviewees point out that when this is combined with little reward and a lot of risk it becomes a difficult path to choose, "there is no reward for being innovative". Apathy may play a role in not pursuing innovation as one evaluator pointed out that it is easier to keep to keep the status quo and keep meeting TBS minimum requirements.

5.6.3 Risk Averse Culture

"Public servants are risk averse and do not want to try things that are new in case they do not work out." As one evaluator points out, "Innovations require the ability to take chances, some managers are good at this, but it is almost an individual personality thing. But generally government evaluation is somewhat risk averse." An interviewee also mentioned that "A risk minimizing mindset restricts the mind in terms of using innovation." This suggests that a risk-averse culture may hinder innovation by eliminating it from thought as a potential option. Some evaluators defended the need for the government to be somewhat risk averse, "It makes sense, the government faces issues of public safety and health, and do you want them to take risks?" As well one evaluator noted that evaluations are paid for by public dollars and there is a responsibility to the tax-payer in terms of resource management.

Despite the typical risk aversion, one interviewee did not feel that there would be huge repercussions to attempting innovation within their division, "It may just be the thing that people are hoping for. Or if it doesn't work out it will just not be used in the evaluation and would have wasted time and money. We give it a try, and at the end of the day, if it does not work out we won't try it again, but at least we tried and we know."

5.6.4 Resistance to change

The challenge of overcoming resistance to change was also cited as a barrier, "People did not want to change the way they do things, they liked how it was done and believed it worked well...people do not want to change." Evaluators had given examples of resistance in all levels including directly in the evaluation division—with other staff or directors, with the program and with senior management.

5.6.5 Failure and the Management Accountability Framework

Management Accountability Framework (MAF) ratings are issued by TBS and are used to assess a division's performance. As described by one evaluator, "MAF ratings are the prime motive of the directors." An interviewee indicated that when it comes to innovation, "We have to be conscience of our MAF ratings." Several evaluators pointed out the adverse effects an innovation would have on their MAF ratings if the intended results failed to be achieved, "If we try something innovative and it does not work we will be penalized in our MAF ratings." One of the unimplemented innovations from the last section was attributed partly due to the potential a failure could have on MAF ratings, "Our DG was concerned whether our attempt at innovation would meet the requirements for that year. To do the innovation would have taken more time and we could not promise that it would not fail."

5.6.6 TBS Policy on Evaluation (2009)

As put by one evaluator, “There are standards, directives and a wide range of established practices. Whenever you have these it creates an environment in which it is challenging to be innovative.” Speaking directly to the 2009 policy, another interviewee pointed out that properly answering the 5 core questions dominates the evaluation and it would be difficult to add to that without some sort of flexibility, “An interview guide that needs to cover a long list of evaluation questions is long, you can’t push your key informant beyond an hour and a half, their quality of information starts suffering, so you would need to schedule a second interview, good luck with that” (it may be hard to get). Several of the interviewees had noted that it’s not that the policy is too limiting in itself, but that the guidelines are often interpreted too strictly and that there is a lack of flexibility in the federal government evaluation culture. One example illustrates this well:

“We had identified some interesting things in our field experiments and had our findings peer reviewed, when we wanted to incorporate it into the evaluation report a senior manager said no because it was not in the evaluation matrix and did not answer an explicit question in the framework, therefore we cannot use it. The evaluation matrix which was intended to serve as guidance had become a strait jacket.”

Similarly—in defense of the policy and as a critique of some fellow evaluators’ interpretation of the policy—one evaluator pointed out that often evaluators stick too close to the 5 core questions and are averse to explore other issues due to the strict interpretation of policy and had this to say, “... Yes we have to cover out these 5 questions, but we can ask other things about other value to the program.”

The challenge of achieving 100% evaluation coverage was also cited as a factor in that it limited the ability of an evaluator to think of new innovations, “Your time to do the thinking and analysis is limited because we are constantly going from one evaluation to the next.”

5.6.7 Policies on data display (Web Standards for the Government of Canada)

Limits were also mentioned by a few evaluators when speaking to the question of data display, “We tend to underutilize data display, there are federal and provincial standards on how the information is delivered, for example, because of the website there is a tight limit on how data can be displayed, there is a big opportunity for improvement here but it is not necessarily done.” As mentioned in the data display and presentation of evaluation results sub-section. For one division that relied heavily on data displays—writing these descriptions and having them translated was more costly than the division could sustain and therefore limited the capacity of evaluators to use certain data displays when publishing results. Some evaluators interviewed admitted that knowing that they could not use a particular data display in the final evaluation report discouraged them from using certain data display tools in the evaluation.

5.6.8 Contracting Policy

Contracting policy presents some challenges as well. From an external consultant perspective, “it is hard to be innovative from outside of government because we are reactive to Requests for Proposals.” One consultant explained how trying to be innovative in their evaluation proposals adversely affected their business, “It is risky to suggest innovation in a response to an RFP. We

have had experiences where we have attempted to be innovative in our bids and we had our scores reduced because we were not meeting the response.” As was mentioned in the sub-section on unimplemented evaluations there is an administrative burden to secure contractors when the expertise for an innovation is not located in house. This can be inconvenient for evaluators who rely on these contracts and “have to prepare them 6 months in advance.” If a contractor is being brought in for an innovative service than the innovation would have to be identified at a very early stage in the evaluation for it to be possible to secure that expertise in line with the overall evaluation plan.

5.6.9 Limited Knowledge and Exposure to Innovations

Some of the interviewees had pointed out that they are limited in using innovations by their lack of knowledge and exposure to other innovations occurring in evaluations as well as the success or failures of these innovations. As one evaluator put it, “we don’t know all of our options” and another interviewee explained that “In terms of leadership we have no limitations...but knowledge is one of the barriers we face; we don’t know what else can be done.”

5.6.10 A Director or Director General who is an auditor

One evaluator explained how it can be hard to get the Director General to understand and approve certain evaluation methodologies—particularly new or innovative ones—if he or she does not come from an evaluation background. Another evaluator also expressed similar challenges at the director level.

5.7 Requirements for Better Innovation Implementation

This sub-section groups all of the requirements identified by the interviewees for better implementation of innovations into three thematic groups.

5.7.1 Internal Capacity:

All of the requirements for better incorporation of innovations grouped into the theme of internal capacity are listed below:

- More resources (time and budget)
- More staff
- Staff who have mastered the basics of evaluation
- Increased capabilities of evaluation staff
- Easier contracting-out processes when there is no internal capacity for the innovation
- Increased awareness of innovations.

More time and budget would be required to incorporate innovations. Some evaluators spoke to this in reference to the amount of time it takes to implement an evaluation as well as any additional costs that might be required such as purchasing a new software license. Several evaluators had indicated that their amount of evaluation staff is either too small or reduced due to budget cuts. The result being that contracting out is required in order to meet 100% evaluation coverage of their departments direct spending. The administrative burden was identified as being uncondusive to quickly securing contractors who are able to execute the innovative techniques.

5.7.2 Culture and Support

All of the requirements for better incorporation of innovations grouped into the theme of culture and support are listed below:

- More supportive environment and less resistance from other staff and senior staff
- Senior management support
- More support and guidance from the TBS CEE
- Evaluation managers willing to take risks
- Evaluation managers with strong personalities that can push innovative ideas through
- More openness to risk broadly and acceptance of failure
- Directors or Director Generals with an evaluation background as opposed to audit.

The interviewees had expressed that there are challenges with other staff, with their direct management, senior management in regards to change and trying new ideas. Some interviewees indicated that it was not outright resistance but sometimes a lack of support to help incorporate these ideas. Some evaluators that looked to the CEE expressed that they would like to see more information coming from TBS. One interviewee pointed out that, “TBS doesn’t really have much to offer in terms of knowledge” and another evaluator stated that, “They produce a limited number of guides and we are left to ourselves.”

Some of the interviewees who were critical of the risk-averse culture believed that more acceptance of risk would be required for innovations to be incorporated in evaluation. As one evaluator explained, “Staff should not be penalized for taking chances and failing, attempts at being innovative don’t have to always work out and that’s okay.”

In regards to having former evaluators for directors as opposed to auditors, one interviewee explained their reason for this suggestion, “The auditor skillset is more aligned with checking the box than thinking about innovations.” This suggestion however does not eliminate the possibility of an audit based director who is willing to consider innovative methods in evaluations. As elaborated by the interviewee the point being made is about finding the right person for the job, “I don’t have a solution for the staffing of directors...I wish they could have tools to get the right person for the job so you could get someone who thinks like an evaluator; and thinks like an innovative evaluator.”

5.7.3 More Freedom and Flexibility

Having more freedom and flexibility was an identified requirement for innovation:

- More flexibility in reporting formats
- Adopting a risk based approach to evaluations
- More freedom regarding evaluation questions.

One evaluator explained that, “to do coverage of the 5 core questions appropriately is quite extensive and hard to add to that. If we wanted to do more questions we would have to do a new evaluation.” In general having the freedom and discretion to focus on the most applicable questions would free up time to allow for a deeper analysis of relevant sections of the evaluation.

5.7.4 Improved Collaboration

All of the requirements for better incorporation of innovations grouped into improved collaboration are listed below:

- Better communication, discussion and relationships between consultants and government
- Evaluation needs to be better involved in decision making and planning committees
- Better communicating with stakeholders as to what the innovation is, its impact on the evaluation and to allay stakeholder concerns
- Increased program understanding of the role of evaluators
- Better collaboration across the department
- Sharing templates and examples of innovations.

Improved communications between consultants and government was emphasized particularly by the consultants interviewed. As explained in the barriers to innovation sub-section several contractors pointed out that the RFP process limits them from including innovative techniques into their submissions for bids. One evaluator explained that to get innovations in the federal government they need to be having conversations with government and “getting the idea and information out to the buyers of evaluation.” One interviewee explained that government employees are sometimes unnecessarily worried about speaking with consultants because they have been afraid that they would be charged for informal chats or short meetings. The interviewee also identified difficulties in communication on both sides of government and contractors, “It’s a two way street, we can’t expect business out of a short meeting, the same way they can’t expect to be charged for informal meetings.”

Communication within the department was also discussed frequently in the interviews. Interviewees explained that there could be benefits to getting more exposure on the planning side of government and to work closer with the programs. Interviewees recounted the misperceptions that programs have of evaluation and pointed out that programs could benefit from a changed perspective, “Evaluation should not be seen by programs as a requirement it should be seen as a consultation, like asking your lawyer.” This is to say, that evaluators need to collaborate better with programs so that program evaluation is something that programs want to have them conduct instead of just having them as a requirement. Better collaboration with other internal services where applicable was also identified especially with the internal corporate services such as any centralized research, financial and audit divisions or other groups involved in research and data collection within the department. This was identified as an innovative way to improve the quantity and or quality of data that feeds into an evaluation.

Particularly what was identified as being useful would be the exchange of ideas with other evaluators as well as successes and failures. One interviewee gave an example of an idea for an innovation that they were considering and found out that a different division was currently implementing the same idea. The idea seemed promising; however, the interviewee had learned later that the organization implementing it had unsuccessful results that revealed the reasons why the innovation was likely to fail. The interviewee pointed out that knowledge of other failed innovations could help end those discussions that are not likely to bear fruit. In regards to learning of other evaluation successes—interviewees stated that knowing that an innovative tool or methodology could be successful in other government contexts would be helpful in getting support to use these methodologies. As well, having templates from other divisions to model

afterward would further contribute to getting the innovation approved and eventually help a division to successfully model these approaches.

5.8 The Ongoing Need for Innovation

Generally the evaluators interviewed strongly agreed that there is a need for innovation (when it was necessary). Some evaluators referred to this need as the ongoing need for change or improvement in evaluation. There were various drivers of need for innovation that were mentioned.

5.8.1 Innovation to Address Resource Constraints

First of all, “Innovation is needed to deal with resource constraints” and “Necessity is the mother of all innovations given that resource parameters are tightened.”

5.8.2 Innovation to be More Useful for Clients and Improve Evaluation Quality

As well innovation is need driven by the client, in order, “To do better justice to their work and the programs to reflect their performance.” One, evaluator implied that something novel is required in order to make the results more useful to stakeholders, “The value of evaluation is being questioned because we are not producing the results that are useful to the Government of Canada.”

5.8.3 Innovation to Stay Relevant with Change and to Overcome Challenges

One evaluator pointed out that sometimes their evaluation reports had not been read by the Departmental Evaluation Committee (DEC) when they presented the results at the DEC meeting. The interviewee believed this was a good indication that something innovative must be done to change this.

As put by one interviewee, “There is always a need for innovation in every field. Interviews, case studies and focus groups were all innovative at one time...sticking on the same path is not going to provide *new* information and show things from a *new* perspective. The basic requirement of evaluators is that we need to look at things in a new way, that’s our job.” Innovation was also identified as needed when current methodologies are insufficient and this change is a driver for the need for innovation. One evaluator explains, “We had been collecting information about people by doing surveys, but survey rates have plummeted and we are not getting the same response rates as before. So we need to find something else to do.” Several evaluators pointed out that innovations do not have to be large or revolutionary, but can be small such as improving internal processes; continuous improvement.

There were objections to the need for innovation in evaluation. Those who shared this view believed that there is room for improvement in evaluation and that this should be the focus, rather than ‘innovation.’ A few evaluators pointed out that there is the foremost need for evaluators to master the evaluation basics before experimenting with innovation, “We aren’t just missing innovation, we are missing concrete basics.” Others admitted that there is a need for innovation but emphasized that this need is only there when the innovation is necessary; when current methodology could not provide the right answers, “If there is no need to try to change the evaluation by trying innovative approaches, then tried and true methods should be used.”

5.9 Additional Thoughts and Suggestions about Innovation in Evaluation

5.9.1 How Evaluators Defined Innovation

Different beliefs and views of innovation were held by the interviewees. This suggests there is a diversity of interpretation as to what is considered innovative by members of the evaluation community. Some of the comments below captured common but sometimes competing views found within the interviews:

- “To be innovative I would say it would have to be a new social science research method”
- “You would need to stick to the literature to prove where its tested”
- “TBE is not necessarily innovative if a few people are using it”
- “Innovation has to have value added”
- “When evaluators are looking at how to be innovative, they should not just look at innovation, but at other sectors for inspiration”
- “There are no gestalt changes in evaluation, I do not think there is innovation, we adapt and we incrementally change; continual improvement is our focus.”

5.9.2 Innovation for the Sake of Innovation

Some interviewees believed that innovation in evaluation was possible and occurring whereas some outright disagreed, “If you are inside government you are innovating simply because you have no money to work with.” This was in reference to evaluation divisions being innovative in regards to becoming an efficient evaluation shop rather than being innovative to be efficient in the field of evaluation, “Much of the innovation that exists are just the attempts to manage in a resource constrained evaluation society and environment. It is hard to know what is innovative and what is cutting corners... I suspect we may have become more efficient and timelier but we are probably more superficial.” One interviewee pointed out that, “Like in any area there are also fads, ‘innovations’ come and go.” Several evaluators spoke out against superficial innovation saying that, “We won’t be innovative for the sake of innovation.”

5.9.3 Innovation as Ongoing Continuous Improvement

In opposition to innovation for the sake of innovation several evaluators defined innovation under the context of continuous improvement which was captured well by an interviewee, “There is an organizational context in regards to innovation, yes, we are always looking at continuous improvement; we want to make sure it’s not about gimmicky innovation.” Several evaluators spoke to this point of evaluation having the need to continuously improve.

6. DISCUSSION

So far this project has demonstrated through the literature review that there is an established body of basic methods for conducting evaluations, however, that in some circumstances there are opportunities for innovation. The literature review and findings from interviews has demonstrated that there are methods being developed and used that offer program evaluators options that can help to better address stakeholder needs for some evaluations. These options are not without drawbacks and there are barriers and limitations that were discussed that would need to be taken into consideration when attempting to utilize these findings.

In regards to these limitations the interview findings corroborated the literature review in that potential challenges and barriers to innovation were:

- Lack of resources and time
- Difficulties in the contracting process
- The risk adverse culture within the federal government.
- Over prescriptive policy framework (As cited by Roy and Dumaine at a CES conference in 2013).

One concept emphasized in the interviews was the distinguishing of innovation in evaluation as a valued improvement or need versus innovation as a gimmick or for the sake of innovation. It was suggested that evaluators do not need to be innovative necessarily; but they do need to continually improve. This is an important interpretation as the value added of the innovation should be that it improves a certain function and not that it just new or different. One of the elusive features of what is considered innovative will depend on the perspective of the beholder. For a division that is trying something new that another division has already done—the division trying this tool or method for the first time may consider it innovative whereas the division that has already been using this practice may have already developed and integrated this method or tool into standard practice and therefore no longer considers it an innovation. The term innovation may be interpreted to refer to the whole field of evaluation and may not be as inclusive as the term continuous improvement when speaking broadly about program evaluation in general. From discussions with evaluators working the federal government context it can be gleaned that what is most important for an evaluation division is that it is contributing to produce better evaluations for itself. Whether these can be seen as contributions to the field of program evaluation in the form of innovative techniques is secondary. Innovation should respond to some sort of need or worthwhile value added as it would be superfluous to innovate for its own sake especially when evaluation divisions are operating in a resource constrained environment.³⁵

Some of the interviewees had pointed out that the basic and required evaluation skills were lacking within their division and suggested that this should be improved upon first before trying some new innovate technique. It is imperative that the most basic evaluation skills are grasped and exemplified before innovations are incorporated into their practices. Knowledge of what innovations are possible and understanding the contexts in which they can be successful will play an important role for an evaluation division that has mastered basic knowledge and skills and aims to improve the usefulness of its evaluation functions. Therefore, an evaluation division

³⁵“While the demand for evaluation and evaluators appears on the rise, the supply of evaluators and the financial resources to conduct evaluations are not” (Gauthier et al., 2012 p. 2).

must maintain and enhance its skill sets as well as keep in touch with the evaluation community at large to stay relevant and continuously improve. When innovative ideas spark interest evaluators should also consider the factors necessary to make that innovation successful and not to become too invested in the innovation too soon. Innovation occurs within a context and it is prudent to not get disillusioned and get caught up in the hype cycle (Azzamet al., 2013, p.17). An illustration of the hype cycle can be found in Appendix 12.

The Federal Government's program evaluation community is a small one, consisting of individuals mainly working in the National Capital Region (NCR). A wealth of evaluation knowledge and experience exists within this community; however, there are just few established processes to share information. There are annual CES events held across the country such as Lunch and Learns which are held regularly. Less frequently regional CES chapter conferences are held as well. Within the government evaluation function are bi-annually held heads of evaluation meetings, however only directors are present and not the evaluators themselves, so detailed methodologies are rarely included in the higher level discussions that occur. Occasionally, the Centre for Evaluation Excellence (CEE) within TBS will hold targeted events on a case by case basis but it is outside the capacity of the CEE to provide all of the answers for evaluators. Outside of these are the more informal methods of evaluator social networking through EvalConnex and Evaltalks. The first step in the innovation diffusion model as mentioned in the literature review section begins with the knowledge of an innovation;³⁶ therefore, it may be prudent to a would-be innovative evaluation division to keep in touch with these networks for the purpose of gaining knowledge regards to innovative processes. Additionally there are courses that can expose evaluators to new ideas as well as to train them in the use of these tools or methods.

All of the innovations discussed in this paper can be applied successfully only when the circumstances are appropriate and should only be used when there is a need that is being addressed. Evaluators should consider the clients and users of evaluation in mind when considering innovations and should help to align these innovations towards their particular needs. That being said, one evaluator pointed out that the client or users may not necessarily know what their own needs are by referring to a famous quote attributed to Henry Ford, "If I had asked people what they wanted, they would have said faster horses." It may be up to evaluators to understand the capabilities of certain innovations or techniques and recognize when they can be employed to take advantage of the possible insights that they might provide. This also brings into the discussion another important point which is that the innovation may not work. Directors of evaluation will have to consider the limitations, costs and effort required and if the risk of failure of the technique or tool being proposed against the potential benefits.

An interesting irony was found in the interviews. Federal government interviewees stated that government will typically look to the private sector to identify innovations. Whereas private sector consultants pointed out that they are constrained in the evaluation plan by the Request for

³⁶ "In Rogers' diffusion of innovations model, the innovation diffusion process passes from (a) first *knowledge* of an innovation, to (b) forming an *attitude* toward the innovation, to (c) taking a *decision* to adopt or reject, then to (d) *implementation* of the new idea, and finally to (e) *confirmation* of this decision As an organized action, innovation is aimed at producing change"(E. M. Rogers, 1995, 2010) and (Desbiens & Dagenais, 2002) as cited in (Rey et al., 2012, p.71).

Proposals (RFPs) prepared for by the federal government department. It is possible that these two points could work together to prevent innovation. If this is the case than the implication is that federal government evaluators cannot reasonably rely on the private sector to identify innovative opportunities for the division given the constraints of the contracting policy. Instead evaluation managers who are responsible for creating the RFPs may be best positioned to identify innovative opportunities within their division the implication being that these managers will need to be kept informed of potential innovative methods occurring in order to seize these opportunities. If contracting methods are not suitable for identifying innovative methods than other forms of communication should be occurring outside of these to identify innovative tools and methodologies.

After mastering the basic skills and getting informed of innovative practices evaluators seeking to improve will have to play close attention to potential opportunities for innovation. Identifying situations where an innovation can be successful will require an experienced evaluator who is familiar with the context in which they are operating and is willing to take on the challenges of exploring innovative options. As put by Roy and Dumaine, in order to have new methods introduced, “We need to have inconvenient evaluators who are willing to look at the advantage of new ways” (as cited from Roy and Dumaine at a 2013 CES Conference). Although there are many examples of innovation in evaluation found within the literature and from the interviews, the unique circumstances that can make an innovation successful will occur within that evaluation or evaluation division and it is up to individual evaluators to identify these opportunities. Likewise innovation requires an agent of change; someone devoted to this idea and willing to see it through, “Therefore, an innovation’s success relies on the presence of liaison-agents (Orlandi, Landers, Weston, & Haley, 1990), or agents of change in the organization (E. M. Rogers, 1995) who devote their personal influence to encouraging the innovation’s adoption” (as cited in Rey et al., 2012, p. 72). As well, its success requires a culture that is ripe for the innovation, “Moreover, the process of innovation diffusion hinges upon organizational, sociocultural, and political factors” (as cited in Rey et al., 2012, p. 72).

Given the rigid requirements of TBS, it may be a challenge to incorporate ‘innovative’ techniques found in the literature and still maintain acceptable MAF ratings if evaluators stray too far from the standards that are elucidated by TBS Policy and Directives. Whether or not the TBS 2009 policy is open to interpretation may be a moot question since—as demonstrated by the interview findings—evaluators certainly have been interpreting it in different ways. If in doubt, evaluators can consult with the CEE and ask for advice in regards to attempting to engage in an innovative activity. Having these conversations is important for both the evaluator and the CEE and can serve as a learning opportunity for all of those involved. These types of consultations were reported in the interviews to have helped to ensure the division was properly addressing the evaluation policy which leads to acceptable MAF ratings for evaluators attempting innovations. Logically the CEE will be better suited to rank MAF ratings as acceptable if they have an understanding of why the innovation was used and may allow for flexibility if kept informed.

Other collaborations have been occurring within federal government departments between evaluation and other counterparts in internal corporate services. Numerous benefits of these relationships had been cited in the interview findings and it may be worth the investment for evaluators to cultivate these relationships to seek efficiencies as drawing on theses as sources is

an innovative way to add more data into an evaluation and can reduce to burden of data reporting on program.

Developing a working relationship with programs in itself can serve a function as well. The literature suggests that there is an important connection between evaluation utilization and evaluation capacity building within an organizational context, such that the more an organization uses their evaluation results, the more likely ECB efforts will continue over time” (Dreolin et al., 2008, p. 41). As some of the interviewees had mentioned that programs unfamiliarity with program evaluation could be a barrier particularly when introducing innovative approaches, developing a report with programs through exercises such as evaluation capacity building could help to facilitate the process of implementing innovative approaches when they are identified.

7. LESSONS LEARNED

This section will put forth four lessons learned regarding the facilitation of innovation in evaluation with the directors of evaluation divisions as the target audience. After briefly discussing the context in which innovation occurs, this section will review some of the innovative tools examined as part of this research and revisit the need for innovation. Next, director support and professional development will be described as a way to support innovation in evaluation. This section will close with some thoughts related to ensuring that innovations are compliant with TBS policy and MAF assessments.

7.1 The Context of Innovation in Evaluation

Federal government program evaluation is required by the TBS policy to address the 5 core questions.³⁷ Some interviewees had indicated that there is little flexibility in this requirement and going beyond this can be arduous for the evaluators and evaluands involved. Asking additional questions or using additional methods other than traditional approaches may also require additional resources. This presents a challenge as interviewees indicated that evaluation divisions have constrained capacity and limited resources. Some of the innovations discussed in this paper require skills or tools that may not be available in a typical evaluation division and would necessitate additional support. This context presents a challenge to being innovative.

7.2 Innovative Tools and the Need for Innovation

This subsection summarizes information learned about some of the innovative tools examined and the need for innovation.

7.2.1 Geographic Information Systems (GIS)

GIS is a tool that has been used successfully within the federal government context and has been reported to add value when it has been used. This tool would not be applicable for all program evaluations; however, there are opportunities for its use such as in proximity analysis or creating visual representations of diverse geographic outcomes (health or economic outcomes by county or province) which may be valuable for informing decision-making. Additional drawbacks of this method are the associated costs for the software, licensing, training or the use of a consultant. Despite these drawbacks, this is a tool that may be underutilized in federal government program evaluations and may be worth exploring in innovations where this tool has the potential to add value.

³⁷ Relevance:

1. Continued Need for the Program
 2. Alignment with Government Priorities
 3. Alignment with Federal Roles and Responsibilities
- Performance – Effectiveness Efficiency and Economy:

4. Achievement of Expected Outcomes
5. Demonstration of Efficiency and Economy (Policy on Evaluation, 2012)

7.2.2 Data display and Presentations of Evaluation Results

When it comes to data display, the literature indicates that simple and clear communication is the best approach (Lysy, 2013, p. 34). Although there were a wide variety of data display options identified in the literature, from the account given by the interviewees most displays beyond simple charts and graphs are typically not used in evaluations. This may be appropriate as the literature points out that “Bar charts, line graphs, and scatterplots are as effective today as they were 15 years ago” (Lysy, 2013, p. 34). The lack of alternative data display and presentation choices was partially explained by interviewees as being related to the limitations on the use of data display.³⁸ In light of these revelations, this paper suggests that there is limited value to pursuing innovations in evaluation that involve data display. One type of presentation and methodology that had been used by several interviewees was Photovoice. Although it has been used successfully Photovoice may require additional personnel and costs and it has a recognized limited applicability to use in most federal government evaluations particularly when addressing the 5 core questions.

7.2.3 Social Networking and Social Media

Evidence of social networking and social media contributing to a specific evaluation both in literature and in the interviews was quite sparse. Furthermore its use as a line of evidence to feed into federal government decision making is preliminary and potentially controversial. The benefit of its use in evaluation may be limited in comparison to traditional lines of evidence. However, Social networking as a method of sharing information within the evaluation has been identified for its potential as a professional development and information sharing tool. As technological developments have allowed for this type of exchange there could be some value for evaluators to access networks such as EVALtalk, GCConnex, and EvalConnex to learn more broadly about the field of program evaluation.

7.2.4 Big Data

Although Big data was heralded in the literature it was seldom used by the evaluators interviewed. Big data may be inapplicable to most evaluations in the federal government context as interviewees indicated that they do not typically have large datasets within the department or related to programs. One evaluator pointed out that there are many changes external to a program that occur, the result being that this type of analysis is exposed to several threats to validity. Big data may also be limited in its ability to narrow analysis down to specific evaluands—which is what evaluation typically tries to do. The skills required to conduct Big data analysis were not something identified as readily available within evaluation divisions and more than likely would require additional support. For these reasons this innovative tool is likely not worth looking into for standard evaluations.

7.2.5 Need for innovation

The information gleaned from this research does not support a strong need for innovation in evaluation generally; rather, the need to continuously improve practice was identified as important by several of the interviewees. Innovation for its own sake—that is—to do something

³⁸ The examples given were that evaluators are not able to use complex displays in Departmental Evaluation Committee meetings and the time and costs to produce long text descriptions and translation that are required with online web display policy.

new because it is available would be a misplaced investment. An innovation should only be invoked when there is clear value that can be added and it is likely to succeed in providing this value. When attempting something new there is uncertainty and of its success and this should be taken into consideration as well as the potential benefit it may offer when deciding if an innovation is worth attempting.

7.3 Director Support and Professional development

Director level support and leadership can be a welcomed aspect in recognizing the effort that is required by an evaluation manager to incorporate an innovation into an evaluation. Innovation occurs in a context and a culture and it is important that this culture is supportive of innovations if they are expected to get off the ground. Directors of evaluation can be supportive of innovation by encouraging a culture of learning and professional development.

Managers who are familiar with the details of an evaluation and have the experience to judge if there is potential for value to be added may be in the best position to make recommendations for innovative tools. It was identified by several interviewees that having knowledge of innovations as well as how they have successfully or unsuccessfully been applied would help to better incorporate these innovations in evaluations. Evaluators can get exposure to innovations through professional development such as participation in evaluation community events, structured training, and conversations with other evaluators in the public and private sectors. Through these networks evaluation managers can share ideas, tools and templates that can help to make innovations successful. By facilitating access to professional development, directors can help ensure that evaluation staff will have the knowledge and skills required for implementing innovations that could add value to an evaluation.

7.4 Communication with TBS CEE regarding MAF

The Management Accountably Framework (MAF) ratings of an evaluation division will reflect on the director so it is critical that the director understands the requirements related to innovation and the criteria within the assessments. Likewise, the director must ensure that the assessors of MAF ratings understand why a particular innovation was used by demonstrating its value added. It may also be incumbent upon directors to be able to demonstrate why an innovation was not used by explaining the limited ability it may have had to add value in an evaluation.

To help ensure that the MAF assessment is positive, directors should consult with the CEE when considering whether to implement an innovation and clarify its intended use. Engaging the CEE is important for both the evaluation division and the CEE and can serve as a learning opportunity for all of those involved. These types of consultations were reported in the interviews to have helped to ensure the division was properly addressing the evaluation policy when innovating. Meeting these policy requirements leads to acceptable MAF ratings so having these conversations could contribute to getting an acceptable MAF rating. Logically the CEE will be better informed when conducting MAF assessments if they have a greater understanding of why the innovation was used and may allow for flexibility if consulted. Likewise directors can be more confident going into an assessment knowing that they have vetted the innovation with these assessors during the course of its implementation.

8. CONCLUSION

The purpose of this project was to summarize current knowledge related to innovation in program evaluation in order to inform future thinking so that federal government evaluators in Natural Resources Canada can continually improve their practice through exposure to these ideas. Through investigating the literature and discussing with evaluators working with the federal government evaluation context this project identified benefits of innovations in evaluations and discussed some of the barriers and limitations to provide a better understanding of innovation in evaluation.

This was done by giving a background, methodology and conceptual framework that provided the historical genesis surround this project as well as described how it was conducted. Relevant texts found in the literature were introduced to discuss the methodologies and tools related to innovation in evaluation and the findings from the interviews were presented. The discussion section highlighted some of the key aspects identified in the literature and interview findings which were used to help inform the four lessons learned regarding the facilitation of innovations in federal government program evaluation.

9. REFERENCES CITED

- Assessing Program Resource Utilization When Evaluating Federal Programs. (2013). *Treasury Board Secretariat of Canada*. Retrieved June 19, 2014, from <http://www.tbs-sct.gc.ca/cee/pubs/ci5-qf5/ci5-qf5tb-eng.asp>
- Azzam, T. & Robinson, D. (2012). GIS in Evaluation: Utilizing the Power of Geographic Information Systems to Represent Evaluation Data. *American Journal of Evaluation*, 34 (2),207-225
- Azzam, T., Evergreen, S., Germuth, A. & Kistler, S. J. (2013). Data visualization and evaluation. In T. Azzam & S. Evergreen(Eds.), *Data visualization, part 1. New Directions for Evaluation*, 139, 7–32.
- Bakker, Paul. (n.d), Canadian Evaluation Reflections Part 2:Photovoice, *Social Impact Squared*. Retrieved June 24, 2014, from <http://www.socialimpactsquared.com/evaluation/canadian-evaluation-conference-reflections-part-2-photovoice/>
- Bamberger, M., & Rugh, J. & Mabry, L. (2006). Real World Evaluation: Working Under Budget, Time, Data, and Political Constraints. p. 1-468. *Sage Publications*.
- Bernier, C. (May 08, 2014). Is the government checking you out on Facebook? *HuffingtonPost*. Retrieved June 19, 2014, from http://www.huffingtonpost.ca/2014/05/08/canada-snooping-social-media-privacy-commissioner_n_5289753.html
- Brown, Ryan. (2011). Evaluation Use at Environment Canada, p. 1-72, Retrieved June 19, 2014 from http://dspace.library.uvic.ca:8080/bitstream/handle/1828/3683/brown_ryan_mpa_2011.pdf?sequence=3
- Canadian Evaluation Society 2011 Conference Program. (2011). *Canadian Evaluation Society*. Retrieved June 19, 2014, from <http://c2011.evaluationcanada.ca/documents/>

[2011_CESFinalProgram.pdf](#)

Capital Assessment Survey. (2012). *Treasury Board Secretariat of Canada*. p. 1-14. Government of Canada Internal Document.

Coryn, C., Noakes, L., Westine C. & Schröter, D. (2011). A Systematic Review of Theory-Driven Evaluation Practice from 1990 to 2009. *American Journal of Evaluation*, 32 (2), 199-226.

Dreolin, N., Christie, C. & LaVelle, K. (2008). Perceptions of Evaluation Capacity Building the United States: A Descriptive Study of American Evaluation Association Members. *Canadian Journal of Program Evaluation*, 23 (3), 37-60.

Dybdal, L., Nielson, S., Lemire, S. & Ramboll Management Consulting. (2011). Contribution Analysis Applied: Reflections on Scope and Methodology. *Canadian Journal of Program Evaluation*, 25 (2), 29-57.

Gauthier, B. et al., (2010). The Lay of the Land: Evaluation Practice in Canada in 2009. *Canadian Journal of Program Evaluation*, 24 (1), 1-49.

Henderson, S. & Segal, E. H. (2013). Visualizing qualitative data in evaluation research. In T. Azzam & S. Evergreen (Eds.), *Data visualization, part 1. New Directions for Evaluation*, 139, 53–71.

Janus, A. (May 08, 2014). Is the government checking you out on Facebook? *CTVNEWS*. Retrieved June 19, 2014 from <http://www.ctvnews.ca/mobile/politics/is-the-government-checking-you-out-on-facebook-1.1812110>

Johnson, J., Hall, J., Greene, J. & Ahn, J. (2013). Exploring Alternative Approaches for Presenting Evaluation Results. *American Journal of Evaluation*, 34 (4), 486-503.

Kistler, S. J., Evergreen, S. & Azzam, T. (2013). Toolography. In T. Azzam & S.

- Evergreen(Eds.), *Data visualization, part 1. New Directions for Evaluation, 139*, 73–84.
- Kwong, M. (May 09, 2014). Government snooping on social media may breach Privacy Act. *CBCNEWS*. Retrieved June 19, 2014 from <http://www.cbc.ca/m/news/#!/content/1.2636905>
- Lysy, C. (2013). Developments in quantitative data display and their implications for evaluation. In T. Azzam & S. Evergreen (Eds.), *Data visualization, part 1. New Directions for Evaluation, 139*,33–51.
- Mason, G., Tereraho, M. (2007). Value-For-Money Analysis of Active Labour Market Programs. *Canadian Journal of Program Evaluation, 22 (1)*,1-29.
- McDavid, J., Hawthorn L. (2006). Program Evaluation and Performance Measurement: An Introduction to Practice. p. 1-477.Thousand Oaks, CA: *Sage Publications*.
- McDavid, J., Huse, I. & Hawthorn L. (2013). Program Evaluation and Performance Measurement: An Introduction to Practice, Second Edition. p. 1-534. *Sage Publications*.
- McDavid, J. (2012). In Search of a Balanced Canadian Federal Evaluation Function: Getting Back to the Future. *Canadian Journal of Program Evaluation, 26 (2)*,55-59.
- Patton, M.Q. (1997). Utilization-focused Evaluation. *Sage Publications*.
- Petrosino, A. (2000). Answering the Why Question in Evaluation: The Causal-Model Approach. *Canadian Journal of Program Evaluation, 25 (1)*, 1-24.
- Policy on Evaluation. (2012). *Treasury Board Secretariat of Canada*. Retrieved June 19, 2014, from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?section=text&id=15024>
- Policy on Management, Resources and Results Structures. (2012). *Treasury Board Secretariat of Canada*. Retrieved June 19, 2014, from <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18218§ion=text>

- Results-Based Management Lexicon. (2012). *Treasury Board Secretariat of Canada*. Retrieved June 19, 2014 from <http://www.tbs-sct.gc.ca/cee/pubs/lex-eng.asp>
- Rey, L., Brouselle, A., Dedobbeleer, N. (2012). Logic Analysis: Testing Program Theory to Better Evaluate Complex Interventions. *Canadian Journal of Program Evaluation*, 26 (3), 61-89.
- Smith, N et al. (2011). Looking Ahead: The Future of Evaluation. *American Journal of Evaluation*, 32 (4), 565-599.
- Snijders, C., Matzat, U &, Reips., U (2012). Big Data: Big Gaps of Knowledge in the Field of Internet Science. *International Journal of Internet Science*, 34 (1), 1-15.
- Sundar, P., Kasprzak S., Halsall, T. & Woltman, H. (2011). Using Web-Based Technologies to Increase Evaluation Capacity in Organizations Providing Youth and Mental Health Services. *Canadian Journal of Program Evaluation*, 25 (1), 91-112.
- Supporting Effective Evaluations: A Guide to Developing Performance Measurement Strategies.(2010). *Treasury Board Secretariat of Canada*. Retrieved June 19, 2014 from <http://www.tbs-sct.gc.ca/cee/dpms-esmr/dpms-esmr05-eng.asp#OverviewLM>
- Szanyi, M., Azzam, T. & Galen, M. (2013). Research on Evaluation: A Needs Assessment. *Canadian Journal of Program Evaluation*, 27 (1), 39-64.
- Theory-Based Approaches to Evaluation: Concepts and Practices. (2012). *Treasury Board Secretariat of Canada*. Retrieved June 19, 2014, from <http://www.tbs-sct.gc.ca/cee/tbae-aeat/tbae-aeat01-eng.asp>
- Vanlandingham, G. (2011). Escaping the Dusty Shelf: Legislative Evaluation Offices' Efforts to Promote Utilization. *American Journal of Evaluation*, 32 (1),85-97.

10. APPENDICES

10.1 Appendix 1 – Validity

Below is a list of validity threats accompanied by brief notes:

- A. Statistical conclusion validity
 - 1. sampling procedures are adequate
 - 2. appropriate statistical tests have been applied
 - 3. outcomes are interpreted correctly
 - 4. measures being used are sufficiently reliable
- B. Internal validity – assessing causal linkages in program evaluations
 - 1. History
 - 2. Maturation
 - 3. Testing
 - 4. Instrumentation
 - 5. Statistical regression
 - 6. Selection
 - 7. Mortality
 - 8. Ambiguous temporal sequence in the cause and effect variable
 - 9. Selection-based interactions
- C. Construct validity – being able to generalize from the variables in the evaluation and their observed relationships back to the constructs and their relationships in the program logic
 - 1. Diffusion of treatments – sometimes people can communicate experiences to control group
 - 2. Compensatory equalization of treatments – the group that is not supposed to get the program is offered it to avoid criticism of withholding potential benefits
 - 3. Compensatory rivalry – the no program group is given a ‘better’ version of the existing program and cancels out differences
 - 4. Resentful demoralization – the control group perceives unfair treatment and reacts negatively
 - 5. **Hawthorne effect** – being part of an experiment produced an effect regardless of the levels of the experimental variables being manipulated (caused by the fact that people know they are part of an experiment)
- D. External validity (evaluation results being able to be reproduced elsewhere)
 - 1. Causal results of a program and the people/participants (some people in the experiment produced more changes in boys than girls)
 - 2. Causal results of a program and the treatment variations (programs are done differently in different places)
 - 3. Causal results of a program and outcome variations (if extended outcomes were desired...such as increasing employment a year, as opposed to 6 months, or increasing wages by 10,000 not 6,000.
 - 4. Causal results of a program and the setting – a school to influence children in poor families (but the setting takes place in a university town- where most families enjoy good incomes)
 - 5. Context-dependent (McDavid, 2006).

10.2 Appendix 2 – Overview of the Three Generations of Contribution Analysis

Overview of the Three Generations of Contribution Analysis

	<i>First-generation CA</i>	<i>Second-generation CA</i>	<i>Third-generation CA</i>
Article	Mayne (1999)	Mayne (2001)	Mayne (2011)
Scope	How monitoring data can be used to address the attribution question through CA (p. 5)	How monitoring data can be used to address the attribution question through CA (p. 5)	How CA can be used in evaluation and/or performance measurement (p. 1)
Problems addressed	Sketches elements in CA	Elaborates operational CA design	Elaborates operational CA design and considers implications in terms of causal claims
Operational approach	CA structured around nine elements	CA's nine elements consolidated into six steps	CA's six steps changed slightly. New Step 1 and Steps 3 & 4 consolidated as new Step 4. Step 7 added
Level of detail for each step	Outline guidelines	General guidelines	Elaborate guidelines
On attribution	Acknowledges experimental design most appropriate for addressing attribution (p. 5)	Acknowledges experimental design most appropriate for addressing attribution (pp. 4–5) Argues that "plausible association" can be inferred (p. 8) Contribution can be inferred when four conditions are met (pp. 21–22)	Argues that other approaches to attribution can be valid (pp. 5–6) Argues that probabilistic causation can be inferred (p. 7) Contribution can be inferred when four conditions are met (p. 7)
On magnitude of contribution	None offered	None offered	Based on evidence available and causal links demonstrated three levels of contribution can be determined (pp. 25–26).
Methodology	Advocates mix of quantitative and qualitative methods. Use of existing data and additional data collected (p. 5)	Same (p. 18)	Same (pp. 17–21)
Epistemology	None offered	None offered	None offered
Ontology	None offered	None offered	None offered

(Dybdal et al., 2011, p.33).

Figure 4. Contribution Analysis Process

Step 1: Set out the cause-effect issue to be addressed

Step 2: Develop the theory of change

Step 3: Assess the resulting contribution story

Step 4: Gather the existing evidence on the theory of change

Step 5: Reassess the contribution story and challenges to it

Step 6: Seek out additional empirical evidence

Step 7: Revise and strengthen the contribution story

(Theory-Based Approaches to Evaluation: Concepts and Practices, 2012).

10.4 Appendix 4 – Interview Instruments

Advanced Notification Email

Dear colleague,

A study on **Innovation in Program Evaluation** is currently underway by a Masters Student from the University Of Victoria's School of Public Administration. This study is being done in collaboration with the Director of Strategic Evaluation at Natural Resources Canada (NRCan) and Environment Canada (EC) in order to develop a better understanding of innovations being used in federal government program evaluations so that evaluators can better meet the needs of clients and management.

Part of this study involves the conduct of interviews with key participants. Tyler Toso, a former Environment Canada Co-op student from the University of Victoria's School of Public Administration, is leading the project and will be conducting the interviews.

As one of a group of potential participants, you may be contacted in the following weeks by Mr. Toso and asked to participate in an interview. Your views would make a valuable contribution to this project.

The purpose of this interview would be to obtain information regarding innovation in evaluation both in your department and more generally. An interview guide would be sent to you to assist in preparation for the interview. These interviews are voluntary and will be scheduled at a time convenient for you. Please note that information obtained in interviews will be confidential, and you will not be required to answer any questions you feel unable to address.

Thank you in advance for your willingness to participate in this process. If you have any questions about this study, please contact Tyler Toso at ttoso@uvic.ca or (613) 612-2333, or Gavin Lemieux, Director of Strategic Evaluation, at Gavin.Lemieux@NRCan-RNCan.gc.ca or (613) 996-9649

Regards,

Gavin Lemieux
Director/Directeur
Strategic Evaluation /Évaluation stratégique
Science and Policy Integration/Intégration des sciences et des politiques
Natural Resources Canada/Ressources naturelles Canada
588 Booth Street
Ottawa, Ontario,
Telephone : 613-996-9649

Invitation to Participate Email

As noted in an earlier email sent by Gavin Lemieux, a study on Innovation in Evaluation is currently being conducted by Natural Resource Canada and Environment Canada's Evaluation Divisions'. The study should help to develop a better understanding of how innovation is occurring in evaluations within the Federal Government in order help Evaluation Divisions better meet the needs of clients and management.

I would like to invite you to participate in an interview as part of this study. The interview is voluntary; results will be used confidentially, and can be withdrawn at any time. The interview should take approximately 30-45 minutes. The interview guide is provided below for your review. You will only be asked to respond to the questions that you feel comfortable answering.

As this project is also being completed for academic purposes, as part of my MPA Master's Project at the University of Victoria's School of Public Administration, a participant consent form for the interview is provided below. Please read over this document. You may provide consent by signing and returning the document by email or at the time of the interview.

I am conducting interviews in May and June 2014. If you are interested in participating in this study please feel free to contact me directly by phone or email in order to schedule a time most convenient to you.

Thank you.

Tyler Toso
MPA Student
University of Victoria
toso@uvic.ca
613-612-2333

Interview Consent Form

Participant Consent Form

Project Title: Understanding and Applying Innovation in Federal Government Program Evaluations

Researcher: Tyler Toso, Graduate Student, Master of Public Administration, University of Victoria, (613) 612-2333, ttoso@uvic.ca

Supervisor: James McDavid, Professor, School of Public Administration, University of Victoria, (250) 472-4293, jmcdavid@uvic.ca

Clients:

Gavin Lemieux, Director, Strategic Evaluation (Natural Resources Canada)
Gavin.Lemieux@NRCan-RNCan.gc.ca, 613-996-9649

William Blois, Director, Audit and Evaluation Branch (Environment Canada)
William.blois@ec.gc.ca, (819) 956 7612.

Description:

As mandated through Treasury Board Policy, all federal government departments must evaluate all program spending. The Strategic Evaluation of Natural Resources Canada and The Evaluation Division, a division of the Audit and Evaluation Branch at Environment Canada (EC) are responsible for producing evaluations of programs that provide information on the relevance, effectiveness, efficiency, and economy of departmental policies, programs, and initiatives. This is mandated through Treasury Board Policy.

The performance of these evaluation divisions is assessed yearly in regards to achievement of certain Management Accountability Framework (MAF) criteria set and evaluated by Treasury Board of Canada Secretariat (Treasury Board Secretariat of Canada, 2012e).

The project seeks to answer these questions: What are the novel trends and innovative tools in evaluation and how are these being used in federal government organizations to improve evaluations? What lessons can be learned by Evaluation Divisions in order to promote continuous improvement and receive an ‘acceptable’ ‘MAF’ assessment?

Participation: You have been selected as a potential participant in this research because of your knowledge of and experience with federal government program evaluations. Your expertise and your opinions are important and will help to increase our understanding of novel methods and tools being used in program evaluation, and how they might be applied to provide more effective and efficient evaluation services.

Participation will require approximately 30-45 minutes for an interview. Interviews will take place either at your office or an agreed upon location. Each participant will receive a copy of the final report. There are little to no risks of participating in the research other than the inconvenience of the time it will take to participate.

Participation is voluntary and you may withdraw at any time. Should you wish to withdraw from an interview, the data you have already will be destroyed and not included in the final report.

Anonymity and Confidentiality: Names of participants and their departments or agencies will remain confidential and will not be attached to the final report. Any information that would make it easy to identify participants (name, contact info) will be removed from the analysis.

Participants will also be given the opportunity to review applicable sections of the report to ensure anonymity prior to publication.

There are some limits to the researcher's ability to protect the confidentiality of participants as there are limits due to context: The nature or size of the sample from which participants are drawn may make it possible to identify individual participants. Given the small community of evaluators in the federal government, complete confidentiality cannot be absolutely certain. Raw data that includes identifiers will be kept only on the researcher's personal home computer and will be password protected and locked up in a desk when not in use. Written notes will be kept in a locked desk until the completion of the final report, at which time they and the raw data will be destroyed using an approved Top Secret Shredder and erased from the computer by deleting and clearing the recycle bin.

Results: The results of the research may be used/disseminated in a report that will be distributed to participants, the University of Victoria, Natural Resources Canada, Environment Canada and any other interested party. The results may also be used in presentations pertaining to the report including the Master's defence.

Questions and Concerns: If you have any questions or concerns, please contact the researcher using the contact information at the top of page 1. You may also contact the Human Research Ethics Office at the University of Victoria, (250) 472-4545, ethics@uvic.ca.

Consent: Your signature below indicates that you understand the conditions of participating in this study, as stated above, and that you have had the opportunity to have your questions answered by the researcher and consent to participate in this research project.

Name of Participant

Signature

Date

The researcher and participant will each receive a copy of the signed consent form.

Innovation in Evaluation Interview Questions Guide

Thank you for taking the time to meet with me today to participate in this interview. As I mentioned in the invitation, I am a Master of Public Administration student at the University of Victoria. I am conducting this project on behalf of Natural Resources Canada and Environment Canada's Evaluation Divisions, with the aim to learn more about innovation occurring in evaluations in the federal government and elsewhere.

The project seeks to answer these questions: What are the novel trends and innovative tools in evaluation and how are these being used in federal government organizations to improve evaluations? What lessons can be learned by these Evaluation Divisions in order to promote continuous improvement and receive an 'acceptable' 'MAF' assessment?

The interview should take approximately 30-45 minutes. You may also choose to end the interview at any point – no reason is required. You may also choose to skip a question at any time. Do you have any questions or concerns before we get started?

Background information

1. I understand that your position is (title) at (name of organization). How long have you been in this position?
2. How long have you been with the organization?
3. What is the nature of your responsibilities?

Innovation in Evaluation

Innovation generally occurs when a new and novel design, approach or method is adopted (perhaps as a pilot project) and if successful has the potential to add value by addressing a user's needs. Examples of innovation might include the adoption of a newly developed project design or approach that reduces the overall time or cost to conduct an evaluation, the use of a method that improved the quality of evaluation results, etc.

4. Considering evaluations you are familiar with in the last 5 years (Since 2009), could you identify any innovative approaches that were used to plan, design or to carry out the work? (Ask to clearly identify the 'value added' of the innovation)
5. How successful have these innovations been for you? Which ones would you say have been the most successful so far? Why?

Divisional Innovation

I want to explore several specific potential innovations and see if any of these have been tried by your unit.

6. Please describe, if any, innovations that were used in **data display** or in **presentations of evaluation results**? (How were these successful? Were there any drawbacks?)
7. Please describe, if any, any innovations that have involved the use of (Geographic Information System) **GIS**? (How were these successful? Were there any drawbacks?)
8. Please describe, if any, any innovations that have involved the use of **Social Networking/Social Media**. (How were these successful? Were there any drawbacks?)
9. Please describe, if any, innovations that have involved **Big Data**? (How were these successful? Were there any drawbacks?)
10. In your organization, what is being done departmentally to generate innovative ideas?

11. Are there are examples of innovative ideas that have been discussed but were never implemented?

Barriers and Resources

12. What barriers or limitations does your organization face in developing or implementing innovation into evaluations?
13. What would be required to better implement innovations into evaluations?

Need for Innovation

14. From your perspective, is there an ongoing need for innovation in program evaluation?
15. Do you have any additional thoughts or suggestions related to innovation in evaluation?

Thank you for taking the time to meet with me today to participate in this interview.

10.5 Appendix 5 – Operational-Efficiency Analysis

Operational-efficiency analysis (OEA) Examples of analytical approaches include:

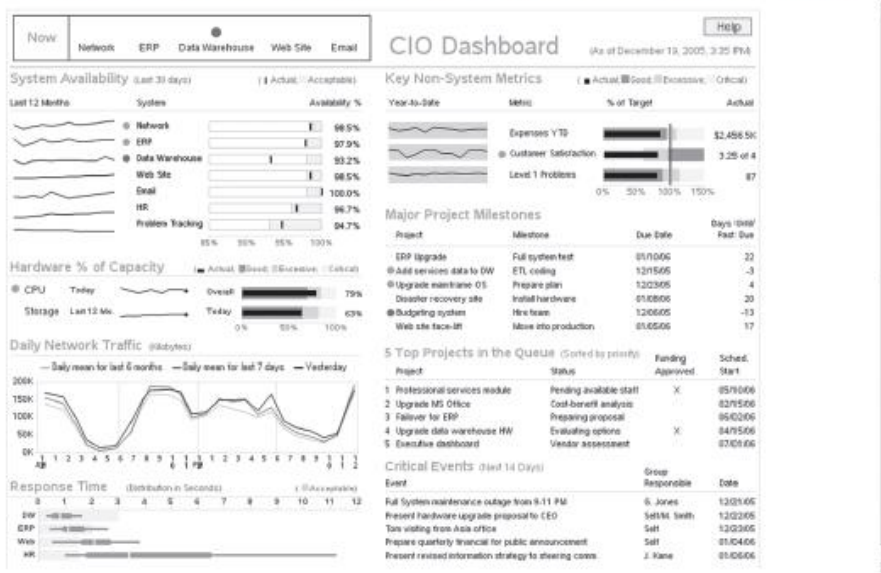
- Benchmarking
- Planned to actual cost comparison and analysis/expenditure tracking and analysis
- Business process mapping and analysis (determining challenges in key processes, e.g., bottlenecks)
- Fidelity Assessment/testing theory of implementation (determine if rationale for variances)
- Participatory Appraisal
- Optimization Analysis/expert opinion
- Comparison with alternative program models (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p.13-14)

Identification of costs:

- Operating costs - (salary and benefits, accommodations, supplies, workstations, professional services, direct corporate overhead)
- Capital Costs – Costs of developing new capital infrastructure and cyclical replacement of capital items
- Services from other government departments (sometimes portions of services are delivered)
- Non-Administrative disbursements (grants/contributions/transfer payments to third parties)
- Non-government costs: costs borne by parties other than the government of Canada: participant's costs (user fees, application costs, compliance costs) (Assessing Program Resource Utilization When Evaluating Federal Programs, 2013, p.17-18).

10.6 Appendix 6 – Dashboard

Figure 1.10. Dashboard for Program Implementation and Outcomes



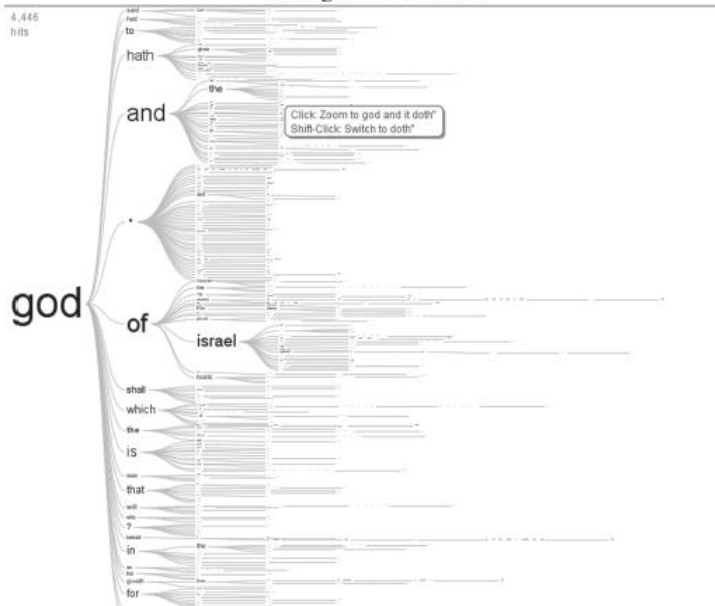
Source: Few (2006).

(As cited in Azzam et al., 2013, p.21).

10.8 Appendix 8 – Interactive Phrase Tree

This is a **static** representation of an interactive phrase tree.

Figure 1.8. An Interactive Phrase Tree Representing the Use of the Word *god* in the Bible



Source: Created with www.many-eyes.com

(As cited in Henderson et al., 2013, p. 58).

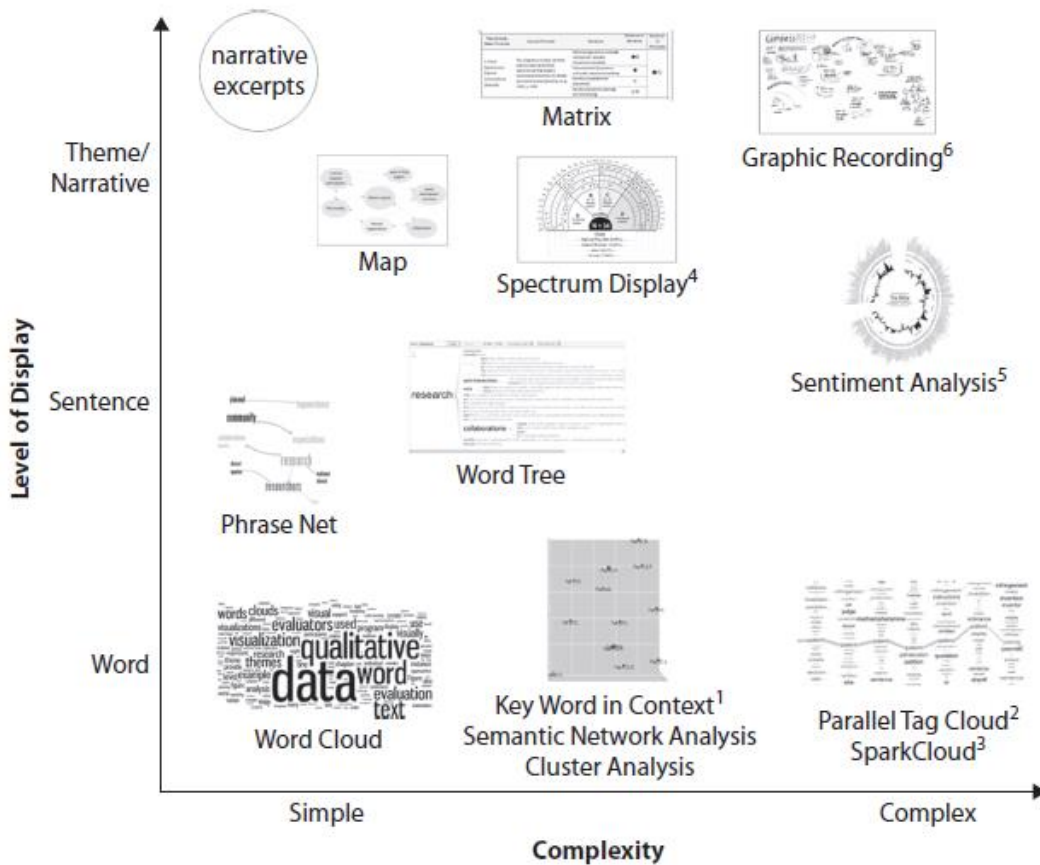
10.9 Appendix 9 – List of Data Visualizations

Below is a list of data visualizations:

- Data Visualization Suites
 - Google Fusion Tables
 - Google’s Public Data Explorer
 - Many Eyes
 - Microsoft Excel
 - Tableau
- Mapping Tools
 - Bing Maps
 - Google Maps
 - ArcGIS
- Finishing Tools
 - Adobe Illustrator
 - Adobe Flash
- Qualitative Visualization Tools
 - MAXQDA
 - NVivo
- Word Cloud Generators
 - Wordle
 - Tagxedo (Kistler et al., 2013).

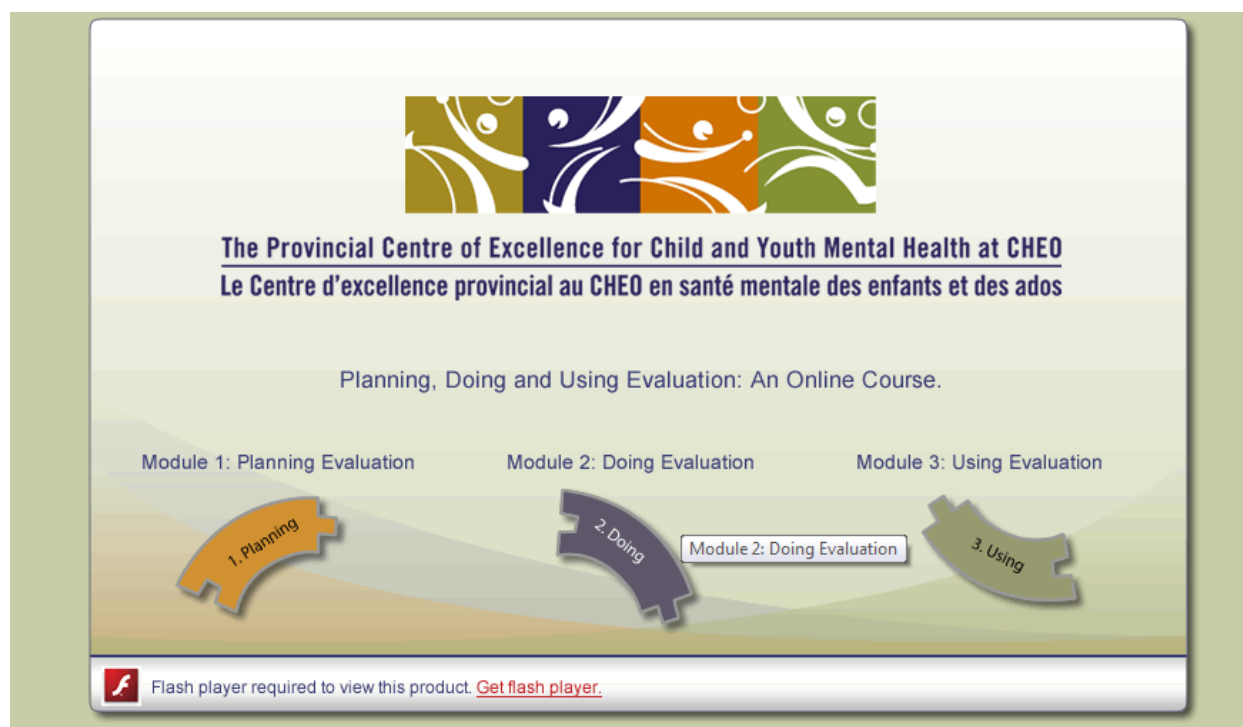
10.10 Appendix 10 – Data Visualizations by Complexity

Figure 3.1. Graph Showing Variety of Qualitative Visualizations by Display and Complexity



“The degree of complexity does not refer to the complexity of what is seen in the visualization, but rather the knowledge or skill level required to develop the visualization” (Henderson et al., 2013, p.56-57).

10.11 Appendix 11 – Online learning Modules



The above site can be accessed through this link:

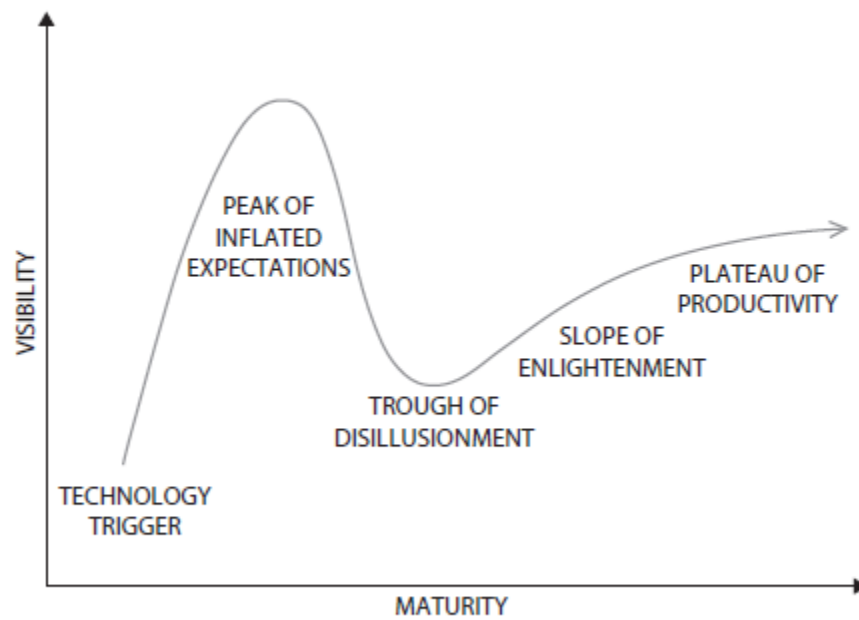
http://www.excellenceforchildandyouth.ca/sites/default/files/docs/olm_e/index.htm

The purpose of creating this module was based on the following premises:

“By engaging with the modules, the hope was that users learn how to:

1. Identify and describe the steps and tasks associated with planning an evaluation project,
2. Describe key evaluation concepts and apply these to their particular settings,
3. Conduct an evaluation project,
4. Use evaluation findings to improve the effectiveness and efficiency of their programs, and
5. Identify resources to build further capacity in evaluation” (Sundar et al., p.99, 2011).

Figure 1.17. Depiction of the Hype Cycle

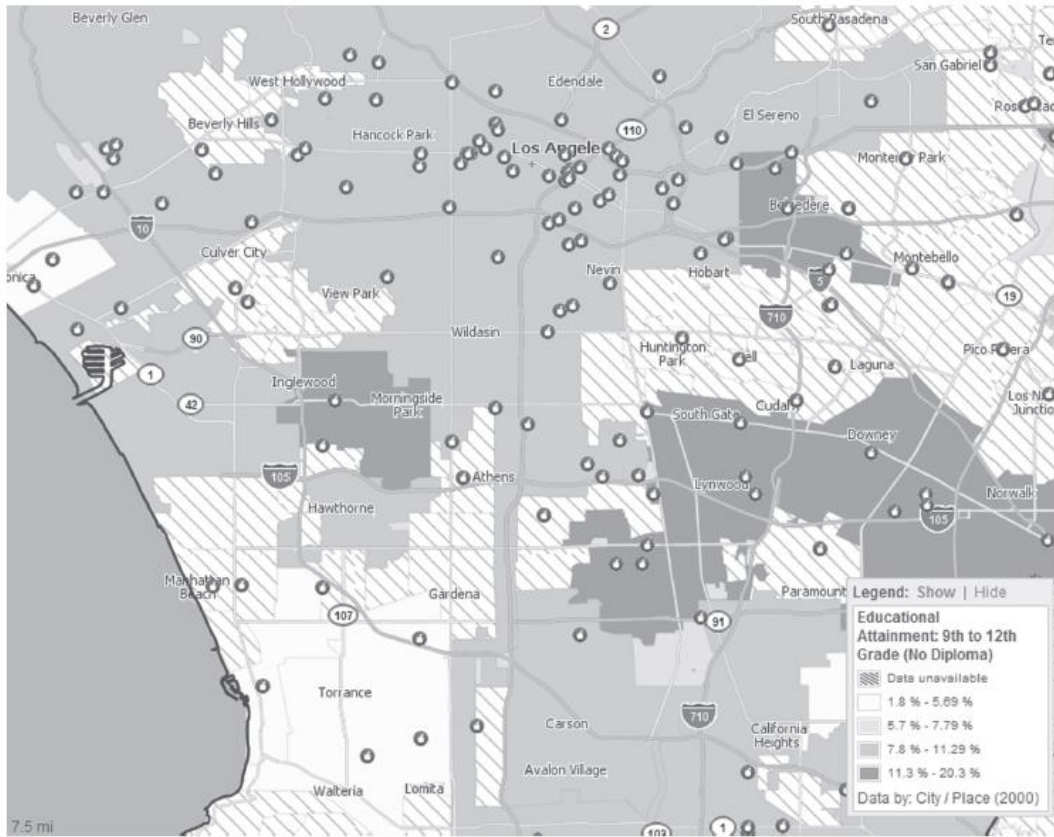


Source: <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp#>

(Azzam et al., 2013, p. 29).

10.13 Appendix 13 – Example of GIS

Figure 1.6. An Example of Using GIS to Conduct a Community Needs Assessment



Source: Created with <http://www.healthycity.org/>

(Azzam et al., 2013, p. 16).