Chapter XI

DQ Options: Evaluating Data Quality Projects Using Real Options

Monica Bobrowski, Pragma Consultores, Argentina

Sabrina Vazquez Soler, Pragma Consultores, Argentina

ABSTRACT

Data plays a critical role in organizations up to the point of being considered a competitive advantage. However, the quality of the organizations' data is often inadequate, affecting strategic and tactical decision-making, and even weakening the organization's image. Nevertheless it is still challenging to encourage management to invest in data quality improvement projects. Performing a traditional feasibility analysis based on return on investment, net present value, and so forth, may not capture the advantages of data quality projects: their benefits are often difficult to quantify and uncertain; also, they are mostly valuable because of the new opportunities they bring about.

Dealing with this problem through a real options approach, in order to model its intrinsic uncertainty, seems to be an interesting starting point. This chapter presents a methodological framework to assess the benefits of a data quality project using a real options approach. Its adequacy is validated with a case study.

INTRODUCTION

Quality control and management have become competitive needs for most businesses today. Approaches range from technical, such as statistical process control, to managerial, such as quality circles. An analogous experience basis is needed for data quality.

Many of the problems that come up when using poor quality data are well-known to software engineers. The NEAT methodology (Bobrowski, Marre & Yankelevich, 2001) provides a systematic way to determine data quality so as to develop an improvement plan. The output is a diagnosis of the present data quality condition and an improvement plan that comprises both corrective and preventive actions (in order to maintain the quality standards finally met). In particular, NEAT bases its approach on the goal question metric framework (Fenton & Pfleeger, 1997), GQM from now and on, a framework used for metrics definition.

However, there are no serious studies aimed at providing a framework to analyze the convenience of investing in data quality improvement. Many organizations come to action when they find they have very poor data (lawsuits filed by clients, returned posts, networks that do not match reality, etc.). The analysis is ad-hoc and, generally speaking, it aims at assessing the initiative cost alone, thus submitting the decision to the resulting amount (high or low) (Loshin, 2001; Trillium Software, 2002).

The question then is how to justify a preventive approach to these issues? The first approach would imply conducting a classic feasibility analysis, using standard techniques: net present value (NPV), profitability index (PI), and internal rate of return (IRR). Nevertheless, there are many limitations when applied to the analysis of quality investment projects: the benefits of this kind of projects are usually difficult to quantify economically, basically because they are not direct: they are related to the opportunities they bring about. In addition, part of the economic impact is associated with cost-saving resulting from prevented problems, which is difficult to measure and is not captured by traditional indicators.

Within this context, it seems interesting to use a real options approach (Brealey & Myers, 2000; Amran & Kulatilaka, 1999; Brach, 2002) to model the uncertainty that exists with respect to the subsequent decision. This model also allows capturing the essence of the NEAT methodology (Bobrowski, Marre & Yankelevich, 2001), which presents the need of making a diagnosis to assess, based on its output, the convenience of implementing a corrective improvement action on the data and also to establish specific improvement expectations. This model would allow assessing the best investment that an organization can make to improve its data, considering the performance evolution of the quality investment and future benefit expectations.

There are records of the use of the real options model to assess different software engineering projects (Boehm & Sullivan, 2000; Sullivan, Chalasani, Jha & Sazawal, 1999). However, their use to assess the benefits of quality investments has not been studied yet. We believe this model offers an interesting potential which is worth exploring.

The objectives of this section are:

- To define a methodological framework to assess the benefit of data quality improvement projects using the real options approach.
- To validate the model proposed by means of a case study.

In this work, we use data and information interchangeably, meaning we use both raw and processed data. Although there are some differences, they are of no significance within this context.

THE NEAT METHODOLOGY

We present the NEAT methodology (Bobrowski, Marre & Yankelevich, 2001), based on theoretical models of the data lifecycle (Huang, Lee & Wang, 1999) as well as on our practical experience in many evaluations and controlled experiments (Bobrowski, Marre & Yankelevich, 1999). The NEAT methodology is a framework for guiding data quality evaluation, consisting of six general stages: elicitation; planning; measurement; diagnosis; treatment; and maintenance.

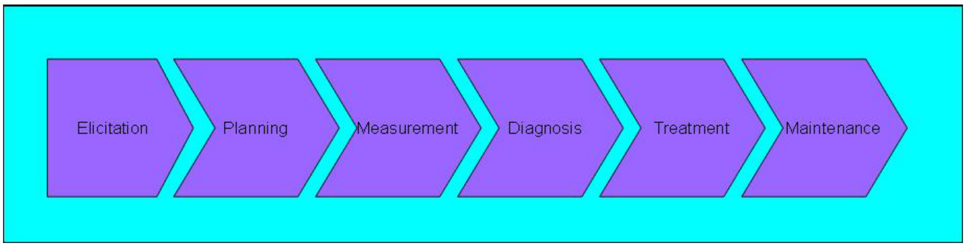
Each step is composed of several tasks that must be carried out in order to achieve particular goals. This is a brief description, but we think that it covers the main components and ideas of the method.

First Step: Elicitation (Data Life Cycle, Actual State and DQ Goals)

The first task during the elicitation step is the acquisition of the information on the data lifecycle in the organization.

During the second task of this step, we perform the evaluation of the state of the data in the organization. The outcome of this activity is a precise description of the actual state (a "snapshot") that will be used as a basis for many decisions during all steps.

Figure 1. NEAT steps



With the information of tasks 1.1 and 1.2, we can now establish which are the relevant quality dimensions for this particular project. Once these dimensions have been chosen, we must define the specific quality requirements for each dimension over different subsets of data. For each subset, we should establish two levels of quality: the target (optimum) level and the minimum level required.

Second Step: Planning

The second step is to develop a plan for assessment of quality. This step has a very concrete deliverable: a document, including the evaluation plan and how the plan is executed. The plan is defined at the very beginning of the step and the document is updated with each task.

Third Step: Measurement

The goal of this step is clearly to measure the quality of certain attributes. Thus, many tools and techniques are available: GQM tables, templates, data analysis tools, etc.

During this step, we perform the measurement, according to the measurement plan.

Fourth Step: Diagnosis

The outcome of this step is a document presenting the diagnosis of data quality. The step involves the interpretation of the measures, as well as reporting of findings. The document might be very useful by itself, but its main goal is to trigger actions that will change the status of data.

Fifth Step: Treatment — Corrective / Preventive

There are two main strategies to improve the quality of data. To improve existing data (through a corrective strategy, changing or deleting wrong data and including the data needed by the organization) or to improve the processes associated with data (data creation, consumption, updates, and so on). The latter prevents the injection of bad quality data into the system.

Sixth Step: Maintenance

When the desired quality goals are met, the project is still not over. We must find a mechanism to maintain the quality obtained.

Maintenance is strongly related to treatment, but clearly there are differences. Preventive actions should be used to prevent quality from decreasing in time, but it is also necessary to monitor the quality of data through time. This monitoring is based on systematic and periodic measurement, which in some cases can be partially automated.

This is an outline of the method used, which is descriptive enough to understand the following proposal and the case study, as well as discuss some conclusions.

IT PROJECT EVALUATION

Introduction

In 2000, Boehm and Sullivan published an article and advanced a new discipline called *software economics*. This field manages the analysis of improvements in software design and engineering, reasoning economically on matters related with products, processes, programs, portfolios and policies. They indicate that so far investigations have been mainly focused on costs, not on benefits.

Therefore, the value added by the investment has not been taken into consideration.

This lack of progress in the subject is based on several reasons (Benaroch & Kauffman, 1999), but the most important is:

- There are no suitable frameworks to model, measure and analyze the connection between technical characteristics, decisions and value creation. Central concepts of software engineering like concealment of information, architecture, spiral model, and so forth, have no clear relation with the value added to the organization.

It has been stated (Sullivan, Chalasani, Jha & Sazawal, 1999) that the real options approach could help to establish some of these relations (staggered design decisions, staged development, etc.).

- Most designers and software engineers are not trained to analyze the concept of value creation as a goal, or to manage technical aspects to add value to the organization. Measurements are usually focused on technical characteristics.
- The set of technologies, the regulatory and tax framework, the market and other structures in which software is developed and used are inadequate. Designers are not able to make decisions that could help to increase the value created by products.

Evolution

In 1969, Sharpe (the "father" of software economics) presented the first application of information economics to software-related issues (cited in Benaroch & Kauffman, 1999). The work included subjects such as purchase, leasing or rent of information systems, definition of service prices and analysis of the scale as an economic aspect. The work includes a small section dedicated to costs, based on the SDC study. This study formulated a model of linear regression to consider software costs. Although not very accurate, it encouraged further research work that gave rise to better cost models in 70's and 80's, some of which are still used nowadays, such as COCOMO, Estimacs, and so forth.

In his book *Software Engineering Economics*, Boehm (1981), in addition to the COCOMO model (Cost Constructive Model), presents a summary of the main concepts and macroeconomic techniques with some examples of their application to software development and software decision making (production functions, economies of scale, net value, present value, marginal analysis, etc.). Other contemporary works include a summary of cost estimates, costs comparisons of computational products and services, computer center management techniques, and so forth.

Some of the most important software engineering techniques implicitly include economic considerations (Benaroch & Kauffman, 1999):

- Software risk management uses principles of the theory of statistical decision to approach questions such as "how much formal verification is required" (to acquire data to reduce risk).
- Spiral or iterative life-cycles model use considerations on product value and risks to organize the increase in the capacity of applications.
- The notion of "design for change" is based on the recognition that a great part of the total product cost is incurred on its evolution (post production), and that a system that is not designed to evolve is going to incur enormous costs of maintenance (even limiting its capacity to adapt to new situations).

At the moment our capacity to analyze software costs exceeds our capacity to analyze its benefits. In recent years, there has been progress on the cost / benefit analysis of software projects, using all standard financial indicators to develop new case studies.

However, experience shows that it is difficult to evaluate a software project in the traditional manner, since they usually give negative numbers as a result. One of the greater problems is the definition of benefits. A great part of the advantages of investing in technology cannot be quantified directly (positioning, small improvements in response time which are of greatest value for the client, etc.) or they may work not as a direct source of income but as a means to open up new possibilities (technological infrastructure improvement that enables taking on more important projects, integrated systems that enable to plan unified, better directed campaigns, etc)

As we see, there is still work to be done as regards the development of a comprehensive framework for the economic assessment of IT projects from the perspective of the value added to the business. This is clearly a multidisciplinary work. It cannot be performed only by technicians who do not know many business aspects, nor exclusively by businessmen, that may disregard the scope and impact of the technical characteristics when assessing which is the most suitable option within a certain context.

Quality Projects

In the last decades, software has become a vital part of most corporate products and services. As Kitchenham and Pfleeger (1996) indicate, such growth brings about the responsibility to establish the contribution of software to the organization. When a telephone company cannot implement a new service because the invoicing system cannot support it, software quality becomes a corporate issue. And these examples abound.

Nevertheless, very little has been made to investigate the relationship between software quality and business efficiency and effectiveness. One of the greater problems is to define the point where defect reduction or flexibility begin to threaten other desirable characteristics of software (time to market, performance, etc.) and cease to be a visible advantage (defects not noticed by users, products that are not going to be extended, etc.). Even in industrial sectors not related to software, returns on improvements in quality are decreasing (Slaughter, Harter, & Krishnan, 1998).

This type of analysis has not been seriously made in the software industry and, in fact, quality investment is usually a political or conjunctural decision (certifications, marketing, and differentiation) rather than economic. Organizations that decide to invest in quality on a systematic and constant basis are those that have experienced serious problems in the past. And, after some time, their expectations quite often decrease.

Some historical errors in software projects assessment go against the organization's goals. For instance, the cost of a software product has been traditionally considered (at least to decide upon its production or to choose among other alternatives) as the cost of its development up to its production, disregarding the statistics that indicate that maintenance (that is the tasks performed once the software is operative) uses 50% of the resources (cost) of a software project.

With a partial evaluation, no investment in quality will ever be justified, since the cost incurred until its operation (at least for the first time) will probably be greater in development than in activities of control and quality assurance. Nevertheless, this difference is compensated with the increase in maintenance-related savings (added to the improved image offered to the client) and to possible reductions of costs in future projects.

Let us remember that quality cost is the price paid for not doing things properly, and not the opposite. Therefore, a quality-oriented culture is going to focus on defining processes that come up rather than on fixing errors.

Quality cost (Slaughter, Harter & Krishnan, 1998) is usually divided into two main categories: cost of conformity and cost of nonconformity. The conformity cost is the amount allocated to obtain quality products. It is then divided into evaluation and prevention costs.

Error Prevention:

- Effort necessary to understand the root cause of the errors
- Process improvement activities
- Tools and quality training

Quality Assessment:

- Inspections, peer reviews and testing
- Product quality metrics

The cost of conformity does not include all the expenses incurred when things go bad — Internal fault of the product (one that takes place before delivering the product to the client):

- Reproduction and fault diagnosis
- Reprocessing
- Requirement changes and program redesigns

External product fault (one that takes place in the client's facility):

- Guarantee (adjustment and replacement)
- Client support for products with faults or usability problems
- Complexity programs or documents
- Legal cost
- Damage cost

When trying to reduce quality costs and maximize benefits, the key is to determine how much and when it is worth investing in specific initiatives of software quality improvement.

If evaluating the benefits of software projects is complex, the case of quality investment is even more so. In a linear way of thinking, a product of better quality is better.

What about quantification? In general, some of the benefits associated to quality investment are:

- Better image before the client due to a smaller amount of errors (nevertheless, it is difficult to know which of the errors preventively eliminated had indeed been detected by the client)
- New projects (extensions, interphases with other systems, migrations)
- Reduce time in future projects (by greater knowledge and systematization of the development processes)

As it may be observed, these benefits are basically intangible and uncertain, which makes it difficult (or at least quite arbitrary) to use traditional indicators to evaluate them. What is clear is that software quality is an investment that must provide some financial return after the expenses incurred.

Slaughter, Harter and Krishnan (1998) propose a series of indicators based on the most popular financial indicators to measure the return of investment of a quality initiative (specifically oriented to reduce the amount of defects in software), and to enable the comparison between different initiatives. They define measures for initial and recurrent costs and measure the benefits in terms of increase in sales (may be projected) or savings of costs considered. In order to apply their model, they compare different quality investment alternatives by means of an ex — post analysis of four initiatives carried out by a company.

Though interesting, this analysis has some limitations:

- It is made once the project is over and the initiatives implemented have been completed. They acknowledge this limitation and say that in order to use it a priori, there would have to be corporate information available as regards to reduction of defects in the products, statistics of defects, and so forth, as well as an estimate model to project improvements.
- When evaluating the obtained benefits, it is difficult to know what percentage can be attributed to quality initiatives.
- The analysis is limited to benefits that are perceived like directly bound to the improvement initiative.

The report shows that the return on quality investment is decreasing. This is particularly problematic when this model is used to compare initiatives that were already implemented sequentially (the effects of which are partially overlapped), since the last ones will be most affected in terms of their return.

Nevertheless, this work aims at understanding that quality-related actions, like investment, must be evaluated in terms of their costs and benefits (or contribution to the organization). The evaluation of data quality projects is much more precarious. There aren't any serious studies that may serve as a framework to analyze the convenience of investing in data quality improvement. Organizations come to action when they find they have very poor data (lawsuits filed by clients, returned posts, networks that do not match reality, etc.) and try to correct such data because they have no other choice (or they stop using them if they are irreparable).

The analysis is ad-hoc and, generally speaking, it aims at assessing the initiative's cost alone, thus submitting the decision to the resulting amount (high or low). Not only does this philosophy hinder the possibility of associating data improvement to the accomplishment of other projects, but it also goes against preventive investment in data quality — it only deals with incorrect data, without trying to identify the root cause of the problems.

Loshin (2001) shows a series of tactical missions of bad data quality and the cost incurred by the organizations. The author proposes the implementation of a Data Quality Scorecard, by means of which it is possible to assess the data's current condition and identify the greater causes of problems as potential improvement areas. Nevertheless, this approach is entirely focused on costs related to poor quality, without providing a framework to analyze if the cost of the improvement is justified (considering that the benefits are uncertain). It only aims at trying to achieve greater benefit with less investment, without describing how this may be done.

The ROI of data quality (2002) shows six cases in which corrective improvement measures are applied to data and tries to calculate the return on investment of these initiatives. Nevertheless, it does not provide a framework to evaluate data quality projects a priori and, within this context of scope uncertainty, it is not possible to define an improvement initiative (this evaluation can only be done a posteriori). In addition, it only contemplates direct benefits (that could not justify the investment) and it aims at corrective and no-preventive improvement.

REAL OPTIONS

Introduction

What is an Option?

An option is the right, but not the obligation, to take an action in the future. Options are valuable when there is uncertainty. For example, one option contract traded on the financial exchanges gives the buyer the opportunity to buy a stock at a specified price on a specific data and will be exercised (used) only if the price of the stock exceeds the specified price. Many strategic investments create subsequent opportunities that may be taken, and so the investment opportunity can be viewed as a stream of cash flow plus a set of options. (Brealey & Myers, 2000)

What are Real Options?

In a narrow sense, the real options approach is the extension of the financial option theory to options on real (nonfinancial) assets. While financial options are detailed in the contract, real options embedded in strategic investments must be identified and specified. Moving from financial option to real options requires a way of thinking, one that brings the discipline of the financial markets to internal strategic investment decisions.
The real options approach works because it helps managers, with the opportunities they have to plan and manage strategic investments. (Brealey & Myers, 2000)

A real option is an option "related to things" (Brach, 2002). Strategic investment and budget decisions are decisions to acquire, exercise, abandon, or let a real option expire. Managerial decisions create put and call options over real assets, which give management the right, but not the obligation, to use those assets in order to achieve strategic goals, and, consequently, maximize the organization's value.

An investment decision is rarely a "now or never" decision, nor a decision that cannot be abandoned or modified during the project. The decision can at least be delayed or accelerated, and it is often organized as a sequence of steps encompassing decision points. All these choices have an impact on the investment value, and constitute real options.

Stewart Myers coined the term "real option" when developing the idea that financial investments generate real options (Brach, 2002). Myers claimed that the valuation of investment opportunities using the DCF traditional approach disregarded the value of options arising in risky and uncertain projects. This idea was later expanded to any kind of investment decision and corporate budgeting.

According to Amran and Kulatilaka (1999), the real options model is a line of thought which comprises three main components rather useful to managers:

- Options are contingent decisions: an option is the opportunity to make a decision once an individual sees how events are taking place.
- Options valuation is aligned with financial markets valuation: the approach uses data and concepts from financial markets to assess complex payments in various kinds of real assets.
- Thinking about options may serve to design and manage strategic investments proactively.

As a matter of fact, the real options approach cannot be used for all investment decisions (Amran & Kulatilaka, 1999). In some cases, the investment is clearly good or bad, and an analysis based on real options would not change that decision. However, many of such decisions fall into a grey area that requires thorough assessment. A real option analysis becomes necessary when:

- There is a contingent investment decision
- Uncertainty is such that it is convenient to wait and gather more information
- Value seems to lie on future growth possibilities rather than on direct cash-flows
- Uncertainty is such that flexibility becomes important
- Updates will take place during its development

The six basic real options which derive from managerial options are (Brach, 2002):

- **Defer:** wait until further information reduces market uncertainty
- **Abandon:** dispose of an unprofitable project
- **Switch:** change input / output parameters or modus operandi
- **Expand / Contract:** alter capacity depending on market conditions
- **Grow:** consider future-related opportunities
- **Stage:** break up investments into incremental, conditional steps

Real Options Valuation

In the beginning, real options valuation models assumed that costs were deterministic, while in practice, costs, as benefits, tend to be uncertain. The time necessary to complete the project is usually uncertain as well. These features are characteristics of the real options model, the valuation of which must incorporate static product life cycles and variable cost structures.

The application of the real options valuation has expanded to appraise intangible assets investments, such as acquisition of knowledge or information and intellectual property, which are usually called virtual options.

Brach (2002) presents an analogy between financial options and real options concepts (see Table 1).

Let's see the assumptions of the Black & Scholes model that are not met in the case of real options:

- The project's volatility is not constant throughout time
- There is no final expiration date for the option
- Both the underlying asset value and the exercise price (i.e., the project's development cost) behave stochastically
- Payoffs are not normally distributed
- Real assets do not follow a "random walk"

Table. 1

Financial Option	**Variable**	**Real Option**
Exercise price	K	Cost to acquire the asset
Stock price	S	PV of future cash flows from the asset
Time to expiration	t	Length of time option is viable
Variance of stock returns	Var	Riskiness of the asset, variance of the best- and worst-case scenario
Risk-free rate of return	r	Risk-free rate of return

Given its discrete nature, in the case of the binomial model, the evolution of parameters can be monitored on a step-by-step basis while "unforeseen" changes can be observed. However, this also hinders the application of the model since, given the stochastic nature of many of its parameters in the case of the real options, it may be necessary to analyze several periods, making the construction of the binomial tree difficult.

Real Options in IT Projects

The application of real options to the information technology field has increased in recent years, the main reasons being twofold:

- A renewed need to justify the convenience of investing in IT.
- The boundaries of traditional project assessment techniques to model an IT investment properly.

Sullivan, Chalasani, Jha, and Sazawal (1999) present an approach to assess design decisions within the context of software development. Authors suggest assessing these principles integrally within the framework of an options analysis, to appreciate their contribution to the project's value. Benaroch (2002) suggests managing IT investment risk within the framework of real options. The author, together with Kauffman (Benaroch & Kauffman, 1999), presents the application of the real options valuation method on an IT investment project: the analysis of the right time to install POS (point of sale) debit services in the Yankee 24 banking network in New England. Schwartz and Zozaya-Gorostiza (2003) propose two models to assess IT projects based on whether they imply infrastructure acquisition or development. The authors also suggest a homogeneous framework to incorporate both types of projects.

Staged Options

Many projects are divided into a number of sequential stages where each step is based on the successful completion of the previous step and the management's possibility to assess the project in each stage (Brach, 2002). The benefit of a staged option will only be appreciated once all its stages have been fulfilled. Some examples are: investments in new technologies and in R&D. Both share the uncertain nature of staged investments, as long as the two main risk sources: private or technical risk (ability of the firm to effectively develop a successful project), and market risk (uncertainty of future demand). Key features of flexibility in a sequential project include the possibility to abandon it once more information is available at each step.

The assessment of this kind of option depends on the knowledge of the costs and benefits' stochastic processes:

- If they are known (or assumed), known closed-form valuation methods may be used.
- If they are unknown, the binomial model offers a viable assessment option. This means:

  - Defining milestones and decision-making points
  - Identifying sources of uncertainty
  - Estimating the costs of each stage
  - Estimating the time needed for the completion of each stage
  - Estimating the success probability of each stage
  - Estimating the best and worst net present value (NPV) case scenarios for the related project.

ASSESSMENT OF DATA QUALITY PROJECTS USING REAL OPTIONS

We present a methodological framework to apply the real options approach to the assessment of data quality projects. We will use the NEAT methodology to illustrate this approach, identifying its various stages.

There are different types of assessments that can be performed on a project of these characteristics:

- To assess the convenience of a specific investment.
- To assess the maximum convenient investment within certain framework.
- To analyze different scenarios to establish which one justifies an investment.

To apply the real options model, it is essential to identify the sources of uncertainty typical of this type of projects:

- State of data — theoretical scope of the improvement
- Cost of the improvement
- Real improvement
- Contingent projects: some projects may depend on the actual quality achieved and therefore, may not be convenient if the improvement does not reach certain levels
- Potential benefits from projects to be implemented

Although the numerical analysis is vital to draw a conclusion, we do not disregard the fact that many of the values used are predictions (with a higher or lower degree of certainty), estimates and even desires. Hence, we will focus on establishing a homogeneous comparison framework of investment alternatives based on the use of real options.

Why Use Real Options?

As mentioned in previous sections, it is difficult to assess the convenience of a quality investment project (whether of software or data) just by looking at the cash-flow directly associated to it. In data quality projects we can highlight the following features:

- They help the organization be ready to perform actions they would otherwise be unable to.
- A great part of the benefits is qualitative and difficult to quantify (what is the value of a robust application?).
- Benefits are not always direct; quality improvement projects establish the starting point to many different initiatives that may or may be not carried out, and they open up new opportunities to organizations that were not even considered before. For instance, a flexible software design allows considering software evolution, which would otherwise be costly or even impossible.
- Although they open up new opportunities, the benefits are still uncertain (quality may not reach the expected level, timing may be wrong, etc.).
- There are some economic issues to consider when investing in quality. Not all investments are cost-effective.

In addition, traditional techniques (Discounted Cash Flow, Net Present Value, Return of Investment, and Internal Rate of Return) are limited:

- They do not capture the possibility to change the investment sequence.
- They do not consider the option to abandon a project.
- They deal with deterministic and known costs (which is not always the case).
- The discount rate and future cash flows may be arbitrarily determined, thus affecting the accuracy of the computation.

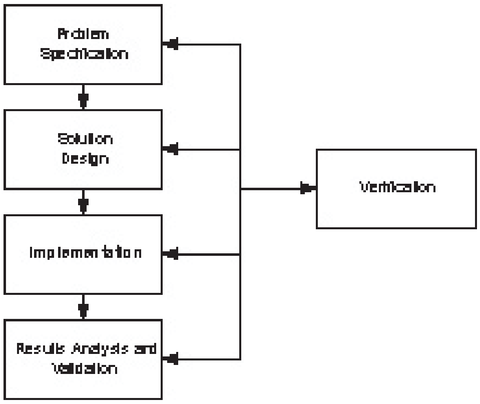
These factors make the real options approach seem a natural alternative for these projects.

Methodological Framework

We propose the *life cycle of a DQ project assessment*. Figure 2 shows a graphic representation where:

- **Problem specification:** Informal description of the project; evaluation goals; sources of uncertainty; sources of benefits and option type (default: staged option).

Figure 2. Life cycle



- **Solution design:** Real option/s and its components, valuation model and key parameters.
- **Implementation:** Assigning values to parameters and computing option price.
- **Results analysis and validation:** Comparison with former projects, sensitivity analysis and results recording.
- **Verification:** tasks are performed during the whole process in order to assess consistency among stages.

This methodology is absolutely simple and general. However, we have listed some basic considerations to make its application successful:

- As in every IT project, the key issue is to understand all requirements properly. In this case, this means understanding the project to be assessed, which components follow stochastic processes and which are deterministic, and identifying all the possible sources of benefits.
- Future benefits are associated to projects, the costs of which are independent from data improvement costs.
- The solution design must contemplate the possibility of assessing more than one option, due to different future projects, diverse investment alternatives, different stages, and so forth.
- The choice of the assessment model is important to ensure the accuracy of the analysis. Nevertheless, given a highly uncertain context with little experience in the use of the technique, it is convenient to promote less complex models (considering their limitations) and, based on the experience, search for models that can better adjust to the problem that is being addressed.

- Analyzing the results is vital, not only for the project under assessment but also for the lessons learned which will enable to improve the application of the model in the future.
- It is convenient to repeat the analysis as you make progress on the project and gain more certainty so as to improve the decision-making process.

EVALUATING DQ PROJECTS USING REAL OPTIONS

To avoid presenting the proposed methodology in rather abstract manner, considering that we are presenting it for a specific type of project (NEAT projects), in this section we will describe how a real option with these characteristics should be produced. However, this does not mean that under different circumstances another type of option could be more convenient. This section may be considered *a partial instance* of stages 1 and 2 of the methodology presented.

Some considerations:

- For the sake of simplicity, this last stage of the methodology, monitoring, is not included in the scope of the projects to be assessed.
- The benefits obtained from this type of project are associated to future projects, which have their own cost.
- An improvement plan may consist of several tasks that are optional to the organization, each of them with its corresponding cost and improvement expectation.
- In the case of poor quality data, the company may be losing money for that reason (as we will show in the case study). In this case, the benefit of the data improvement will be to prevent such loss and therefore it will be direct and positive.

At this point, we will present a two-stage analysis by means of which it is possible to make a preliminary assessment so as to determine if it is worth implementing a data quality improvement project within the NEAT framework, and then, a more detailed analysis that covers specific tasks to be carried out. The application of the model adds a few requirements to the methodology which will be explained later. In the next section, a complete case study is presented.

First Stage

Before elaborating the diagnosis, it may be necessary to consider under what circumstances diagnosis may be justified. That is to say, assessing different possibilities with various probabilities, even without knowing the improvement cost, as wells as considering the success probability of the improvement and considering different contingent projects. In this case, and to make matters simple, the binomial valuation model is recommended as well as creating an option for the cost of each improvement plan.

Questions that should be addressed during this stage:

- How much is it worth to invest in data quality?
- Is it worth to proceed with the diagnosis phase?

Solution design consists partially of identifying the set of questions to be answered. In fact, design is driven by these questions. We need to establish (or at least estimate):

- Cost, duration and success probability of the improvement: recorded information regarding previous experiences can be used.
- Cost, benefits, duration and success probability of the different projects: future project cash flow pessimistic and optimistic estimates
- Project volatility: inferred from the scenarios presented, based on the success probability of the project (if project expected value (E(S)) is $E(S) = S_{max} * q_{max} + S_{min} * q_{min}$ then $Var(S) = (S_{max} - E(S))^2 * q_{max} + (S_{min} - (E(S))^2 * q_{min}$ and $\sigma = Var(S)^{1/2}$) (see Table 2)

- Diagnosis cost and duration
- Risk-free rate of return (company data)
- WACC or rate of return uses for NPV computations

Additional requirements imposed over the NEAT methodology:

- If this stage takes place before the diagnosis phase, there are no additional requirements

Table 2.

Variable	**Real Option**
K	Cost to acquire the asset
S	PV of future cash flows from the asset
t	Length of time option is viable
Var	Riskiness of the asset, variance of the best and worst case scenario
r	Risk-free rate of return
E	Expected Value
q	Probability
σ	Variance

- If it is performed again after the diagnosis, second stage requirements apply

In order to be worthwhile, the resulting price should be equal to or lower than the diagnosis cost. If more than one option is defined, a more detailed analysis is needed.

Second Stage

The second stage aims at conducting an in-depth analysis to determine the suitability of the improvement. The steps to follow are similar to those of the first stage, with the exception that the improvement has already been broken down and various tasks may be combined. It is necessary to determine:

- Different improvement configurations (if any)
- Cost and success probability of each improvement configuration
- Possible cost and success probability of the project/s and minimum quality requirements, to establish under which improvement configuration they are feasible.
- Benefits of the project/s

Questions to answer:

- Is the improvement process worthwhile?
- Which projects and improvement tasks should be performed?

Data required:

- Diagnosis cost and duration
- Improvement plan: including, for each task
  - Activity
  - Cost
  - Success probability
  - Duration
  - Relation to other tasks

- Future projects:
  - Project
  - Optimistic cost and benefits (Net Present Value, expected cash flow, etc.)
  - Pessimistic cost and benefits (Net Present Value, expected cash flow, etc.)
  - Pessimistic scenario probability
  - Optimistic scenario probability

- ◦ Duration
- ◦ Previous activities required
- ◦ Volatility

- Different combinations of future projects (to construct different scenarios)
- Risk-free rate of return
- Weighted-average cost of capital (WACC) or discount rate used for net present value (NPV) calculations

Additional requirements over the NEAT methodology:

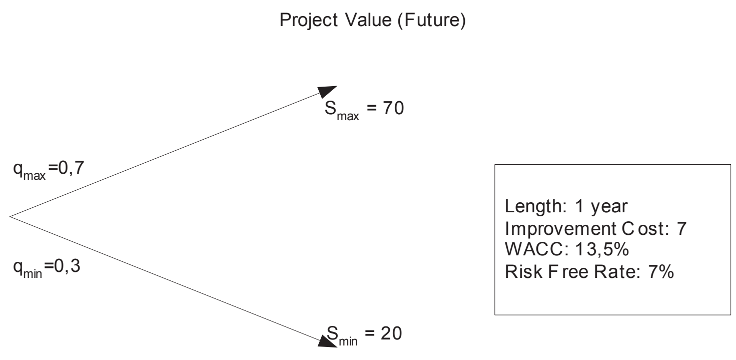
- Success probability of the improvement plan
- Success probability of the improvement tasks
- Evaluation criteria (to assess whether or not activity goals have been accomplished)
- Quality requirements for each contingent project
- Dependencies among tasks
- Worst- and best-case scenario for each future project; probability of each scenario

Option Valuation Model

For this assessment, we also suggest using the binomial model as proposed by Brach (2002). It is worth pointing out that, though simple, this approach poses an essential constraint: it does not contemplate that the success of an improvement action (or of the plan as a whole) may be partial. However, it is a useful approach, a trade-off between simplicity and clarity. The valuation model should be further improved in the future.

We present a simple application of the binomial valuation model.

Figure 3. First stage analysis



Take the scenario in Figure 3 for a first stage analysis. We can compute (Brach, 2002):

Project expected value:

$V = (q_{max} * S_{max} + q_{min} * S_{min})$
$V = (0.7*70+0.3*20)=55$
$\sigma= 40.31$

Risk-free probability (Brach, 2002) uses this formula to compute p because it is less complex than the one used to valuate financial options and comprises all the properties p should have (i.e., be a probability and represent the value q would have in equilibrium in a risk-free scenario. That is why we are using it here, too).

$p = ((1+r_{free} *V)- S_{min}) / (S_{max} - S_{min})$
$p = ((1.07*55)-20) / (70-20) = 0.777$

Option value:

$C = ((p * S_{max} + (1-p)* S_{min})/ (1+r_{free})^t) -K*(1+r_{wacc})^t)= 47.055$

Where K is the improvement cost, t is the diagnosis duration + improvement
$C = ((0.777*70 + (1-0.777)*20)/1.07) -7*1.135= 47.055$

The option value is calculated after the improvement has been performed, but before the so-called future project is executed. That is why the future project's NPV is subtracted and the improvement cost is to be considered in the future (we are assuming it to be one year).

If we plan to spend 7 in improvement tasks, and the diagnosis cost is below 47.055/1.135 (and it requires a very short time), the investment is worthwhile (notice that data here is only presented to exemplify the valuation method). If we want to compute the critical investment level K, we should solve the last equation making it 0 and leaving the improvement cost variable. The result would include diagnosis plus improvement costs.

Assuming that the diagnosis and improvement last a year, at most:

$C = ((0.777*70 + (1-0.777)*20)/1.07) - K*1.135= 0$

Or

$(((0.777*70 + (1-0.777)*20)/1.07)) /1.135 = K$

so K = 48.458

So diagnosis and improvement costs together have to be at most 48.458.

Knowing the diagnosis cost in advance, this analysis becomes a crucial input for stage 2, because it helps to determine whether to implement the improvement plan completely or partially, even before a detailed analysis is made.

CASE STUDY

In this section, we will apply the methodology to evaluate a real DQ project, as described in earlier sections. At present, the project is in the "lessons learned" phase, so the analysis allows the upper management to assess its progress.

Actual names and figures were changed in order to preserve confidentiality. These changes do not affect the results of the analysis.

Problem Specification

International Petroleum is implementing an information quality assurance program on its geology and geophysics (G&G) information. This program relates with a parallel project that aims at classifying physical data.

The aims of the project are: perform an information diagnosis; ensure that the classification of information meets minimum quality standards; ensure that the tools that are being used are adequate; standardize the input data of all applications; and implement a data quality improvement process.

Solution Design

To evaluate the case, we will apply both stages proposed. In this section we will design the components to be included in each of the evaluations.

The variables risk-free ratio (7%), WACC — capital cost — (13.5%), diagnosis' cost ($0.1) and diagnosis' length (2 months) are required for both stages (money figures in MM US dollars).

The company only identified the physical data classification project. This is a high yield project. However, the project success depends on the availability of high quality information.

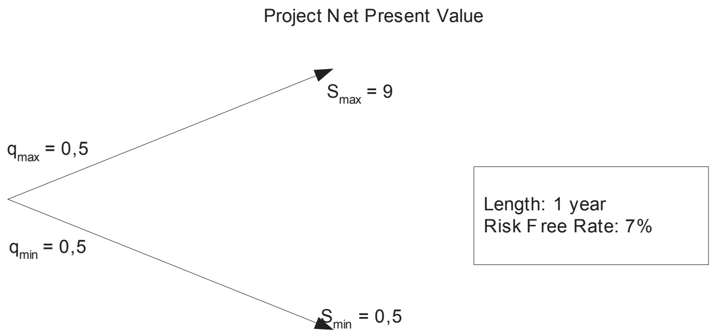
The general **Project** variables are the *diagnosis' cost* ($0.1), *length* (5 years), *scenario probability* (50%) and *total cost* ($4.21).

For the **Project Yield,** the optimistic variables were defined as *NPV* -$9 and *revenue* — $15; and the pessimistic scenario were defined as *NPV* — 0.5% and *Revenue* — $5.35.

First Stage

In this stage of the analysis we need to know the kind of data quality investment the company is planning to make. This level of investment will enable to assess

Figure 3. Project feasibility



the project's feasibility, even if the total cost of the data improvement program still remains unknown.

The variables of the **Improvement Project** are *length* (diagnosis + improvement): 1 year and *success probability*: 60%.

Second Stage

This analysis is done after the diagnosis is completed and the improvement plan tasks are identified and estimated.

This first analysis will be similar to the previous one, with the addition of the improvement cost and success probability. A subsequent analysis will identify sequential tasks.

The required data for the First Stage are *length, cost and success probability*; and for the Second Stage are *improvement plan (activity-based)* and *dependencies among tasks*.

Diagnosis Result

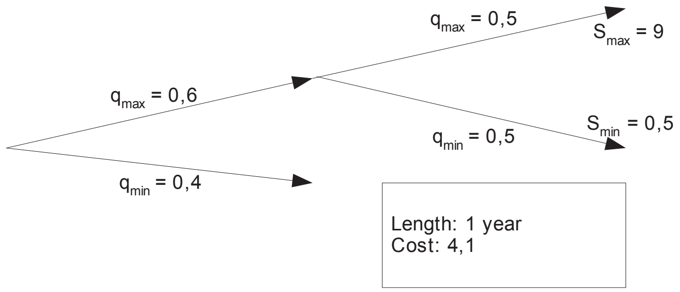
The following issues were detected during the analysis: problems in the definition of roles and responsibilities; lack of definitions for data loading criteria; and Data inconsistencies among different applications.

Improvement Plan

The *cost* of improvement plan for data loading criteria definition and implementation is *$1.2*; the *success probability 70%* and *length 4 months*.

The *cost* of the data-cleansing activities is *$1.5*, the *success probability*: 50% and the *length: 3 months*.

Figure 4. Overall improvement



Global Improvement Plan

The *cost* of the Global Improvement Plan is *$4*, the *success probability*: 60% and the *length: 10 months*. (Success of the improvement plan depends on the success of the data cleansing process, as well as on the success of the remaining tasks.) The diagram in Figure 4 shows the design of the first stage.

This graph represents the overall improvement project: The first node shows that the improvement will have a 60% success probability. If the improvement fails, the overall project will be abandoned. If it is successful, then the physical data classification project will be implemented, with a 50% success probability.

Implementation

To calculate the value of the options a binomial model was used. All the necessary functions were defined in a Microsoft Excel spreadsheet.

First Stage

Project NPV

$V = (q_{max}*S_{max}+q_{min}*S_{min})$
$V = (0.5*9+0.5*0.5)=4.75$
$\sigma= 4.25$

Risk-free probability:

$p = ((1+r_{free}*V)- S_{min}) / (S_{max} - S_{min})$
$p = ((1.07*4.75)-0.5) / (9-0.5) = 0.539$

Critical investment level:

$$K = ((p * S_{max} + (1-p)* S_{min})/ (1+r_{free})^{t}) -K*(1+r_{wacc})^{t})= 0$$
$$K = ((0.539*9 + (1-0.539)*0.5)/1.07) / 1.135= 4.18$$

Second Stage

We will use these formulae to assess the staged options.

- **First analysis:** The previous node evaluation is completed in the same way. First we calculate the node NPV assuming that the immediate node was successful, subtracting the maximum value for the WACC for the number of periods (15 in this case). The minimum value is always 0, because if the task fails the project is abandoned). The expected value is calculated using the success probability. To see details of this case, refer to Table 3.

Max Value:

$$S_{max} = S_{max'} /(1+r_{wacc})^{t}$$
Where $S_{max'}$ is the maximum value of the immediate node
$$S_{max} = 15/1.135=13.2$$

Project NPV:

$$V = (q_{max} *S_{max} +q_{min} *S_{min})$$
$$V = (0.6*15+0.4*0)=5.4$$

Risk-free probability:

$$p = ((1+r_{free} *V)- S_{min}) / (S_{max} - S_{min})$$
$$p = (1.07*5.4) / (13.2) = 0.435$$

Table 3.

Stage	**NPV**	**q (success)**	**length**	**Cost**	**Option Value**	**P**
Project	15 (max) 5.35 (min) 10.175 (exp)	0.5	1 year	4.21	5.96	0.573
Improvement + diagnosis	13.2 (max) 0 (min) 5.4 (exp)	0.6	1 year	4.1	1.27	0.435

Table 4.

Stage	NPV	q (success)	length	Cost	Option Value	P
Project	15 (max) 5.35 (min) 10.175 (exp)	0.5	1 year	4.21	5.965	0.573
Data cleansing	13.2 (max) 0 (min) 4.48 (exp)	0.5	3 months	1.5	3.225	0.363
Dataflow, roles & responsibilities	12.78 (max) 0 (min) 2.76 (exp)	0.7	4 months	1.3	1.6	0.231
Data loading criteria	12.22 (max) 0 (min) 1.95(exp)	0.8	3 months	1.2	0.734	0.171

Option Value:

$$C = ((p * S_{max} + (1-p) * S_{min}) / (1+r_{free})^t) - K$$
$$C = ((0.435*13.2 / 1.07) - 4.1 = 1.27$$

- **Second analysis:** Refer to Table 4.

Analysis

After the first stage we can conclude that diagnosis is worthwhile and, if improvement costs are below 4, it is convenient to proceed.

The second stage gives us a more detailed view. We observe that it is still convenient to perform the improvement, but also we can see how the option value increases as long as intermediate milestones are met and uncertainty is reduced.

A sensibility analysis shows that if the estimated probabilities are modified, the results can be substantially affected. Having in mind the fact that those values are estimations, a close tracking of the tasks involved may help improving the model calibration.

In this particular example we haven't obtained negative values. If that was the case, a tasks subset could have been chosen, in order to maximize the option value but, at the same time, achieving project quality requirements.

Our model also facilitates the process of staging the improvement project, defining checkpoints to decide whether to abandon the project or not, thus reducing any potential loss.

Conducting a NPV analysis (setting the expected DCF):

NPV = -4.1 + (4.75/ 1.135) = 0.08

Although the NPV is positive (which is not always the case), based on this number alone we would not be able to decide to abandon the project if one of the stages fails. In addition, since the value is relatively low, we could be tempted to terminate the project. We see that the NPV analysis may force us to abandon valuable projects, but also to proceed with a project without reviewing the decision. Finally, the option value is significantly higher than the NPV, because the option analysis captures the "time flow" that reduces uncertainty and helps to decide whether to abandon the project or not.

Conclusion

After developing the case study we may conclude:

- The proposed analysis gives us a global view of the overall project and how much we would be willing to invest in it, considering future expectations.
- The binomial valuation model facilitates the definition of simple spreadsheets to perform the calculations, even when specific software may not be available. However, tracking the valuation is a complex process.
- The proposed model facilitates the definition of go / no go stages, which add flexibility to the management decision process, enabling to take risks and delay decisions in a controlled framework.
- It would be valuable to enrich the model by adding sensitivity analysis, scenario analysis, project combinations, investment portfolios, etc.
- The proposed model shows some advantages over traditional indicators, showing some limitations of the classical view. However, further research on this field is required.
- The chosen valuation method considers neither partial success, nor the possibility to paralyze stages. The steps are sequential and the results are either success or failure.
- More experimentation with complex cases is needed.
- Comparison with NPV: the results obtained with the most traditional analysis were compared using NPV, thus giving clear proof of their own limitations.
- Support Tools: although the model can be applied, the use of support tools is critical to extend the scope of the evaluation.
- Complexity of the case: Since a simple case was used, it is essential to experiment with situations of greater complexity.
- Time management: for short-duration tasks, time management makes the use of discount rates more complex.

Detailed Calculations

Tables 5 through 8 show the spreadsheets used for these calculations. The nodes are numbered in their evaluation order (in reverse chronological order).

Table 5.

First node		
Concept	**Value**	**Formula**
qmax	0,5	Data
qmin	0,5	Data
Smax	15	Data
Smin	5,35	Data
rfree	0,07	Data
rwacc	0,135	Data
Length (t)	1	Data
Stage Cost (K)	4,21	Data
V expected	**10,175**	qmax*Smax+qmin*Smin
p free	**0,574**	((1+rfree *Vesp)- Smin) / (Smax- Smin)
Vop	**5,965**	((p * Smax + (1-p)* Smin)/ (1+rfree)t) –K

Table 6.

Second Node		
Concept	**Value**	**Formula**
qmax	0,5	Data
qmin	0,5	Data
Smax	**13,216**	Smax = 15 /(1+rwacc)
Smin	0	Data
rfree (monthly)	0,005	Data
rwacc	0,135	Data
Length (t)	3	Data (months)
Stage Cost (K)	1,5	Data
V expected	**4,482**	V esp= (qmax*Smax+qmin*Smin)
p free	**0,363**	p = (1+rfree*Vesp) / Smax
Vop	**3,225**	(p * Smax / (1+rfree monthly)t) –K
rwacc (monthly)	0,011	Data
rfree	0,07	Data

Table 7.

Third Node		
Concept	**Value**	**Formula**
qmax	0,7	Data
qmin	0,3	Data
Smax	**12,780**	13.216 /(1+rwacc monthly)3
Smin	0	Data
rfree (monthly)	0,005	Data
rwacc	0,135	Data
Length (t)	4	Data(monthly)
Stage Cost (K)	1,3	Data
V expected	**2,764**	Vesp = qmax*Smax+qmin*Smin
p free	**0,231**	(1+rfree*Vesp) / Smax
Vop	**1,600**	(p * Smax/ (1+rfree monthly)t) –K
rwacc (monthly)	0,011	Data
rfree	0,07	Data

Table 8.

Fourth Node		
Concept	**Value**	**Formula**
qmax	0,8	Data
qmin	0,2	Data
Smax	**12,220**	12.780 /(1+rwacc monthly)4
Smin	0	Data
rfree(monthly)	0,005	Data
rwacc	0,135	Data
Length (t)	5	Data (monthly)
Stage Cost (K)	1,3	Data
V expected	**1,949**	Vesp = qmax*Smax+qmin*Smin
p free	**0,171**	(1+rfree*Vesp) / Smax
Vop	**0,734**	(p * Smax/ (1+rfree monthly)t) –K
rwacc (monthly)	0,011	Data
rfree	0,07	Data

FUTURE TRENDS

We believe the proposed model is not only useful to evaluate data quality projects, but to help organizations to analyze and optimize systematically quality investments in general. In order to achieve that goal several issues have to be cover:

- Validate the suitability of the model by applying it to more case studies in different industries
- Extend the proposal to other quality investment projects (i.e., apply this methodology to other fields, not only data quality). It will only be necessary to review how to define some stages, since the methodological frame is the same.
- Develop tools to automate the estimation process and the recording of results.
- Consider the analysis of the combination of future projects, avoiding the simplifications presented in this work.
- Study how to incorporate intangible benefits to the analysis in systematic manner, considering the frequency of its appearance in this type of projects.
- Systematize the scenario analysis and facilitate comparisons.
- Consider other valuation models
- Extend the model to consider different quality investment combinations (software + data).
- Incorporate additional requirements to the NEAT methodology.

CONCLUSION

This chapter was aimed at defining a methodological framework to assess the benefit of a data quality improvement project using real options and to validate the proposal with a case study. We have drawn the following conclusions:

- Despite the fact that information is highly important, it is difficult for organizations to find an economic justification to invest in data improvement.
- Traditional techniques are very limited, since they do not consider the possibility of changes in investment sequences, they consider deterministic and already known costs, and may be somewhat arbitrary in the choice of a discount rate, in addition to the determination of future flows.
- Real options are a suitable approach for quality projects in general and data quality in particular, since they allow to consider uncertainties in terms of costs and benefits, flexibility to decide whether to continue with a project or not, different open opportunities, etc.
- We have proposed a methodological framework to use real options for the assessment of data quality projects. The framework is general and we have instantiated it for this type of project specifically. We have not made emphasis

on the valuation method, since we'd rather prioritize the underlying reasoning model.

- Applying this methodology to a concrete case enabled to prove its simplicity, its swift implementation and also some of its constraints (binary result of activities, arbitrary estimation of some probabilities, etc.).
- The methodology proposed does not solve the difficulties that arise when trying to quantify benefits and opportunities. Moreover, though possible, the analysis of different scenarios may be highly complex and bothersome.

ACKNOWLEDGMENTS

We'd like to thank Daniel Yankelevich and Martin Patrici for their helpful comments.

REFERENCES

Amran, M., & Kulatilaka, N. (1999). *Real options: Managing strategic investment in an uncertain world*. Boston, MA: Harvard Business School Press.

Benaroch, M. (2002). Managing information technology investment risk: A real options perspective. *Journal of MIS*, *19*(2), 43-84.

Benaroch, M., & Kauffman, R. J. (1999). A case for using real options pricing analysis to evaluate information technology project investments. *Information Systems Research*, *10*(1), 70-87.

Bobrowski , M., Marre, M., & Yankelevich, D. (1999, November). *An homogeneous framework to measure data quality*. Paper presented at the Information Quality Conference, Boston, MA.

Bobrowski, M., Marre, M., & Yankelevich, D. (2001). A NEAT approach for data quality assessment. In M. G. Piattini, C. Calero, & M. Genero (Eds.), *Information & Database Quality* (pp. 135-162). Norwell, MA: Kluwer Academic Publishers.

Boehm, B., & Sullivan, K. (2000, June). *Software economics, a roadmap*. Paper presented at the 22nd International Conference on Software Engineering, Limerick, Ireland.

Brach, M. (2002). *Real options in practice*. Hoboken, NJ: Wiley & Sons.

Brealey, R. A., & Myers, S. C. (2000). *Principles of corporate finance* (6th ed.). Boston, MA: Irwin McGraw-Hill.

COCOMO (2005). Retrieved January 20, 2005, from http://sunset.usc.edu/research/COCOMOII/

Fenton, N., & Pfleeger, S. (1997). *Software metrics, a rigorous and practical approach* (2nd ed.). London: PWS Publishing.

GQM. (2005). Retrieved February 10, 2005, from http://sel.gsfc.nasa.gov/website/exp-factory/gqm.htm

Huang, K., Lee, Y., & Wang, R. (1999). *Quality information and knowledge*. Englewood Cliffs, NJ: Prentice Hall.
Kitchenham, B., & Pfleeger, S. L. (1996). Software quality — The elusive target. *IEEE Computer*, *13*(1), 12-21.
Loshin, D. (2001, June). *The cost of poor data quality*. In DM direct Newsletter. Retrieved November 21, 2001, from http://www.dmreview.com
McKnight, W. (2003). *Overall approach to data quality ROI* (white paper). Retrieved October 20, 2003, from http://www.firstlogic.com/papers
PI. (2005). Retrieved January 18, 2005, from http://www.investopedia.com
Redman, T. D. (2001). *Data quality: The field guide*. Boston, MA: Butterworth-Heinemann.
ROI. (2005). Retrieved February 3, 2005, from http://www.mrlease.com/leasedict.htm
Schwartz, E. S., & Zozaya-Gorostiza, C. (2003). Investment under uncertainty in information technology: Acquisition and development projects. *Management Science*, *49*(1), 57-70.
Slaughter, S. A., Harter, D., & Krishnan, M. S. (1998): Evaluating the cost of software quality. *Communications of the ACM*, *41*(8), 67-73.
Sullivan, K. J., Chalasani, P., Jha, S., & Sazawal, V. (1999). Software design as an investment activity: A real options perspective. In L. Trigerorgis (Ed.), *Real options and business strategy: Applications to decision making* (pp. 215-261). London: Risk Books.
The ROI of Data Quality. (2002). Retrieved October 15, 2002, from http:// www.trilliumsoftware.com
WACC. (2005). Retrieved January 15, 2005, from http://www.invesco.com.au

GLOSSARY

GQM: "The Goal-Question-Metric (GQM) method is used to define measurement on the software project, process, and product in such a way that:

- Resulting metrics are tailored to the organization and its goal.
- Resulting measurement data play a constructive and instructive role in the organization.
- Metrics and their interpretation reflect the values and the viewpoints of the different groups affected (e.g., developers, users, operators).

GQM defines a measurement model on three levels:

- **conceptual level (goal):** A goal is defined for an object, for a variety of reasons, with respect to various models of quality, from various points of view, and relative to a particular environment.

- Operational level (question): A set of questions is used to define models of the object of study and then focuses on that object to characterize the assessment or achievement of a specific goal.
- Quantitative level (metric): A set of metrics, based on the models, is associated with every question in order to answer it in a measurable way."

(GQM, 2005).

weighted average cost of capital (WACC): "Expected return on a portfolio of all the firm's securities. Used as hurdle rate for capital investment" (WACC, 2005).

net present value (NPV): "The current value of a stream of income and principal (if any) discounted by an interest rate over the period of an investment." (WACC, 2005).

internal rate of return (IRR): "The rate of interest that needs to be applied to make the net present value of an investment equal to the price paid." (WACC, 2005).

return on investment (ROI): "The yield. The interest rate earned by the lessor in a lease, which is measured by the rated at which the excess cash flows permit recovery of investments. The rate at which the cash flows not needed for debt service or payment of taxes amortize the investment of the equity participation." (ROI, 2005).

profitability index (PI): "An index that attempts to identify the relationship between the costs and benefits of a proposed project through the use of a ratio calculated as:

$$= \frac{\text{PV of Future Cash Flows}}{\text{Initial Investment}}$$ " (PI, 2005).

yield: "(a) The return on an investment expressed as a percentage; (b) the profit or income that an investment or property will return; (c) the money derived from any given business venture, usually expressed as an annual percentage of the initial investment. Straight yield (or running yield) relates cash flow to price paid and does not take into account any gain or loss of principal. Redemption yield relates to the sum of both cash flow, for example interest payments, over the life of the security and any gain or loss at maturity on the initial amount invested." (WACC, 2005).

risk-free rate of return: "A theoretical return that is earned with perfect certainty; it is without risk. In Australia, the risk-free return is generally the government bond rate." (WACC, 2005).

expected value: "A statistical term denoting a predicted value of a variable in the future." (WACC, 2005).

expected rate of return: "The weighted arithmetic average of all possible returns on an asset or portfolio, where the weights represent the probabilities that the outcomes will occur. It is the expected value or mean of a probability distribution." (WACC, 2005).

COCOMO "is a model that allows one to estimate the cost, effort, and schedule when planning a new software development activity. It consists of three submodels, each one offering increased fidelity the further along one is in the project planning and design process. Listed in increasing fidelity, these submodels are called the Applications Composition, Early Design, and Post-architecture models (COCOMO, 2005).