Software Engineering Institute

The Swiss eGov Case: “Metadata 2010”

Philip Boxer
Hans Sassenburg

January 2010

SPECIAL REPORT
CMU/SEI-2010-SR-003

Research, Technology, and System Solutions Program
Unlimited distribution subject to the copyright.

http://www.sei.cmu.edu

Carnegie Mellon

This report was prepared for the

SEI Administrative Agent
ESC/XPK
5 Eglin Street
Hanscom AFB, MA 01731-2100

The ideas and findings in this report should not be construed as an official DoD position. It is published in the interest of scientific and technical information exchange.

This work is sponsored by the U.S. Department of Defense. The Software Engineering Institute is a federally funded research and development center sponsored by the U.S. Department of Defense.

Copyright 2010 Carnegie Mellon University.

NO WARRANTY

THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

Use of any trademarks in this report is not intended in any way to infringe on the rights of the trademark holder.

Internal use. Permission to reproduce this document and to prepare derivative works from this document for internal use is granted, provided the copyright and "No Warranty" statements are included with all reproductions and derivative works.

External use. This document may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

This work was created in the performance of Federal Government Contract Number FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. The Government of the United States has a royalty-free government-purpose license to use, duplicate, or disclose the work, in whole or in part and in any manner, and to have or permit others to do so, for government purposes pursuant to the copyright license under the clause at 252.227-7013.

For information about SEI publications, please visit the library on the SEI website (http://www.sei.cmu.edu/library).

Table of Contents

Abstract **v**

1 Introduction **1**

2 Demand Analysis **2**
2.1 Themes, Complexity and Costs of Questions 2
2.2 Preliminary Conclusion 4
2.3 Selected Investment Options 4

3 Main Analysis **5**
3.1 A Model of the Number and Variety of Questions 5
3.2 The Costs of Different Complexities of Questions 5
3.3 Value of the Different Investment Options 6

4 Conclusions **10**
4.1 On the Faceted Option 10
4.2 On the Advanced Option 10

5 Recommendations **11**
5.1 Summary 11

List of Figures

Figure 1:	Transient versus Stable Themes	2
Figure 2:	Complexity of Questions	3
Figure 3:	Annualized Cohesion Costs (= Sum of Direct costs and Alignment Costs)	4
Figure 4:	Distribution of Questions	5
Figure 5:	Costs Analysis for Different Categories of Questions	6
Figure 6:	Average and Variance on Savings for Faceted Option	6
Figure 7:	Average and Variance on Savings for Advanced Option	7
Figure 8:	NPV Analysis – Faceted Option	7
Figure 9:	NPV Analysis – Advanced Option	8
Figure 10:	Results Real Option Analysis	8
Figure 11:	Graphical Summary	11

Abstract

eGov projects are designed to impact on the relationships possible between the citizen and government, typically including relationships with multiple parts of government at different levels of government. As a result, the effects of eGov projects cannot be justified based on the direct effects of a project because the main benefits are indirect, intended as they are to affect the ways in which government activities can be aligned to the needs of the particular citizen. Multi-sided markets provide a way of describing the relationship between a supplier of services and the value of indirect benefits.

The traditional basis for justifying an investment is in terms of the supply-side economies of scale or scope generated by the direct effects of the investment on existing working practices and/or with respect to new revenues that can be captured. In the case of eGov projects, no such justifications are possible because all the effects are indirect, given the multi-sided nature of the demands being met. In their place demandside economies have to be identified, being economies in the costs of aligning government activities to the needs of particular citizens. A way of meeting the need to express monetary effects can be found in valuing the impact of introducing new eGov-investments on these costs of alignment, through the use of real option analysis

This special report describes the results of a feasibility study executed in October-December 2009 for the Swiss Government. Fundamental objective was to be able to assess the value of different investment options aiming to increase the responsiveness of the Swiss Government to questions asked of it by its citizens and enterprises. The four different steps in the study are described, including their outcomes: (1) demand analysis regarding the questions raised, (2) main analysis to compare the different investment options considered using Monte Carlo simulation and real option analysis, (3) conclusions regarding the study, and (4) recommendations for a follow-up project.

1 Introduction

The Swiss Federal Chancellery provides services to the Federal Council and the Federal Administration as a whole, as well as to the general public. With the use of information and communication technologies, the activities of the administration should become as efficient and as close to citizens and enterprises as possible. This is what is meant by the term e-government. This potential can only be fully developed when synergies between the various services and between the different federal levels are fully exploited. This represents a special challenge for Switzerland due to its state structure.

Users should be able to fulfill their information demand by using the websites of the Federal Government. If users do not find the information this objective is not met (and this is a common scenario that has been verified in a usability study). The project "Metadata 2010" tries to improve the search engine with new technology, improved usability and the use of metadata in order to ease access to the information (improve "findability" of information).

In order to estimate the benefits of such an investment, more quantitative insight is needed with respect to the current demand as well as incurred costs to fulfill this demand. A preliminary study was executed during the period October-December 2009 to gain this insight and use the analyzed results to evaluate different investment options. Available data regarding the swine flu was used as the main reference. The results of this study are summarized in this report.

2 Demand Analysis

2.1 Themes, Complexity and Costs of Questions

Current perception of Swiss Government is that e-government is about automating human efforts regarding questions ("we need a better search engine"). The underlying assumption is that questions that arise are relatively simple, relate mostly to stable themes and are easy to standardize. In this study however, it was revealed that a significant percentage of the questions dealt with were transient (= not stable) as in Figure 1.

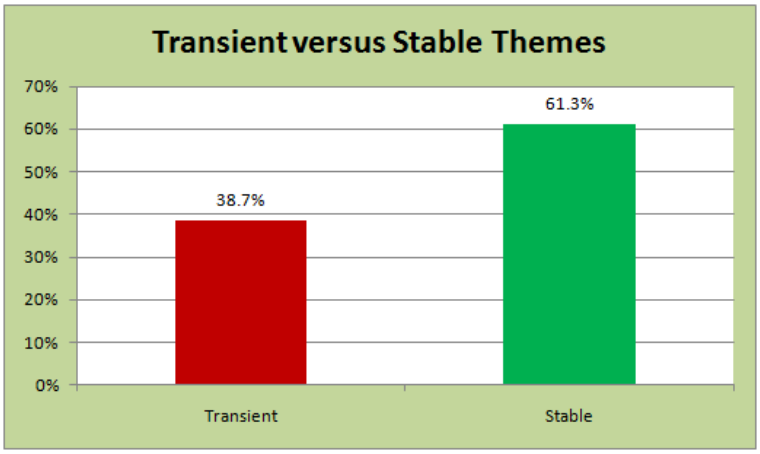

Figure 1: Transient versus Stable Themes[i]

In addition, it was found that many questions are complex (= not simple), and therefore not always that easy to standardize. A distinction was made between 4 categories of questions/responses (in ascending level of complexity and associated costs):

- Wholly *standardized* responses
  - There is a standard response available to the question, usually in the form of Frequently Asked Questions (FAQs)
  - Example: does the government provide face-masks?
- *Customized* responses
  - The response involves adapting previous responses to the particular circumstances of the questioner
  - Example: where can I get a vaccination?
- Responses dependent on specialist *knowledge*
  - The response involves orchestrating a number of different sources to address the particular circumstances of the questioner
  - Example: can I be vaccinated with my current medication?
- Problems needing wholly *new* responses

- In order to respond, the question as presented has to be re-formulated in a way that depends on the particular circumstances of the questioner before it is possible to develop a response
- Example: what should I do about preventing an epidemic at my school?

Results regarding complexity distribution for both stable and transient themes are summarized in Figure 2.

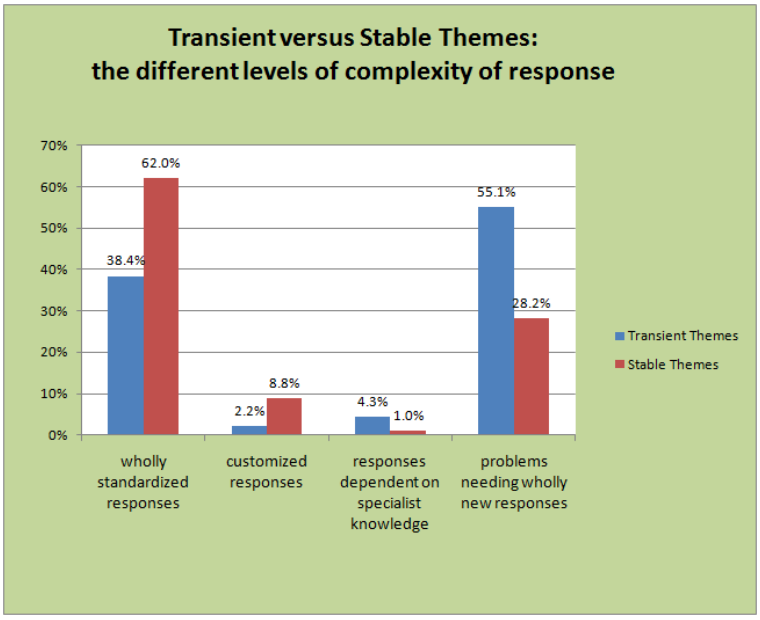

Figure 2: Complexity of Questions[ii]

In addition, the study revealed that dealing with these transient and complex themes incurred high *cohesion costs*, being the sum of direct costs incurred by departments and costs of aligning the work of multiple departments to provide a complete response (referred to in this study as the costs of alignment). See Figure 3.

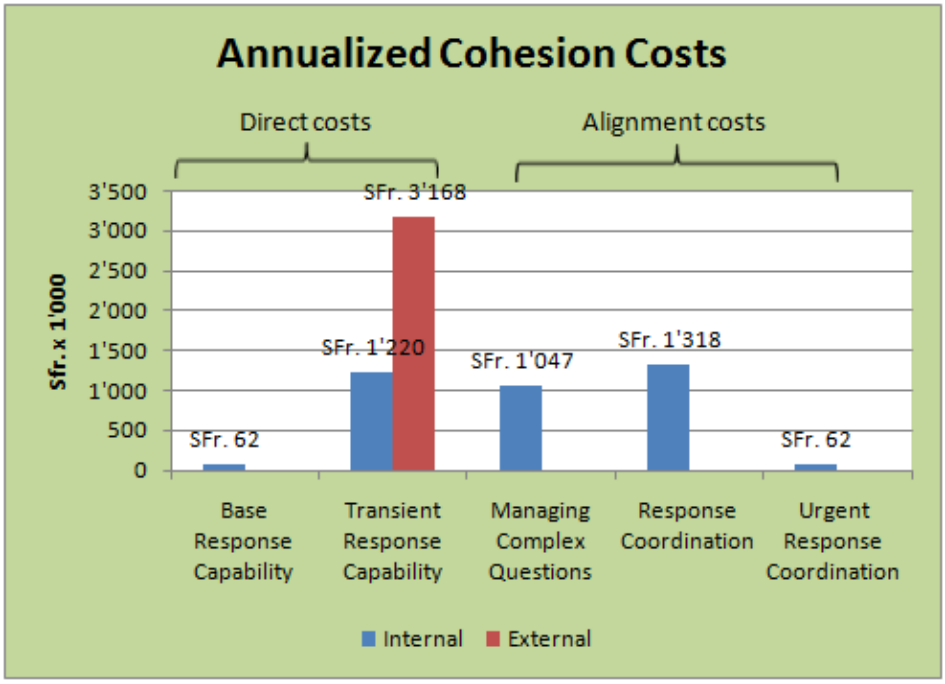

Figure 3: Annualized Cohesion Costs (= Sum of Direct costs and Alignment Costs)[iii]

2.2 Preliminary Conclusion

The preliminary conclusion was therefore that the Swiss Government should look for a technical approach that helped them to cope better with transient themes, at the same time offering the possibility to standardize responses to questions faster and more easily. Both attributes would help reducing cohesion costs. The solution should also be designed to support the variety of forms of collaboration needed between Swiss Government departments in responding to the variety of types of question from its citizens and enterprises.

2.3 Selected Investment Options

It followed that two investment options needed to be evaluated against the base case ("status quo"):

1. A search engine using a *Faceted* approach for identifying content, based on the manual codification of documents.

2. A search engine using an *Advanced* approach for identifying content, in which the coding of documents is determined by an automated semantic analysis of their content.

3 Main Analysis

To consider both investment options, there were three steps to the main analysis:

3.1 A Model of the Number and Variety of Questions

A model of the number and variety of questions having to be responded to by Government was built, using Monte Carlo simulation based on data gathered. Detailed quantitative data for this simulation and subsequent analysis are available, results being summarized in Figure 4.

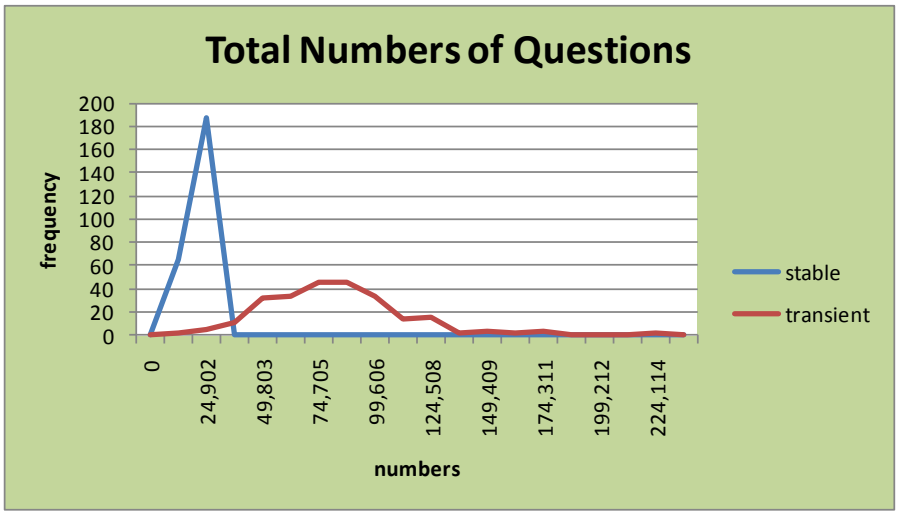

Figure 4: Distribution of Questions[iv]

3.2 The Costs of Different Complexities of Questions

The same set of question categories (based on varying complexity and the cohesion costs) were used to identify an initial estimate of the impact of each investment option on those costs.

Costs identified for each category are summarized in Figure 5. The cohesion cost multiple is the ratio of the total cohesion cost to the direct costs incurred by departments, indicating the extent of the costs of alignment needed to respond to the question. For the non-standardized responses, it can be seen that this ratio ranges between 2x and 4x.

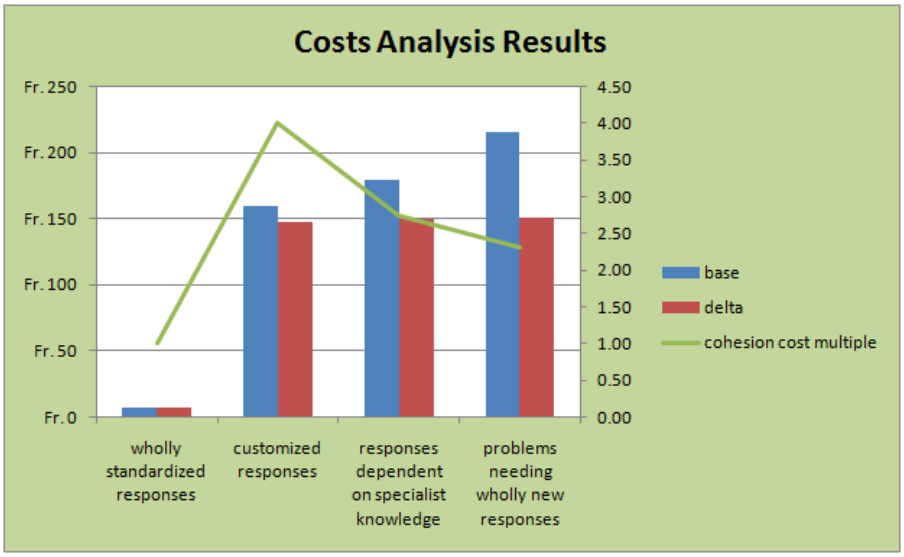

Figure 5: Costs Analysis for Different Categories of Questions[v]

3.3 Value of the Different Investment Options

There were three sub steps to this analysis step.

1. The average and variance on savings for each option was calculated against the base case. See Figure 6 and Figure 7. The average saving for the Advanced option was Sfr. 2M, compared with an average saving of Sfr. 250K for the Faceted option. The difference reflects the ability of the advanced option to enable transient questions to be dealt with more effectively using the search engine.

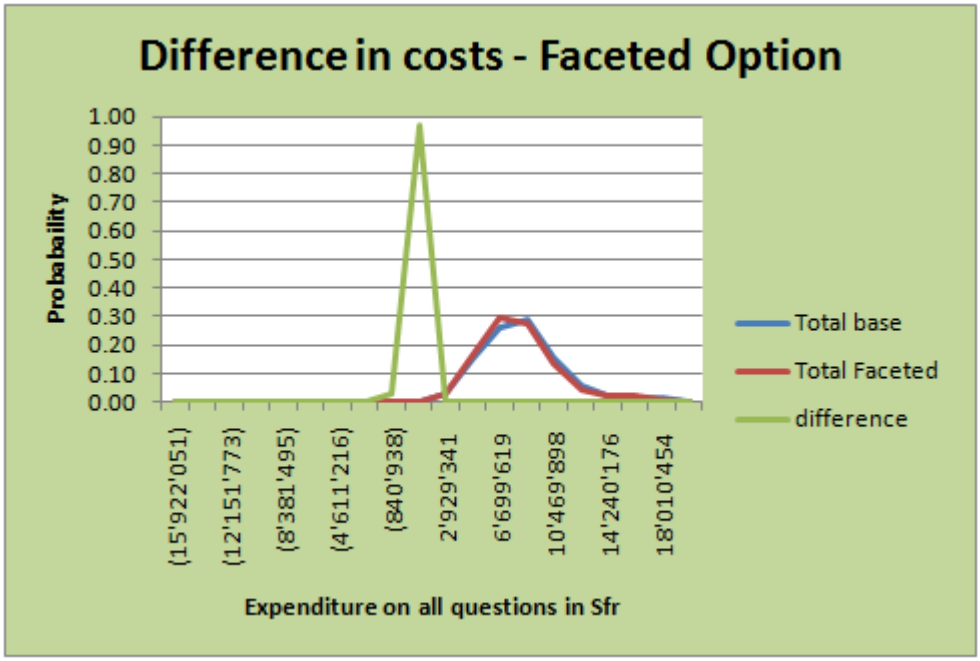

Figure 6: Average and Variance on Savings for Faceted Option[vi]

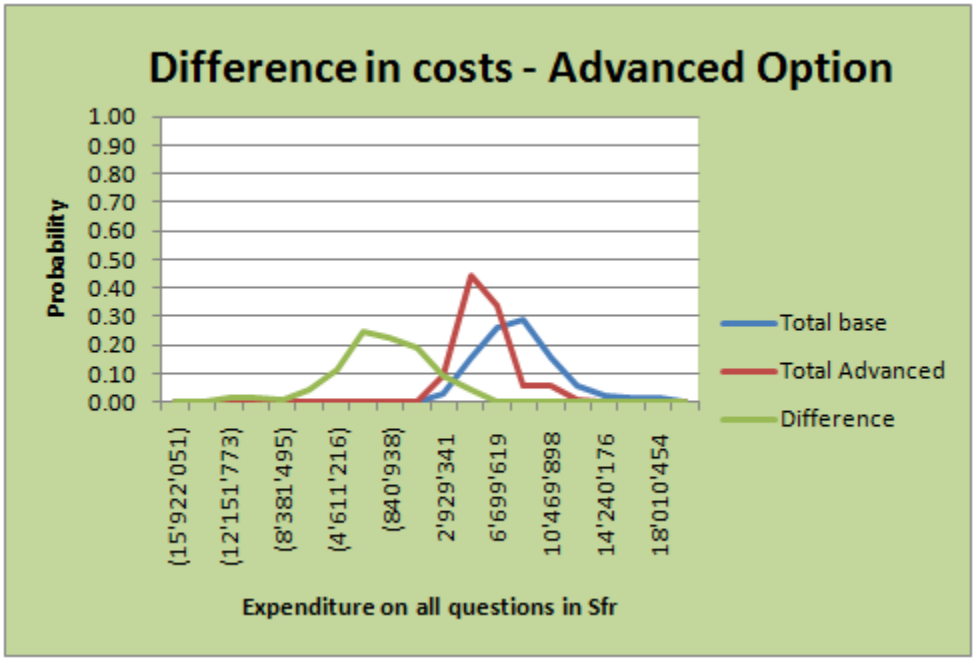

Figure 7: Average and Variance on Savings for Advanced Option

2. Using average values for costs and savings, the fully discounted Net Present Value (NPV) was calculated for each investment option. This included securing the current system Convera (Phase 1) till the transition to a new system, either Faceted or Advanced (Phase 2). For each option, Phase 2 included the savings on the Phase I base case, and the incremental costs in 2016 include a terminal value for the project based on another 10 years' costs and savings. Results are given in Figure 8 and Figure 9 for both investment options.

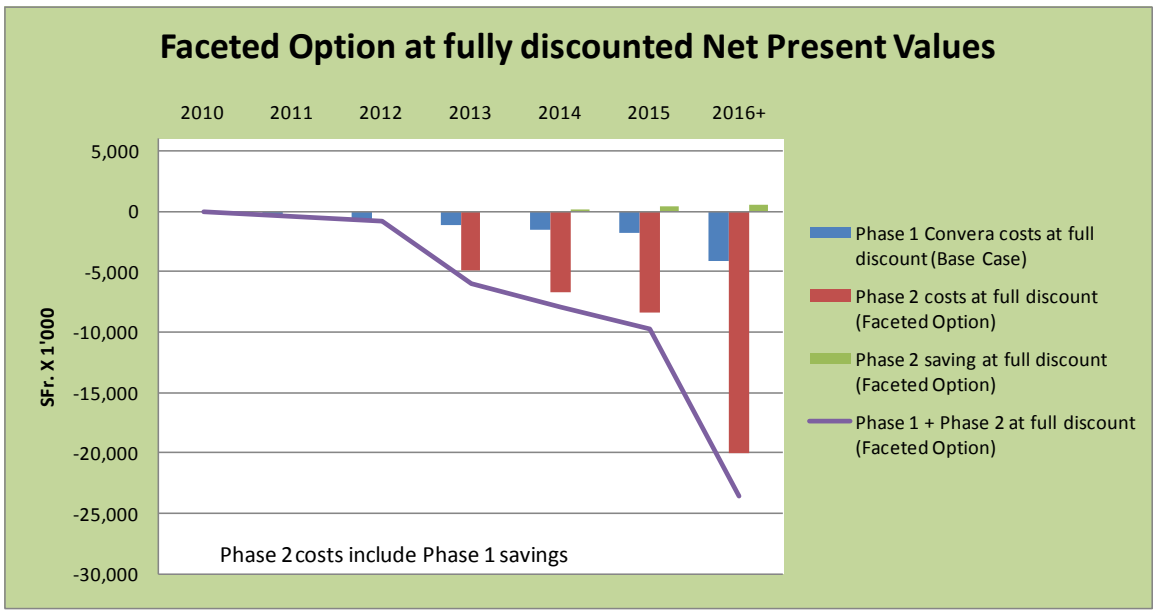

Figure 8: NPV Analysis – Faceted Option[vii]

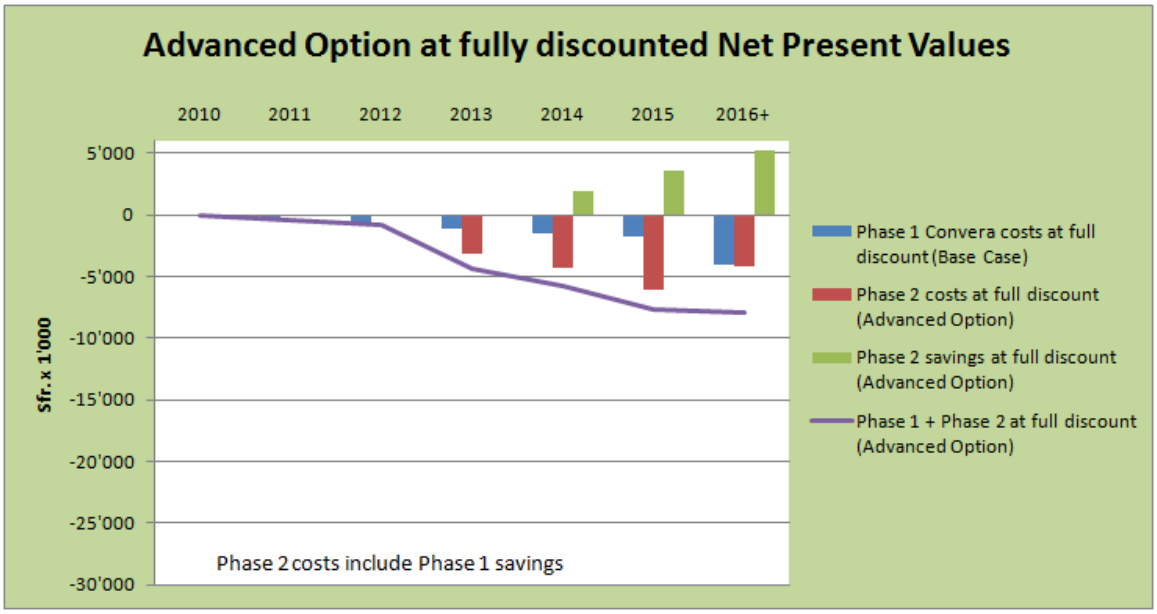

Figure 9: NPV Analysis – Advanced Option

This NPV analysis clearly shows the superiority of the Advanced option (NPV = ~Sfr. 4.5M) over the Faceted Option (NPV = ~Sfr. 23.5M), but does not take into account the impact of the variability of the savings driven by the variability in the numbers of questions.

3. Real Option Analysis was then used to attach a value to the way costs varied with the variations in the numbers of questions, arising as a consequence of the investment. These variations are shown in Figure 7 for the Advanced option, and the additional value is shown in Figure 10.

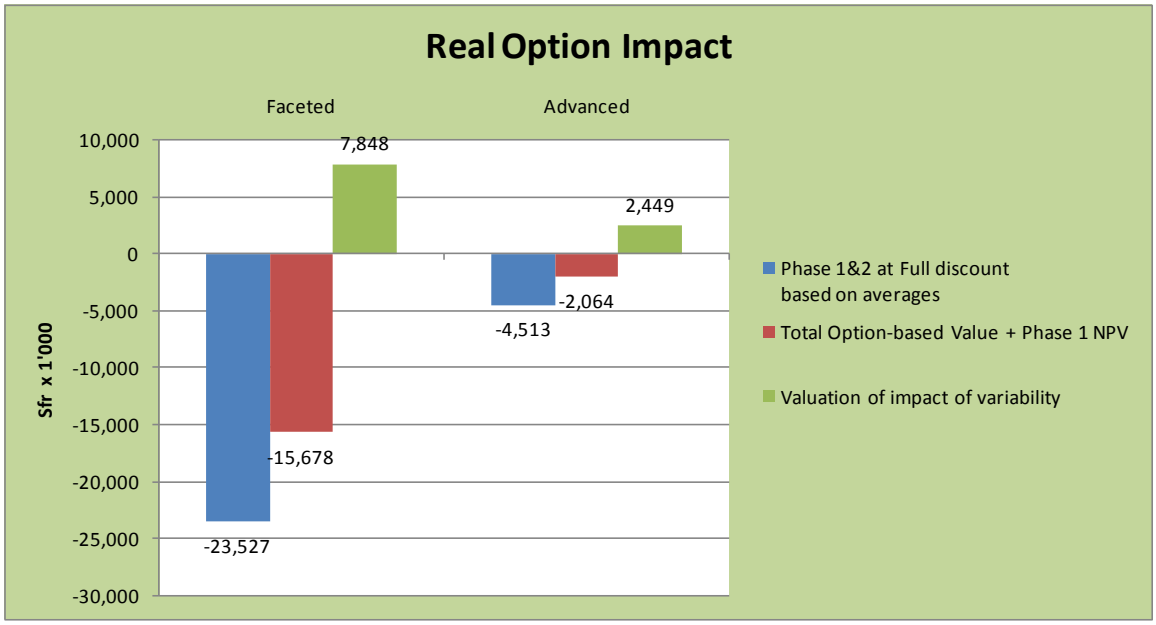

Figure 10: Results Real Option Analysis[viii]

When the impact of this expected variation is included, it further reduces the expected overall cost of both options considered, reducing the expected cost of the Advanced Option to ~Sfr. 2M.

4 Conclusions

The model used for demand analysis, costs analysis and investment option evaluation is considered powerful. Although the quality of available data available during this study was incomplete and limitedly validated (the correction of which would be one of the objectives of a follow-up project), some preliminary conclusions are available when comparing the two investment options considered.

4.1 On the Faceted Option

It became apparent that the Faceted option was the less attractive option for two reasons:

1. The higher initial investment cost made the Faceted option less attractive compared to Advanced option, using a NPV comparison of average costs and savings.
2. Because of the costs of encoding additional documents as themes changed, the Faceted option impacts on stable themes only, and does not accelerate standardization.

4.2 On the Advanced Option

The Advanced option was considered the most attractive candidate as it would:

1. Cost less on the basis of a NPV analysis using average costs and savings.
2. Reduce the variability in costs through increasing the agility of government's responsiveness to citizens and enterprises, thus further increasing the value of this option.

5 Recommendations

Final recommendations are therefore:

1. The analysis approach used by the study should be adopted. This approach addresses both the supply-side economics of the investment itself, and also the demand-side economics associated with the impact of the investment on the costs of aligning government activities to the needs of citizens and businesses. The resulting analysis expresses the consequence of an investment as monetary effects on the Government's costs of responding to the questions of citizens and enterprises.

2. This study was a feasibility study only, with access to incomplete and partial data based on sampling. Therefore, it should be re-executed on a broader and more detailed level.

3. For the selected investment option (Advanced), an analysis of its quality attributes should be used to establish detailed architectural recommendations. This will assure that cost and saving advantages are maximized.

5.1 Summary

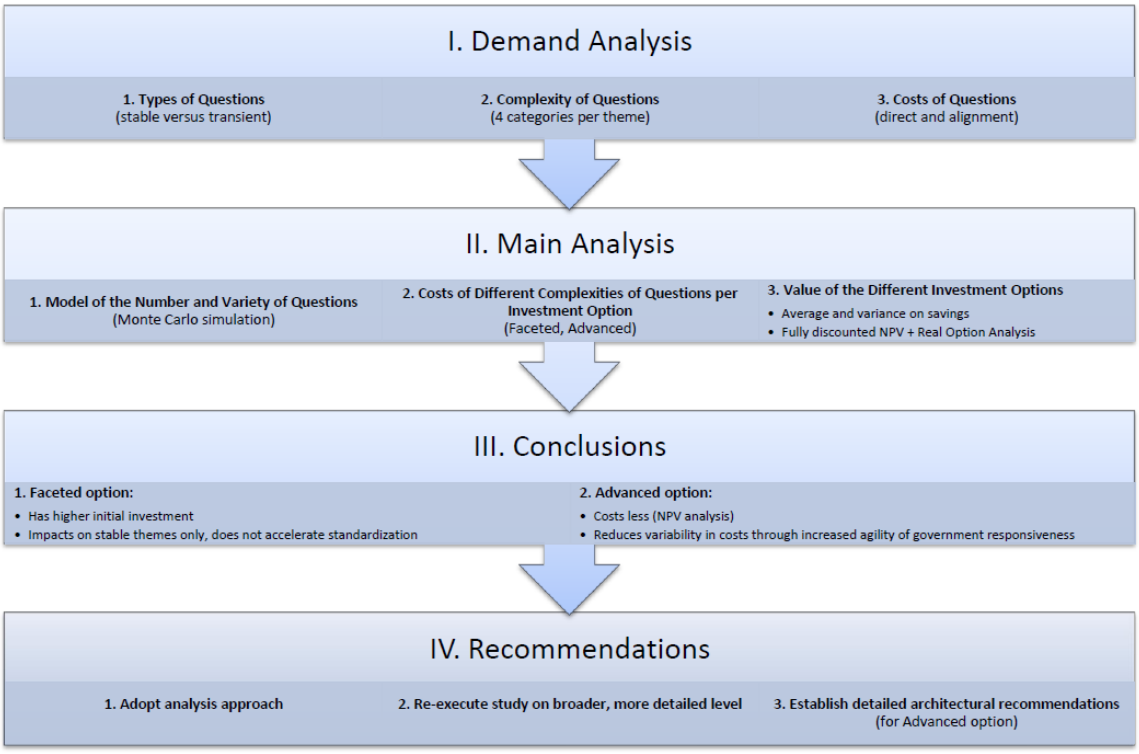

Figure 11: Graphical Summary

Endnotes

i Taken from *Option Analysis.xlsx [Summary variation + costs]*, this shows the total number of stable themes vs transient themes. This gives a different picture to the numbers of questions in Figure 4 because the numbers of questions per theme are much higher for transient themes.

ii Taken from *Press Conferences (Analysis).xlsx [Frequency Press Conferences]*, this shows the relative occurrence of different kinds of question. The differences reflect the number of different departments involved in responding to the question, in which departments may be both internal and external to Government. In general, more departments involved means more time managing the alignment of their partial answers in relation to the original question.

iii Taken from *Control Questions.xlsx [Control Figures]*, these are based on top-down estimates of the total capacity being used in dealing with transient questions. They include the costs of providing hotlines using external resources.

iv Taken from *Option Analysis.xlsx [Summary variation + costs]*, this shows the variation in the total numbers of stable and transient questions in the Monte Carlo simulation.

v Taken from *Cost Model.xlsx [Summary]*, this shows the costs for the base case and the delta for each type of response. It also shows the cohesion cost multiple in each case, being the ratio between the direct cost and the total cohesion cost.

vi Taken from *Option Analysis.xlsx [Summary variation + costs]*, this shows the variation in the total costs for the base case and for the faceted case, together with the difference between them. In this case the difference curve is high and tight because the difference is small and (in absolute terms) the variation around that difference is not great. This contrasts with the next Figure 7 in which we see a much bigger shift to the left of the red curve, and the difference is much more substantial.

vii Taken from *Option Analysis.xlsx [Realoption]*, this and Figure 9 show the costs of each option at net present values. The 2016 figure includes a terminal value based on a further 10 years, and the Phase 2 costs and savings include the savings from discontinuing the Phase I investment in Convera. These net present values are based on averages, and do not take account of the impact of the potential variation in the levels of saving.

viii NPV calculations are based on a single series of figures into the future (shown in Figure 8 and Figure 9) for which a net present value can be calculated. This ignores the potential future variation in the range of possible future savings, shown as the 'Difference' in Figure 7. What real option analysis does is to modify the NPV based on averages to take into account this variation. The result is an adjusted NPV. In Figure 10, the impact of this adjustment on the NPV based on averages is show as 'Valuation of impact of variability'.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
(Leave Blank)	January 2010	Final

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
The Swiss eGov Case: "Metadata 2010"	FA8721-05-C-0003

6. AUTHOR(S)

Philip Boxer & Hans Sassenburg

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213	CMU/SEI-2010-SR-003

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
HQ ESC/XPK 5 Eglin Street Hanscom AFB, MA 01731-2116	

11. SUPPLEMENTARY NOTES

12A DISTRIBUTION/AVAILABILITY STATEMENT	12B DISTRIBUTION CODE
Unclassified/Unlimited, DTIC, NTIS	

13. ABSTRACT (MAXIMUM 200 WORDS)

14. SUBJECT TERMS	15. NUMBER OF PAGES
	21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL