



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A Risk Based Model For Quantifying The Impact Of Information Quality

Citation for published version:

Borek, A, Parlikad, AK, Woodall, P & Tomasella, M 2014, 'A Risk Based Model For Quantifying The Impact Of Information Quality', *Computers in Industry*, vol. 65, no. 2, pp. 354-366.
<https://doi.org/10.1016/j.compind.2013.12.004>

Digital Object Identifier (DOI):

[10.1016/j.compind.2013.12.004](https://doi.org/10.1016/j.compind.2013.12.004)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Computers in Industry

Publisher Rights Statement:

© Borek, A., Parlikad, A. K., Woodall, P., & Tomasella, M. (2014). A Risk Based Model For Quantifying The Impact Of Information Quality. *Computers in Industry*, 65(2), 354-366. 10.1016/j.compind.2013.12.004

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



A Risk Based Model for Quantifying the Impact of Information Quality

Alexander Borek, Ajith Kumar Parlikad, Philip Woodall, Maurizio Tomasella

Cambridge University Engineering Department

(This manuscript has been published in Computers in Industry - DOI:10.1016/j.compind.2013.12.004)

Abstract

Information quality is one of the key determinants of information system success. When information quality is poor, it can cause a variety of risks in an organization. To manage resources for information quality improvement effectively, it is necessary to understand where, how, and how much information quality impacts an organization's ability to successfully deliver its objectives. So far, existing models have mostly focused on the measurement of information quality but not on the impact that information quality causes. This paper presents a model to quantify the business impact that arises through poor information quality in an organization by using a risk based approach. It hence addresses the inherent uncertainty in the relationship between information quality and organizational impact. The model can help information managers to obtain quantitative figures which can be used to build reliable and convincing business cases for information quality improvement.

Keywords: Information Quality, Data Governance, Business Impact, Risk Based Modeling, Risk Management

1. Introduction

Information quality has been identified and confirmed as a key determinant for information system (IS) success (DeLone and McLean 1992; Delone and McLean 2003). Information quality is the fitness for use of information and is a multi-dimensional concept (Wang and Strong 1996, D. Ballou et al. 1998, English 1999) with dimensions such as, for example, accuracy, consistency, interpretability, timeliness and completeness. The terms “data quality” and “information quality” are usually used interchangeably. Information of poor quality can endanger the competitiveness and success of organizations. It can lead to poor decision making (Raghunathan 1999; Keller and Staelin 1987; O'Reilly III 1982; Jung et al. 2005; Shankaranarayanan and Cai 2006; Chengalur-Smith, Ballou, and Pazer 1999; Ge and Helfert 2008) and can in many ways create risks that hinder organizational performance (Redman 1998; Slone 2006; Eppler and Helfert 2004; Fisher and Kingma 2001).

Effective methods for assessing how poor information quality impacts the business are therefore crucial to enable an organization to focus information quality improvement, where information quality affects the business goals most severely. They are also needed to build a convincing business case for information quality improvement initiatives in organizations. However, current methods to assess the business impact of information quality are inadequate when it comes to thoroughly

quantifying the business impact of information quality. Some methods have been proposed by management consultants without a solid model as a basis (English 1999; McGilvray 2008; Loshin 2001; Loshin 2010). Models proposed in academia are either focusing on information quality processes and technical metrics without characterizing the impact of information quality, e.g. (D. P. Ballou and Pazer 1985; D. Ballou et al. 1998; Askira Gelman 2011; Parssian, Sarkar, and Jacob 2004), or use utility theory to characterize the impact of information quality (Even and Shankaranarayanan 2007; Ahituv 1980), which is difficult to apply in practice as utilities are difficult to link to key business performance indicators. In particular, the models do not consider the uncertainty that is inherent in the relationship between information quality and business outcomes.

Understanding and assessing the risks arising from information quality would allow organizations to focus on information quality improvement where it matters most and also help to justify the costs for information quality improvement. The technical challenge is that an adequate model has to be as easy to use as possible to be applicable by practitioners in an industrial context and it has to be feasible to collect the data input. At the same time, it needs to provide the necessary rigour to build a believable business case for information quality improvement. Furthermore, the business impact of information quality must relate to business metrics that are accepted and meaningful in the given industrial organization.

This paper presents a risk based approach to quantify the business impact of information quality that achieves the balance between usability and rigor. It has been extensively applied and refined in six industrial case studies and successfully tested for feasibility, usability and utility. The model enables managers to quantify how information quality affects their organizational performance and use the results to build solid business cases for information quality improvement.

The novelty of this research is to apply risk management concepts and methods to information quality management in order to quantitatively assess the business impact of information quality in a comprehensive manner. It adds the probabilistic dimension to the current state of the art of information quality business impact assessment methods. The model provides therefore an improved solution for industrialists to understand and assess the risks arising from information quality on an organization-wide scale. The results of these assessments can be useful to guide information quality management strategy based on a solid factual basis of the actual impact of information quality in the organization. The research shows a further way to utilise risk management effectively in the information systems discipline, in addition to the previous usages in the areas of information security and IT failure management.

2. Research Methodology

A design science approach as defined by (A. R. Hevner et al. 2004; March and Smith 1995) following a pragmatic philosophical paradigm is taken as philosophical underpinning and methodological approach in this research. Design research follows a cycle of develop/build and justify/evaluate of the research artifact. The artifact in form of the model was therefore designed and tested in four consecutive research phases, as shown in Figure 1. The model explicitly considers the uncertainty in the impact. The model presented in this paper is an instance of a cause-to-effect operational risk quantification model, which can “preserve the cause-to-effect relationship that shows how operational risk can be reduced, managed, and controlled” (Supatgiat, Kenyon, and Heusler 2006, 16). In phase 1, a review of relevant literature in information quality and risk management and 15 semi-structured interviews with management professionals were conducted to understand the current needs and practices in the industry. This helped to prepare the six in-depth case studies (summarized in Table 1), which had a focus on the manufacturing and utility sectors and which aimed to develop, refine and test the model. Every case study involved a 1-2 weeks visit on the company’s site to facilitate at least six data collection workshops with managers and employees, which took half a day each. In phase 2, information risks were identified and analyzed in three industrial case studies in a number of core business processes. The collected data was used to develop the initial model. Case company A was a large semiconductor manufacturer, case company B a family owned medium-sized steel manufacturer, and case company C was a medium-sized energy company. The initial studies were not aimed at fully understanding the investigated phenomena, but should rather assist in the development of the model. The model was then applied in two industrial case studies D and E by a final year MSc student for testing and refinement at two manufacturers. Using an independent facilitator shows that the model can be applied without the presence of the researcher. After each workshop, a feedback discussion took place to evaluate how the model and methodology can be improved and refined. The final version of the model was then applied in a last case study F, this time with the researcher as study facilitator.

Case	Industry	Number of employees	Annual turnover in £
A	Semiconductor	25,000	2.6 Billion
B	Steel	1,300	450 Million
C	Energy	256	75 Million
D	Electric, Electronics, Connectivity & Networks	3,500	380 Million
E	Electric, Electronics, Connectivity & Networks	440	47 Million
F	Water	2,000	1.047 Billion £

Table 1: Summary of Case Study Backgrounds

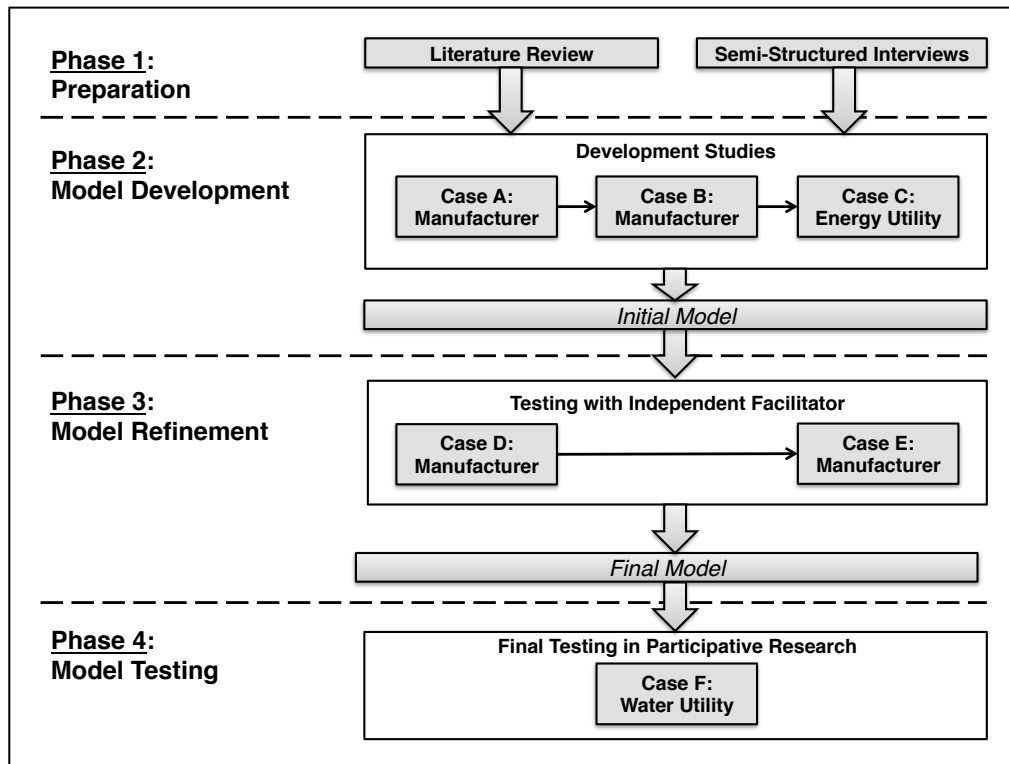


Figure 1: Research Methodology

A questionnaire was given at the end of each of case studies D, E and F to all workshop participants to judge the utility of the model. Five criteria were used to evaluate the model output: feasibility, usability, relevance, usefulness and confidence. Relevance, usefulness and confidence are three different aspects and measures of overall utility. The criteria ensure that the model strikes the right balance between industrial applicability and rigor of the model output. Each criterion was evaluated on a five step scale from 1 (very high), 2 (high), 3 (neutral), 4 (low) to 5 (very low). The results from the questionnaires are shown in Table 2. The five criteria were evaluated consistently high or very high in all three case studies. The refined version of the model used in case study F received a slightly better evaluation in terms of usefulness and confidence.

	Case D	Case E	Case F
Feasibility	1.67	1.50	1.90
Usability	2.28	1.58	1.70
Relevance	1.67	1.83	1.80
Usefulness	2.00	1.83	1.20
Confidence	2.00	1.58	1.40

Table 2: Evaluation Questionnaire Results in Case Studies D, E, F

2.1. Model Assumptions

The purpose of the model is to be feasible, usable and useful in an industrial environment. This requires a number of limiting assumptions. Risk changes over time, however, the model is static, as otherwise the data input would be by far too complex for the intended purpose. An information resource can have one or more information quality problems. Out of practical reasons, the model assumes that information quality problems / consequences are independent from each other. A metric for each business objective has to be defined, which can be financial or non-financial, but which has to be an interval scale in order to allow the statistical calculations needed, such as, for example, calculating the mean and standard deviation. A ratio scale is an affine line of ordered points without a true zero point, which can be used to calculate the size of the intervals in between data points. Absolute frequencies are used in the model to determine the number of times a task is executed. A time unit has to be defined for the model (as default, the model uses “year” as a time unit) that the absolute frequencies of task execution can refer to. This also determine the time unit of any risk calculation output subsequently. The model assumes that information risk treatments are aimed at modifying the causes of risk, hence the probability of information quality problems is only modified by an information risk treatment option.

3. Development of Model Based On Case Studies A, B, C

70 individual information risks were found in case studies A, B, C that have a significant level of risk. Some illustrative examples are shown in Table 3. Some patterns could be consistently observed and allow to formulate the following learnings. In many ways, there is uncertainty when it comes to the impact of information quality. This is due to the variety of external factors that influence the impact. The same faulty piece of information used for a task by a human decision maker can, in one case, lead to a poor decision, and, in another case, might not affect the quality of the decision at all. Even, if the decision is poor, the consequences of the poor decision can differ from one time to the other substantially. Take, for example, a poor maintenance decision, which can lead to machine failure in one case and not lead to any problems in another case.

Information Resource	IQ Problem	Direct Consequence	Intermediate Consequences	Overall Risk Evaluation
<i>Monthly volume forecast</i>	Investment planning needs to adjusted when there are errors in the forecast	Poor investment decisions	Too high investments lead to unnecessary costs Not enough investments lead to problems in production and delivery	Medium

<i>Information about personnel requirements</i>	Personnel planning decisions are based on wrong assumptions	Personnel planning decisions are wrong	There is not enough personnel or there is too much personnel to do the jobs. In the first instance, this can lead to delays in production, in the second instance, this causes unnecessary costs.	Medium
<i>Detailed technical vendor-specific information about assets</i>	The selection decision is sometimes based on poor technical information about asset	Equipment design and vendor are selected sub-optimally or simply wrong	There are problems in deployment and usage of asset, e.g. performance is not as high as expected.	Medium
<i>Asset user manual and description of the interface (SECS)</i>	Incorrect user manuals lead to wrong integration decisions	Poor asset integration decisions	Delays in completion of the asset integration, which leads to Complications after installation Higher costs	High
<i>Pricing of materials</i>	Order decisions are based on outdated pricing information	Order decisions are not optimal	Higher prices of ordered materials, around \$60000 yearly.	Medium
<i>Spare part usage data</i>	Spare parts are difficult to identify based on spare part usage data	It becomes very difficult and time consuming to find a cheaper supplier for a spare part.	The organization uses the old supplier for a spare part, although there are suppliers that are much cheaper. Up to 10% is paid more for materials (approx. 1 million dollar per production site yearly).	Very High

Table 3: Examples of Information Risks Identified in Case Studies

Furthermore, when information of poor quality is used for a task, it can lead to different direct consequences. These consequences can cause further consequences, and so fourth. In theory, there can be a never ending chain of consequences, however, in practice, we have never encountered such a case so far.

The consequences of information quality problems can affect the organization's ability to successfully deliver its business objectives, such as, for example, sales targets, costs, health & safety, customer satisfaction, employee satisfaction, etc.

The learnings from case studies A, B, C were used to build the model. After the model was tested in case studies D and E, a minor change was made to improve the usability and precision of the model. The final version of the model is presented in the next section.

4. Model Constructs

The model constructs are defined and explained in this section. A simplified illustration of the model with an example for each construct is shown in Figure 2.

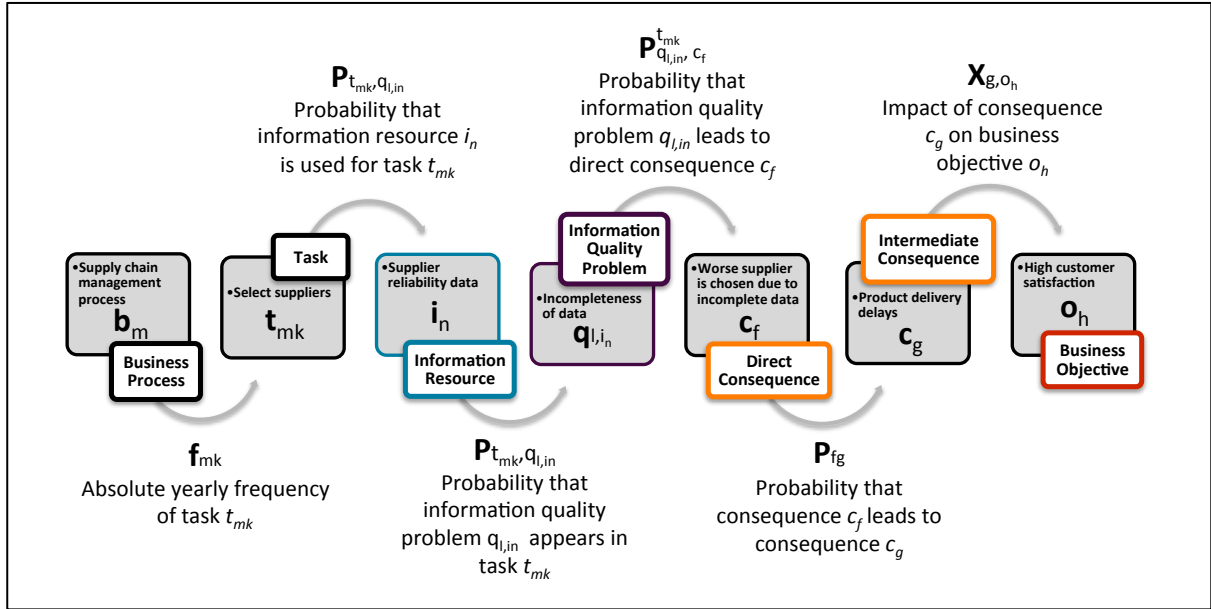


Figure 2: Simplified Illustration of Model with Examples

A *business process* is a sequence of activities with a defined input and output. Let b be the vector of business processes b_m for $m = 1 \dots B$. For instance, the vector b in a company consists of two business processes b_1 and b_2 , where the b_1 is the “supply chain management process” and b_2 is the “sales and customer support process”.

A *task* is a key activity in a business process, which is essential for the success of the business process. Each business process b_m has $k = 1 \dots T(b_m)$ tasks t_{mk} . Let $t(b_m)$ be the vector of tasks t_{mk} for each business process b_m . Let $f(b_m)$ be the vector of yearly (absolute) frequencies f_{mk} of how often a task t_{mk} of a business process b_m is executed. For example, the business process b_1 “supply chain management process” has a task vector $t(b_1)$, which consists of the task t_{11} “select suppliers” and task t_{12} “order supplies”. The task “select suppliers” is executed four times per year, while “order supplies” is executed 24 times per year, which leads to the frequency vector

$$f(b_1) = \begin{pmatrix} 4 \text{ per year} \\ 24 \text{ per year} \end{pmatrix}$$

To execute a task, information resources are needed. Let i be the vector of $n = 1 \dots I$ information resources i_n . Moreover, let $p^I(t_{mk})$ be the vector of the probabilities p_{t_{mk}, i_n} that an information resource i_n is required for the task t_{mk} . Note that for readability all $p^I(t_{mk})$ that are not explicitly specified are assumed to be a vector filled with 0 values. For example, task t requires the information resource i_1 “supplier reliability data” with a probability p_{t_{11}, i_1} of 20%.

An *information quality problem* (example in Figure 2: “Incompleteness of data”) arises when an information resource is not fit for the specific purpose of a task in a business process and the outcome

of the task is potentially influenced by this. Each information resource i_n has a vector of $l = 1 \dots Q(i_n)$ information quality problems q_{l,i_n} , which is denoted as $q(i_n)$. Furthermore, let $p^Q(t_{mk})$ be the vector of the probabilities $p_{t_{mk},q_{l,i_n}}$ that a information quality problem q_{l,i_n} appears in a task t_{mk} . Again, for readability reasons, all $p^I(t_{mk})$ that are not explicitly specified are assumed to be a vector filled with 0 values. For instance, information resource i_1 “supplier reliability data” suffers from information quality problem q_{1,i_1} “reliability data about a supplier does not exist” with a probability $p_{t_{11},q_{1,i_1}}$ of 8% when the information resource is needed for task t_{11} “select suppliers”.

An information quality problem can have one or more *direct consequences* (e.g. consequence c_1 “worse supplier is chosen due to incompleteness of data”) and each direct consequence can have one or more *intermediate consequences* (e.g. consequence c_2 “supply and delivery delays”). Let c therefore be the vector of $f = 1 \dots C$ direct and intermediate consequences c_f .

A *direct consequence* is the immediate effect of an IQ problem with a likelihood attached, which might directly impact one or more business objectives. Let $P_{Q(i_n) \times C}(i_n, t_{mk})$ be the matrix of probabilities $p_{q_{l,i_n},c_f}^{t_{mk}}$ that a quality problem q_{l,i_n} leads to the direct consequence c_f which has to be defined for each tuple (i_n, t_{mk}) :

$$\forall i_n, t_{mk}: P_{Q(i_n) \times C}(i_n, t_{mk}) = \begin{pmatrix} p_{q_{1,i_n},c_1}^{t_{mk}} & \dots & p_{q_{1,i_n},c_C}^{t_{mk}} \\ \vdots & \ddots & \vdots \\ p_{q_{Q(i_n),i_n},c_1}^{t_{mk}} & \dots & p_{q_{Q(i_n),i_n},c_C}^{t_{mk}} \end{pmatrix}$$

For example, if the consequence c_1 “worse supplier is chosen due to incompleteness of data” appears in 10% of times when information resource i_1 “supplier reliability data” is used for task t_{11} “select suppliers” and the probability of consequence c_2 “supply and delivery delays” resulting as a direct consequence is zero, it would create the following probability matrix:

$$P_{1 \times 2}(i_1: \text{Supplier reliability data}, t_{11}: \text{Select suppliers}) = (0.1 \quad 0)$$

An *intermediate consequence* is a consequence of a consequence with a likelihood attached, which might directly impact one or more business objectives. Note that intermediate consequences can cause further intermediate consequences. A consequence c_f can lead to an intermediate consequence c_g with a probability of p_{fg} defined in the matrix $P_{C \times C}$.

$$P_{C \times C} = \begin{pmatrix} p_{11} & \dots & p_{1C} \\ \vdots & \ddots & \vdots \\ p_{C1} & \dots & p_{CC} \end{pmatrix}$$

For instance, direct consequence c_1 “worse supplier is chosen due to incompleteness of data” can lead to intermediate consequences c_2 “supply and delivery delays” with a probability of 30%, which would lead to the following probability matrix:

$$P_{2 \times 2} = \begin{pmatrix} 0 & 0.3 \\ 0 & 0 \end{pmatrix}$$

Some of the consequences have an impact on one or more business objectives. *Business objectives* (example in Figure 2: “Customer satisfaction”) are the desired results of an organization, which are set by the executive leadership and typically formulated in the corporate mission. Business objectives are context-specific. They can be financial goals, e.g. maximizing revenues, but may also include other aspects like product quality, delivery times, customer satisfaction and environmental objectives. More formally, each organization has a vector o of $h = 1 \dots H$ objectives o_h that it aims to achieve, which are measured quantitatively, potentially using different measurement units, for instance o_1 might be measured in Pounds whereas o_2 could be measured in number of dissatisfied customers. A consequence c_f might directly impact an objective o_h . Let $x(c_f)$ be the vector of the direct impacts of each consequence c_f on each of the objectives o_1 to o_H . For example, the company has only one business objective o_1 which is to “maximize revenues”. Consequence c_2 “supply and delivery delays” reduces revenues by \$30,000 USD, hence $x(c_2) = (\$30,000)$.

An *information risk treatment* is a deliberate change in technology, organization or people behavior that modifies the probability of information quality problems with the aim to reduce the level of information risk. Let δ be the vector of $r = 1 \dots R$ information risk treatments δ_r . An information risk treatment δ_r can modify the probability vector $p^Q(t_{mk})$ of an information quality problems q_{l,i_n} in a task t_{mk} , which leads to the changed probability vectors $p^{Q,\delta_r}(t_{mk})$:

$$\forall t_{mk} \text{ modified by } \delta_r: \quad p^{Q,\delta_r}(t_{mk}) = \begin{pmatrix} p_{t_{mk},q_{1,i_n}}^{\delta_r} \\ \vdots \\ p_{t_{mk},q_{Q(i_n),i_n}}^{\delta_r} \end{pmatrix}$$

$$\forall t_{mk} \text{ not modified by } \delta_r: \quad p^{Q,\delta_r}(t_{mk}) = p^Q(t_{mk})$$

For instance, after the information risk treatment δ_1 “improve data collection of supplier reliability data”, information resource i_1 “supplier reliability data” suffers from information quality problem q_{1,i_1} “reliability data about a supplier does not exist” with a reduced probability $p_{t_{11},q_{1,i_1}}^{\delta_1}$ of 3% when the information resource is needed for task t_{11} “select suppliers” (which used to be 8% before treatment).

5. Calculating Information Risks and Benefits of Risk Treatments

The vector of yearly total risk for each objective o_h over all business processes b_m can be calculated using the following equation:

(1)

$$\Omega = \sum_{m=1}^B \sum_{k=1}^{T(b_m)} f_{mk} * \left(\sum_{n=1}^I p_{t_{mk}, i_n} * \left[\sum_{l=1}^{Q(i_n)} p_{t_{mk}, q_{l, i_n}} * \left(\sum_{f=1}^C p_{q_{l, i_n}, c_f}^{t_{mk}} * \omega_f \right) \right] \right)$$

A supporting function ω_f is needed, which calculates the vector of impact of a consequence c_f for each objective o_h . Note that ω_f is not dependent on a task t_{mk} and that ω_g has to be calculated recursively.

(2)

$$\forall c_f: \omega_f = x(c_f) + \sum_{g=1}^C p_{fg} * \omega_g$$

Similarly, the vector of yearly risk for each objective o_h for a single business process b_m can be calculated as follows:

(3)

$$\Omega_m = \sum_{k=1}^{T(b_m)} f_{mk} * \left(\sum_{n=1}^I p_{t_{mk}, i_n} * \left[\sum_{l=1}^{Q(i_n)} p_{t_{mk}, q_{l, i_n}} * \left(\sum_{f=1}^C p_{q_{l, i_n}, c_f}^{t_{mk}} * \omega_f \right) \right] \right)$$

Furthermore, the vector of yearly risk for each objective caused by an information resource i_n can be calculated using the equation:

(4)

$$\Omega_{i_n} = \sum_{m=1}^B \sum_{k=1}^{T(b_m)} f_{mk} * \left(p_{t_{mk}, i_n} * \left[\sum_{l=1}^{Q(i_n)} p_{t_{mk}, q_{l, i_n}} * \left(\sum_{f=1}^C p_{q_{l, i_n}, c_f}^{t_{mk}} * \omega_f \right) \right] \right)$$

If more granular details are needed, the vector of yearly risk for each objective caused by a single information quality problem q_{l, i_n} can be calculated as follows:

(5)

$$\Omega_{q_{l, i_n}} = \sum_{m=1}^B \sum_{k=1}^{T(b_m)} f_{mk} * \left(p_{t_{mk}, i_n} * \left[p_{t_{mk}, q_{l, i_n}} * \left(\sum_{f=1}^C p_{q_{l, i_n}, c_f}^{t_{mk}} * \omega_f \right) \right] \right)$$

Moreover, the vector of yearly total information risk for each objective o_h after the implementation of an information risk treatment can be calculated using the following equation:

(6)

$$\Omega_{\delta_r} = \sum_{m=1}^B \sum_{k=1}^{T(b_m)} f_{mk} * \left(\sum_{n=1}^I p_{t_{mk}, i_n} * \left[\sum_{l=1}^{Q(i_n)} p_{t_{mk}, q_{l, i_n}}^{\delta_r} * \left(\sum_{f=1}^C p_{q_{l, i_n}, c_f}^{t_{mk}} * \omega_f \right) \right] \right)$$

Finally, the vector of yearly benefits Δ_r for each objective o_h of implementing an information risk treatment δ_r can be calculated by subtracting the total information risk Ω_{δ_r} after the treatment from the total information risk Ω before treatment:

(7)

$$\Delta_r = \Omega - \Omega_{\delta_r}$$

6. Application Example

This is an illustrative example to illustrate the model, also shown in Figure 3. The example is similar to the data collected in the case studies, but is much smaller in terms of scope. We begin with the data input and then present the calculated data output by using the equations from the previous sections.

6.1. Data Input

In our example, there are two business processes: the “maintenance” process and the “purchasing” process. The vector of business processes is hence:

$$b = \begin{pmatrix} b_1: \text{Maintenance} \\ b_2: \text{Purchasing} \end{pmatrix}$$

The “maintenance” process consists of the three tasks “plan”, “execute” and “repair”:

$$t(b_1: \text{Maintenance}) = \begin{pmatrix} t_{11}: \text{Plan} \\ t_{12}: \text{Execute} \\ t_{13}: \text{Repair} \end{pmatrix}$$

The tasks are executed with a yearly frequency defined in vector f . The task “plan” is executed four times per year, the task “execute” 100 times per year, and the task “repair” 50 times per year.

$$f(b_1) = \begin{pmatrix} 4 \text{ per year} \\ 100 \text{ per year} \\ 50 \text{ per year} \end{pmatrix}$$

The “purchasing” process consists only of two tasks, which are “select suppliers” and “order supplies”

$$t(b_2) = \begin{pmatrix} t_{21}: \text{Select supplier} \\ t_{22}: \text{Order supplies} \end{pmatrix}$$

and the tasks are executed with the following yearly frequency:

$$f(b_2) = \begin{pmatrix} 2 \text{ per year} \\ 12 \text{ per year} \end{pmatrix}$$

In our scenario, there are five information resources that are needed for the two business processes:

$$i = \begin{pmatrix} i_1: \text{Asset manual} \\ i_2: \text{Asset condition data} \\ i_3: \text{Maintenance plan} \\ i_4: \text{Supplier ratings} \\ i_5: \text{Level of supplies} \end{pmatrix}$$

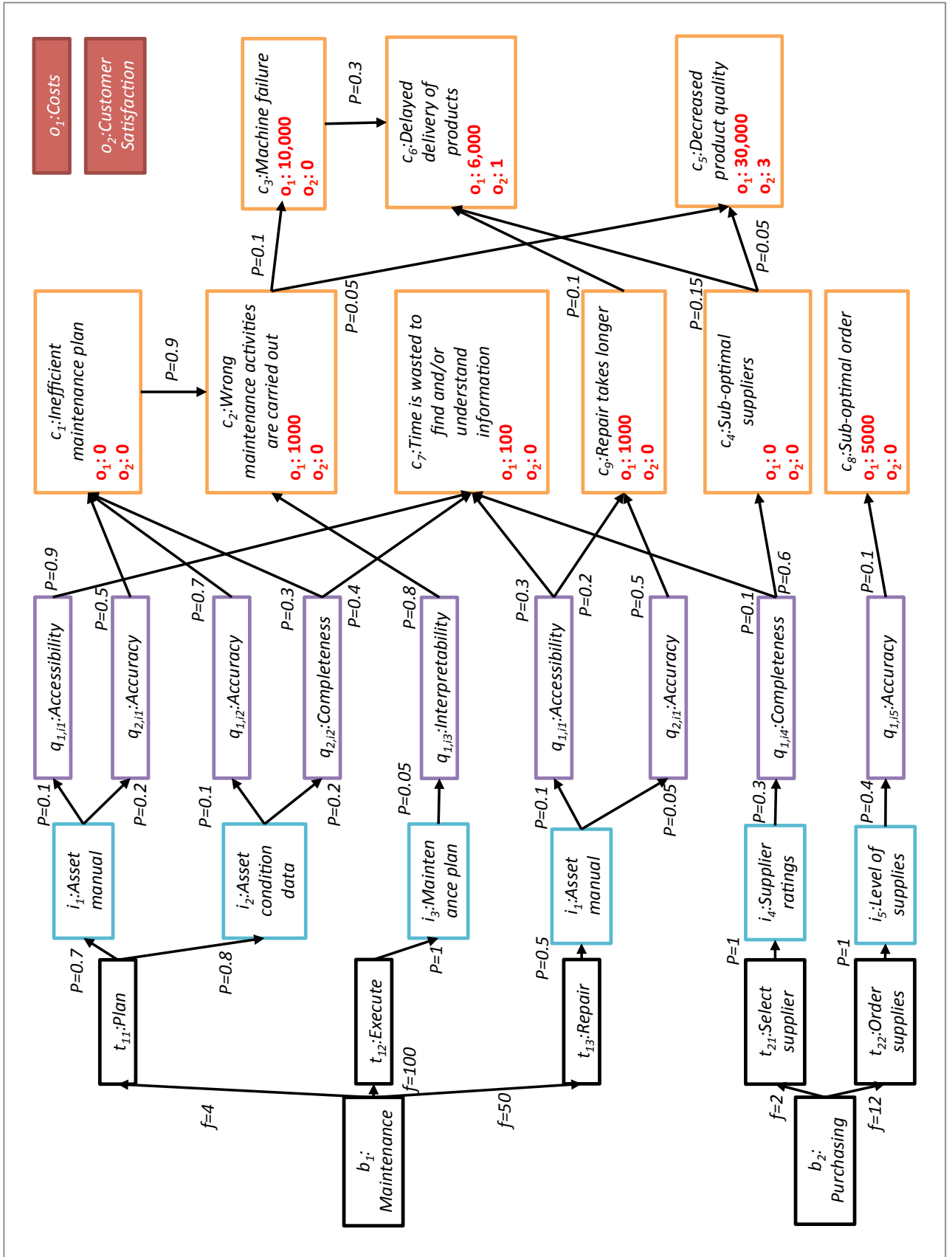


Figure 3: Illustration of Example

Information resource “asset manual” is required with a probability of 70% and the information resource “asset condition data” is needed in 80% of cases when the task “plan” is executed. Other information resources are not needed for this task. This leads to the following vector of probabilities:

$$p(t_{11}: Plan) = \begin{pmatrix} 0.7 \\ 0.8 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The probabilities for the other tasks are captured analogously:

$$\begin{aligned} p(t_{12}: Execute) &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} & p(t_{13}: Repair) &= \begin{pmatrix} 0.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ p(t_{21}: Select suppliers) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} & p(t_{22}: Order supplies) &= \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

Some of the information resources suffer from information quality problems. Information resource “asset manual”, for instance, has problems regarding accessibility and accuracy:

$$q(i_1: Asset manual) = \begin{pmatrix} q_{1,i_1}: Accessibility \\ q_{2,i_1}: Accuracy \end{pmatrix}$$

The information resource “asset condition data” suffers from inaccuracy and incompleteness:

$$q(i_2: Asset condition data) = \begin{pmatrix} q_{1,i_2}: Accuracy \\ q_{2,i_2}: Completeness \end{pmatrix}$$

The information resource “maintenance plan” has problems regarding interpretability:

$$q(i_3: Maintenance plan) = (q_{1,i_3}: Interpretability)$$

The information resource “supplier ratings” has problems regarding completeness:

$$q(i_4: Supplier ratings) = (q_{1,i_4}: Completeness)$$

And, finally, the information resources “level of supplies” also suffers from inaccuracy.

$$q(i_5: Level of supplies) = (q_{1,i_5}: Accuracy)$$

The probability that the information quality problems appear during the “plan” task when the information resource “asset manual” is used is 10% for accessibility and 20% for accuracy:

$$p(i_1: Asset manual, t_{11}: Plan) = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}$$

The other probabilities that information quality problems appear in the tasks during information usage are defined in the following vectors accordingly:

$$p(i_1: Asset manual, t_{13}: Repair) = \begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix}$$

$$p(i_2: Asset condition data, t_{11}: Plan) = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}$$

$$p(i_3: \text{Maintenance plan}, t_{12}: \text{Execute}) = (0.05)$$

$$p(i_4: \text{Supplier ratings}, t_{21}: \text{Select suppliers}) = (0.3)$$

$$p(i_5: \text{Level of supplies}, t_{22}: \text{Order supplies}) = (0.4)$$

All consequences of information quality problems, both direct and intermediate, are presented in vector c :

$$c = \begin{pmatrix} c_1: \text{Inefficient maintenance plan} \\ c_2: \text{Wrong maintenance activities are carried out} \\ c_3: \text{Machine failure} \\ c_4: \text{Sub – optimal suppliers} \\ c_5: \text{Decreased product quality} \\ c_6: \text{Delayed delivery of products} \\ c_7: \text{Time is wasted to find and/or understand information} \\ c_8: \text{Sub – optimal order} \\ c_9: \text{Repair takes longer} \end{pmatrix}$$

The probabilities that information quality problems lead to a direct consequence c_j are defined in the following matrices:

$$P_{2 \times 9}(i_1: \text{Asset manual}, t_{11}: \text{Plan}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.9 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$P_{2 \times 9}(i_1: \text{Asset manual}, t_{13}: \text{Repair}) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 \end{pmatrix}$$

$$P_{2 \times 9}(i_2: \text{Asset condition data}, t_{11}: \text{Plan}) = \begin{pmatrix} 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0.4 & 0 & 0 \end{pmatrix}$$

$$P_{1 \times 9}(i_3: \text{Maintenance plan}, t_{12}: \text{Execute}) = (0 \quad 0.8 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0)$$

$$P_{1 \times 9}(i_4: \text{Supplier ratings}, t_{21}: \text{Select suppliers}) = (0 \quad 0 \quad 0 \quad 0.1 \quad 0 \quad 0 \quad 0.6 \quad 0 \quad 0)$$

$$P_{1 \times 9}(i_5: \text{Level of supplies}, t_{22}: \text{Order supplies}) = (0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0.1 \quad 0)$$

For instance, if the information resource “asset manual” is inaccessible when it is needed for the task “plan”, there is a 90% probability that it leads to consequence c_7 (i.e. that “time is wasted to find and/or understand information”).

The following matrix shows the probabilities that a consequence follows a consequence. For example, there is a 90% probability that consequence c_1 leads to c_2 and a 10% probability that consequence c_2 leads to c_3 .

$$P_{9 \times 9} = \begin{pmatrix} 0 & 0.9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0.05 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.05 & 0.15 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 \end{pmatrix}$$

The organization in our example has two business objectives: minimizing costs and reducing the number of dissatisfied customers. The vector o is therefore:

$$o = \begin{pmatrix} o_1: Costs \\ o_2: Customer Satisfaction \end{pmatrix}$$

$o_1: Costs$ is measured in additional cost generated in thousands of US Dollars. $o_2: Customer Satisfaction$ is measured in thousands of Dissatisfied Customers (a higher value means a lower customer satisfaction; i.e. 1 means 1,000 more dissatisfied customers).

Consequences c_1 and c_4 have no direct impact on any of the objectives and hence:

$$x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad x_4 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The other consequences have a direct impact on the ability of the organization to deliver successfully its business objectives:

$$\begin{aligned} x_2 &= \begin{pmatrix} \$1,000t \\ 0t \end{pmatrix} & x_3 &= \begin{pmatrix} \$10,000t \\ 0t \end{pmatrix} & x_5 &= \begin{pmatrix} \$30,000t \\ 30t \end{pmatrix} \\ x_6 &= \begin{pmatrix} \$6,000t \\ 10t \end{pmatrix} & x_7 &= \begin{pmatrix} \$100t \\ 0t \end{pmatrix} & x_8 &= \begin{pmatrix} \$5,000t \\ 0t \end{pmatrix} & x_9 &= \begin{pmatrix} \$1,000t \\ 0t \end{pmatrix} \end{aligned}$$

For instance, consequence c_2 creates \$1,000,000 impact on the objective o_1 and has no impact on objective o_2 . Consequence c_5 creates \$30,000,000 additional costs and leads to 30,000 additional customers who are dissatisfied.

The organization has identified two potential information risk treatments. The first treatment option is to redesign the data collection processes. The second treatment option is to improve the understandability of the maintenance report. The vector of information risk treatments is hence:

$$\delta = \begin{pmatrix} \delta_1: Data collection process redesign \\ \delta_2: Improve understandability of maintenance report \end{pmatrix}$$

Once the information risk treatments are implemented, it will change the likelihood of information quality problems. Redesigning the data collection processes will reduce the probability of accuracy problems regarding the asset manual, asset condition data and level of supplies. It will not change the likelihood of the accessibility problem of the asset manual and the incompleteness of the asset condition data. The modified vectors of probabilities of quality problems in tasks are presented in the following and the changed probabilities are highlighted in bold letters.

$$p^{\delta_1}(i_1: Asset manual, t_{11}: Plan) = \begin{pmatrix} 0.1 \\ \mathbf{0.1} \end{pmatrix}$$

$$p^{\delta_1}(i_1: Asset manual, t_{13}: Repair) = \begin{pmatrix} 0.1 \\ \mathbf{0.02} \end{pmatrix}$$

$$p^{\delta_1}(i_2: Asset condition data, t_{11}: Plan) = \begin{pmatrix} \mathbf{0.06} \\ 0.2 \end{pmatrix}$$

$$p^{\delta_1}(i_5: Level of supplies, t_{22}: Order supplies) = (\mathbf{0.2})$$

Changing the data collection processes has no effect at all on the interpretability of the maintenance plan and the completeness of supplier ratings in our example, hence:

$$p^{\delta_1}(i_3: Maintenance plan, t_{12}: Execute) = p(i_3: Maintenance plan, t_{12}: Execute)$$

$$p^{\delta_1}(i_4: Supplier ratings, t_{21}: Select suppliers) = p(i_4: Supplier ratings, t_{21}: Select suppliers)$$

The second information risk treatment is to improve the understandability of the maintenance report and, thus, fully eliminates the interpretability problems connected to the maintenance report:

$$p^{\delta_2}(i_3: \textit{Maintenance plan}, t_{12}: \textit{Execute}) = (0)$$

The probabilities of all other information quality problems remain unaffected by this treatment:

$$p^{\delta_2}(i_1: \textit{Asset manual}, t_{11}: \textit{Plan}) = p(i_1: \textit{Asset manual}, t_{11}: \textit{Plan})$$

$$p^{\delta_2}(i_1: \textit{Asset manual}, t_{13}: \textit{Repair}) = p(i_1: \textit{Asset manual}, t_{13}: \textit{Repair})$$

$$p^{\delta_2}(i_2: \textit{Asset condition data}, t_{11}: \textit{Plan}) = p(i_2: \textit{Asset condition data}, t_{11}: \textit{Plan})$$

$$p^{\delta_2}(i_4: \textit{Supplier ratings}, t_{21}: \textit{Select suppliers}) = p(i_4: \textit{Supplier ratings}, t_{21}: \textit{Select suppliers})$$

$$p^{\delta_2}(i_5: \textit{Level of supplies}, t_{22}: \textit{Order supplies}) = p(i_5: \textit{Level of supplies}, t_{22}: \textit{Order supplies})$$

6.2. Data Output

The organization is now able to calculate the level of information risks per annum and the benefits of the proposed information risk treatments. The total yearly information risk can be calculated using equation (1). The results are shown in Table 4.

Symbol	o₁:Costs (additional USD yearly)	o₂:Customer Satisfaction (# of dissatisfied customers yearly)
Ω	\$21,221,000 additional costs	10,530 additional dissatisfied customers

Table 4: Total Expected Information Risk (Per Year)

Expected information risk in individual business processes

Sometimes, it can be useful to know the level of information risk for each individual business process to understand if action is required. This can be calculated using equation (2) and the results of this calculation are shown in Table 5.

Symbol	Business process	o₁:Costs (additional USD yearly)	o₂:Customer Satisfaction (# of dissatisfied customers yearly)
Ω_1	b_1 : Maintenance process	\$18,951,000	9,450
Ω_2	b_2 : Purchasing process	\$3,270,000	1,080

Table 5: Expected Information Risk in Individual Business Processes (Per Year)

Expected information risk caused by individual information quality problems

In many situations, it is important to understand how much information risk is caused across all business processes by each of the information quality problems. This can be used to decide which

information quality problems should get more attention. The calculation can be done by using equation (3). The results are presented in Table 6.

Symbol	Information Resource	Information Quality Problem	$\mathbf{o_1}$:Costs (additional USD yearly)	$\mathbf{o_2}$:Customer Satisfaction (# of dissatisfied customers yearly)
Ω_{q_1,i_1}	i_1 : Asset manual	q_{1,i_1} : Accessibility	\$900,000	500
Ω_{q_2,i_1}	i_1 : Asset manual	q_{2,i_1} : Accuracy	\$1,927,000	1,100
Ω_{q_1,i_2}	i_2 : Asset condition data	q_{1,i_2} : Accuracy	\$741,000	400
Ω_{q_2,i_2}	i_2 : Asset condition data	q_{2,i_2} : Completeness	\$661,000	300
Ω_{q_1,i_3}	i_3 : Maintenance plan	q_{1,i_3} : Interpretability	\$14,720,000	7200
Ω_{q_1,i_4}	i_4 : Supplier rating	q_{1,i_4} : Completeness	\$870,000	1100
Ω_{q_1,i_5}	i_5 : Level of supplies	q_{1,i_5} : Accuracy	\$2,400,000	0

Table 6: Expected Information Risk Caused By Individual Information Quality Problems (Per Year)

Total expected risk after implementation of information risk treatment options

The organization in our example has identified two potential information risk treatment options. Each of the options modifies the level of information risk in a different way. The expected yearly total risk after implementation of the information risk treatments can be calculated with equation (4). The calculated figures are shown in Table 7 for each information risk treatment.

Symbol	Information Risk Treatment	$\mathbf{o_1}$:Costs (additional USD yearly)	$\mathbf{o_2}$:Customer Satisfaction (# of dissatisfied customers yearly)
Ω_{δ_1}	δ_1 : Data collection process redesign	\$19,660,000	9,900
Ω_{δ_2}	δ_1 : Improve understandability of maintenance report	\$7,500,000	3,400

Table 7: Total Expected Information Risk After Implementation of Information Risk Treatments (Per Year)

Expected benefits of information risk treatments

Finally, if the results should be used to build a business case for the information risk treatment, it is important to quantify the yearly benefits of implementing the each of the proposed information risk treatment options by using equation (5), as shown in Table 8.

Symbol	o₁:Costs (savings in USD yearly)	o₂:Customer Satisfaction (reduced # of dissatisfied customers yearly)
Δ_1	\$1,561,000	630
Δ_2	\$13,721,000	7130

Table 8: Expected Benefits of Information Risk Treatments (Per Year)

In practice, the benefits are usually realized only gradually. Thus, it needs to be predicted, when and how quickly the benefits of the information risk treatment will be obtained. For example, data on the geographical location of physical infrastructure assets can be improved by giving handheld devices to the field engineers, who can then update the data. In the first year, approx. 10% of the data will be improved per year, because the field engineers must still learn how to use the handhelds effectively, and in the second year, already 20% of the data can be improved. As less data is faulty, the improvement rate will slow down in the years after.

7. Methodology to Collect Input for Model in Industrial Organizations

7.1. Identify tasks in each business process

A number of business processes has to be selected to be included in the scope of the analysis. In this step, each business process b_m in the scope has to be analysed as a preparation to identify information quality problems and risks in the business process in a workshop with one or more business process representatives (which were chosen in step A1). This step requires to identify key tasks in the business process as input to the model:

$$t(b_m) = \begin{pmatrix} t_{m1} \\ \vdots \\ t_{mT(b_m)} \end{pmatrix}$$

Each task and the people involved in the task should be described. Moreover, the absolute frequency of task execution for each task should be estimated in the timeframe set, based on previous experiences and available data as input to the model:

$$f(b_m) = \begin{pmatrix} f_{m1} \\ \vdots \\ f_{mT(b_m)} \end{pmatrix}$$

7.2. Examine information resources needed for each task

This step requires to identify information resources that are needed (and used if available) for each task t_{mk} of a business process b_m in the scope, which are represented in the model as:

$$i = \begin{pmatrix} i_1 \\ \vdots \\ i_I \end{pmatrix}$$

Each information resource should be described, in particular, how the information resource is created, processed and accessed to make exactly sure which information resource is meant and to give relevant context. Note that not every information resource that is required for a task has to be available. Some information resources might not be considered as they are out of the process scope. The probability that information is used for this task has to be estimated, which is the following parameter in the model:

$$p^I(t_{mk}) = \begin{pmatrix} p_{t_{mk}, i_1} \\ \vdots \\ p_{t_{mk}, i_I} \end{pmatrix}$$

If an information resource is not used, the probability is 0 per definition.

7.3. Identify information quality problems during task execution

In order to identify information quality problems, each information resources that is required for a task is evaluated from an information user's perspective along the chosen dimensions in A1 by the business process representatives. A five level scale can be used that indicates how fit for use an information resource is (i.e. very low, low, medium, high, very high). If information quality is "high" or "very high", the business process representatives should be asked if they are sure that there are no problems. If the information quality is medium or lower, the business process representatives are asked to describe the problem. Each identified information quality problem has to be documented and provides the following input to the model:

$$\forall i_n: \quad q(i_n) = \begin{pmatrix} q_{1, i_n} \\ \vdots \\ q_{Q(i_n), i_n} \end{pmatrix}$$

Moreover, the probability for each information quality problem that it appears when information is used for the task has to be estimated:

$$\forall t_{mk}: \quad p^Q(t_{mk}) = \begin{pmatrix} p_{t_{mk}, q_{1, i_n}} \\ \vdots \\ p_{t_{mk}, q_{Q(i_n), i_n}} \end{pmatrix}$$

Each information quality problem should be described in more detail to make sure that other users refer to the right information quality problem.

7.4. Identify consequences of information quality problems

In this step, the direct and intermediate consequences of information quality problems are identified, which are the following model input:

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_C \end{pmatrix}$$

The direct consequences should be directly related to the activity in the business process in which an information quality problem appears. A consequence can lead to several further consequences. The intermediate consequences can appear in other activities or even outside the business process or organization. Each consequence should be given a textual description.

7.5. Identify for each consequence the business objectives that are affected

For each consequence, it has to be identified if the consequence has a direct impact on one or more business objectives, represented as the vector o in the model:

$$o = \begin{pmatrix} o_1 \\ \vdots \\ o_H \end{pmatrix}$$

This can be done by adding a description to each consequence node in the visual representation. It should also be explained why and how the consequence has an impact on the business objective

7.6. Examine existing risk controls

This step examines if there are existing risk controls in place to prevent information quality problems and/or its consequences. An example of a risk control can be an engineer who reads the opinions of other engineers in the Internet before he uses an asset manual provided by a new supplier, because the asset manuals can be unreliable.

7.7. Estimate probabilities and impact of each consequence

For each consequence, a probability and the impact is determined under consideration of the current risk controls that are in place, which is based on previous experiences, existing data and expert judgments. First, for each direct consequence the probabilities have to be estimated for each information quality problem and each task, which is captured as the following in the model:

$$\forall i_n, t_{mk}: P_{Q(i_n) \times C}(i_n, t_{mk}) = \begin{pmatrix} p_{q1, i_n, c1}^{t_{mk}} & \cdots & p_{q1, i_n, cC}^{t_{mk}} \\ \vdots & \ddots & \vdots \\ p_{qQ(i_n), i_n, c1}^{t_{mk}} & \cdots & p_{qQ(i_n), i_n, cC}^{t_{mk}} \end{pmatrix}$$

Afterwards, the conditional probabilities of the intermediate consequences has to be estimated, which form the matrix $P_{C \times C}$:

$$P_{C \times C} = \begin{pmatrix} p_{11} & \cdots & p_{1C} \\ \vdots & \ddots & \vdots \\ p_{C1} & \cdots & p_{CC} \end{pmatrix}$$

Eventually, when a consequence has an impact on one or more business objectives, the impact has to be estimated in the measurement unit specified for the business objective, captured as the vector $x(c_f)$ which has to be specified for each consequence:

$$\forall c_f: x(c_f) = \begin{pmatrix} x_{fo_1} \\ \vdots \\ x_{fo_H} \end{pmatrix}$$

If a consequence does not have an impact, for example, on business objective 3, the value at position 3 in the vector would be zero. Moreover, if the consequence does not have any impact on any of the business objectives, all values in the vector would be 0 as per definition.

7.8. Refine numbers and verify results

For each business process, further subject matter experts should be chosen to refine the numbers and verify the results from information risk analysis to reduce the bias in the input. If possible, historical data should be used to improve the reliability of the data input. Sometimes, it is advisable to collect further data where feasible or execute enhanced data analysis to improve the numbers input.

8. Operationalization of Model in Case Studies D, E, F

The model was tested in case studies D, E, F using the methodology described in section 7. The major challenge to operationalize the model was to capture the quantitative inputs. Whenever exact values were available for the quantitative input, they were used directly as model input. Both the qualitative and quantitative inputs yet relied to a large degree on expert judgment, as historical data was often not readily available, especially for the probabilities. Expert judgment is a widely used as data input in risk management, e.g. (Cooke and Goossens 2004; Boholm 2010; Celeux et al. 2006; Cooke and Goossens 2004; Evans et al. 1994; Otway and Winterfeldt 1992). Furthermore, numerical ranges were used if it was difficult for the experts to provide precise numbers, as suggested by (Hubbard 2010).

This required to determine or assume a probability distribution. A distribution frequently used in expert-based risk assessment is the triangular distribution (Johnson 1997; René van Dorp and Kotz 2002; Stein and Kebelis 2009; International Organization for Standardization 2009b). It has three input

parameters: the lower boundary, the upper boundary and the mode, which is the most frequently occurring value. The triangular distribution was hence used as a default for the operationalization of the model as a default whenever an exact value was difficult to determine for a quantitative model input in the process. In cases that a mode could not be determined by the expert and the expert did not know the shape of the curve (which is very likely in this case), an uniform distribution was used, as it requires only a lower and upper boundary and it needs no information about the shape of the probability curve is available. A Monte Carlo simulation was then used to make the calculations (1) to (7). This applies to all the quantitative inputs to the model that were collected as model input. The model input was checked and verified by additional subject matter experts to improve the reliability of the results. An example of the model output is presented in Table 9. It shows the yearly impact of the nine most severe information quality problems. For each business objective, a defined measurement metric was used.

Rank	Information Resources (Information Quality Problems)	Operational Efficiency (higher costs in USD yearly)	Customer Satisfaction (# of industry regulatory penalty points)	Health and Safety (# of people under health & safety risk)	Employee Satisfaction (# of employee dissatisfied)
1	<i>Completeness and Accuracy of Asset Spatial Location Data</i>	\$ 1,248,000	0.51	450	1230
2	<i>Accessibility and Interpretability of the Network Incident Situational Awareness Data</i>	\$ 455,000	0.52	2650	0
3	<i>Accuracy and Completeness of the Asset Criticality Data</i>	\$ 312,000	0.63	0	0
4	<i>Accuracy and Completeness of the Asset Material Data</i>	\$ 169,000	0.34	100	0
5	<i>Accuracy and Completeness of the Asset Sizing Data</i>	\$ 52,000	0.35	0	0
6	<i>Interpretability of Asset Ownership Data</i>	\$ 78,000	0	0	2104
7	<i>Accuracy and Completeness of the Asset Performance Assessment Data</i>	\$ 104,000	0.03	0	0
8	<i>Completeness of the Asset Rehabilitation History Data</i>	\$ 65,000	0	0	0
9	<i>Accuracy and Completeness of the Asset Condition Data</i>	\$ 39,000	0	0	0

Table 9: Model Output Example - List of Information Risks and Their Yearly Business Impact

A spreadsheet based tool was used to capture the model input in case studies D and E. After case study E, a minor change in the model was made. In the initial model, the absolute frequency of the information quality problem was estimated instead of a probability. A subject matter expert who wants to estimate this number needs to think about how often they execute a task, how likely it is that they use an information and how often the information usage leads to an information quality problem. This calculation has to be done in the head of the business process representatives when giving an answer to the question, how often an information quality problem appears. By explicitly considering these elements as model input for the calculations, the complexity of giving input to the model is reduced which makes it easier to provide reliable inputs for the experts.

Moreover, for case study F, a software tool was developed that implements the model, including the Monte Carlo, and allows to capture the model input in form of a mind map in a user friendly manner. A screenshot of the user interface of the software tool is shown in Figure 4. The Monte Carlo simulation run by the software tool generates 10,000 random numbers for each node as basis for calculations as a default value as this produced results in a satisfying time and small enough 95% confidence interval for the purpose of usage. Alternatively, it can be set up to produce random numbers until a defined 95% interval is reached.

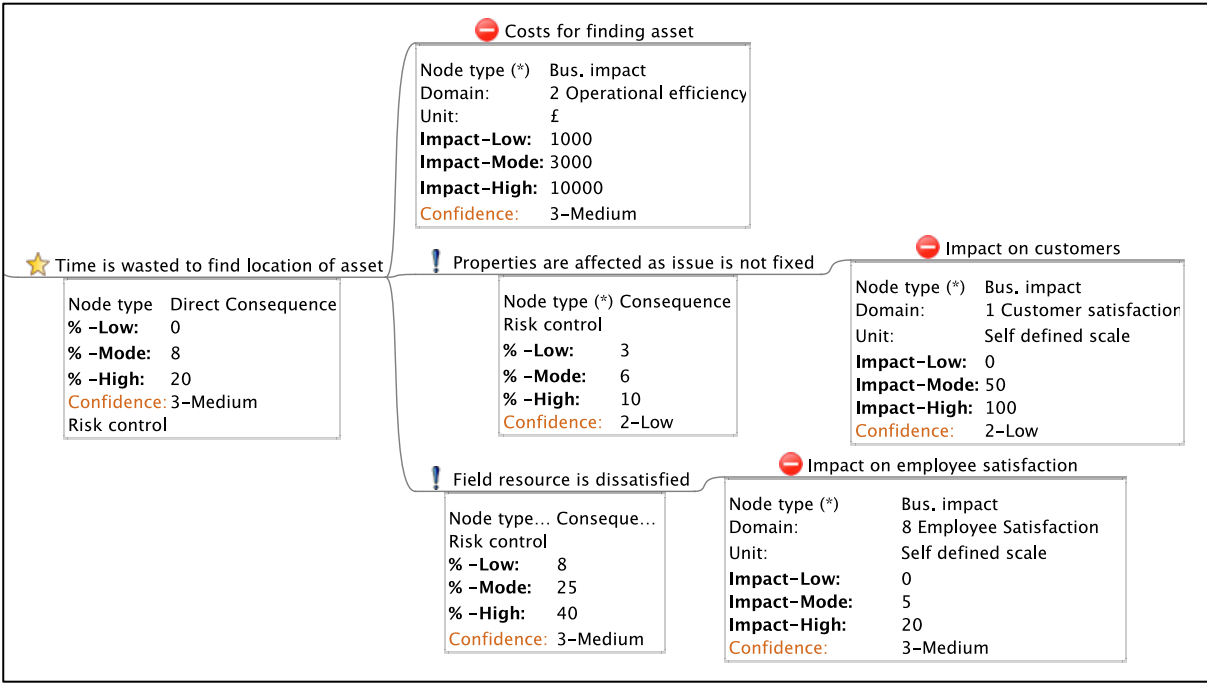


Figure 4: Information Risk Modelling As Displayed in Software Tool in Case F

The model output was used by the managers of case company F to build a business case for an information governance initiative and a new IT system for infrastructure assets. The project manager

at company F emphasized that it “was particularly pleasing to make significant tangible benefits from the analysis” and that he would “be happy to recommend this approach to anyone trying to understand the value of their data and information assets.” Another manager at company F highlighted that “the techniques and model used have exceeded our expectations in terms of the usefulness of the outputs” and that they are “well worthwhile and directly applicable to addressing real current challenges”.

9. Conclusion

Due to the massive amount of data available and the rising capabilities and complexity of information systems and organizations, information quality creates increasingly risks in every organization. A prerequisite of effective information risk management is the assessment of these risks. This paper has presented a risk-based approach to model and calculate the financial impact of information quality problems in organizations, which, so far, has been a very unexplored area that is yet of high relevance for data managers and information quality practitioners. The model is a contribution to the existing knowledge base as it provides an effective approach to quantitatively assess risks that arise from information quality and to use these results for information quality improvement. A limitation of this research is that the validation of the utility of the model has followed mostly from the perceptions of users and experts. A stronger validation would have been to verify in a few years time if the expected benefits of information risk treatment options could have been actually achieved in the case study organizations. As with all models, the quality of the output is highly depended on the quality of the input data, known as the “garbage in – garbage out” principle. This paper is, however, primarily concerned with providing a mathematical foundation to calculate the impact of information risks. More generally, this research has also shown that the information quality business impact is not as intangible and immeasurable as often assumed. Detailed quantitative input can be obtained from experts by using ranges for the numerical inputs and by conducting the calculations using a Monte Carlo simulation. The model could be potentially applied to predict and measure benefits of information quality improvement projects. Finally, the research shows that transferring risk management concepts and techniques to information quality management can be, overall, a very fruitful endeavour. The concepts and techniques might need to be refined to be used in an information quality context, but the general principles do not seem to change substantially.

10. References

- Ahituv, N. 1980. “A Systematic Approach Toward Assessing the Value of an Information System.” *MIS Quarterly* 4 (4): 61–75.
- Askira Gelman, I. 2011. “GIGO or Not GIGO: The Accuracy of Multi-Criteria Satisficing Decisions.” *Journal of Data and Information Quality (JDIQ)* 2 (2): 9.
- Ballou, D. P., and H. L. Pazer. 1985. “Modeling Data and Process Quality in Multi-input, Multi-output Information Systems.” *Management Science* 31 (2): 150–162.

- Ballou, D., R. Wang, H. Pazer, and G. Kumar. 1998. "Modeling Information Manufacturing Systems to Determine Information Product Quality." *Management Science* 44 (4): 462–484.
- Boholm, A.A. 2010. "On the Organizational Practice of Expert-based Risk Management: A Case of Railway Planning." *Risk Management* 12 (4): 235–255.
- Borek, A., A.K. Parlikad, and P. Woodall. 2011. "Towards A Process For Total Information Risk Management (TIRM)." In *Proceedings of the 16th International Conference on Information Quality (ICIQ 2011)*. Adelaide, Australia.
- Celeux, G., F. Corset, A. Lannoy, and B. Ricard. 2006. "Designing a Bayesian Network for Preventive Maintenance from Expert Opinions in a Rapid and Reliable Way." *Reliability Engineering and System Safety* 91 (7): 849–856.
- Chengalur-Smith, I. N., D. P. Ballou, and H. L. Pazer. 1999. "The Impact of Data Quality Information on Decision Making: An Exploratory Analysis." *IEEE Transactions on Knowledge and Data Engineering* 11 (6): 853–864.
- Cooke, R. M., and L. H. J. Goossens. 2004. "Expert Judgement Elicitation for Risk Assessments of Critical Infrastructures." *Journal of Risk Research* 7 (6): 643–656.
- DeLone, W. H, and E. R McLean. 2003. "The DeLone and McLean Model of Information Systems Success: A Ten-year Update." *Journal of Management Information Systems* 19 (4): 9–30.
- DeLone, W. H, and E. R McLean. 1992. "Information Systems Success: The Quest for the Dependent Variable." *Information Systems Research* 3 (1): 60–95.
- English, L. P. 1999. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. John Wiley & Sons.
- Eppler, M., and M. Helfert. 2004. "A Classification and Analysis of Data Quality Costs." In *Proceedings of the 9th International Conference on Information Quality (ICIQ-04)*, 311–325.
- Evans, J. S., G. M. Gray, R. L. Sielken, A. E. Smith, C. Valdezflores, and J. D. Graham. 1994. "Use of Probabilistic Expert Judgment in Uncertainty Analysis of Carcinogenic Potency." *Regulatory Toxicology and Pharmacology* 20 (1): 15–36.
- Even, Adir, and G. Shankaranarayanan. 2007. "Utility-driven Assessment of Data Quality." *SIGMIS Database* 38 (2): 75–93. doi:10.1145/1240616.1240623.
- Fisher, C. W, and B. R Kingma. 2001. "Criticality of Data Quality as Exemplified in Two Disasters." *Information & Management* 39 (2): 109–116.
- Ge, M., and M. Helfert. 2008. "Effects of Information Quality on Inventory Management." *International Journal of Information Quality* 2 (2): 177–191.
- Hevner, A. R, S. T March, J. Park, and S. Ram. 2004. "Design Science in Information Systems Research." *MIS Quarterly* 28 (1): 75–105.
- Hubbard, D. W. 2010. *How to Measure Anything: Finding the Value of Intangibles in Business*. John Wiley & Sons.
- International Organization for Standardization. 2009a. "ISO 31000:2009 Risk Management – Principles and Guidelines on Implementation." http://www.iso.org/iso/catalogue_detail?csnumber=43170.
- . 2009b. "ISO/IEC 31010:2009 - Risk Management -- Risk Assessment Techniques." http://www.iso.org/iso/catalogue_detail?csnumber=43170.
- Johnson, D. 1997. "The Triangular Distribution as a Proxy for the Beta Distribution in Risk Analysis." *Journal of the Royal Statistical Society: Series D (The Statistician)* 46 (3): 387–398.
- Jung, W., L. Olfman, T. Ryan, and Y. T Park. 2005. "An Experimental Study of the Effects of Representational Data Quality on Decision Performance." *AMCIS 2005 Proceedings*: 298.
- Keller, Kevin Lane, and Richard Staelin. 1987. "Effects of Quality and Quantity of Information on Decision Effectiveness." *The Journal of Consumer Research* 14 (2) (September): 200–213.
- March, S. T, and G. F Smith. 1995. "Design and Natural Science Research on Information Technology." *Decision Support Systems* 15 (4): 251–266.
- Loshin, D. 2001. *Enterprise Knowledge Management: The Data Quality Approach*. Morgan Kaufmann.
- . 2010. *The Practitioner's Guide to Data Quality Improvement*. Morgan Kaufmann.
- McGilvray, D. 2008. *Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*. Morgan Kaufmann.

- O'Reilly III, C. A. 1982. "Variations in Decision Makers' Use of Information Sources: The Impact of Quality and Accessibility of Information." *Academy of Management Journal* 25 (4): 756–771.
- Otway, H., and D. Winterfeldt. 1992. "Expert Judgment in Risk Analysis and Management: Process, Context, and Pitfalls." *Risk Analysis* 12 (1): 83–93.
- Parssian, A., S. Sarkar, and V. S. Jacob. 2004. "Assessing Data Quality for Information Products: Impact of Selection, Projection, and Cartesian Product." *Management Science* 50 (7): 967–982.
- Raghunathan, Srinivasan. 1999. "Impact of Information Quality and Decision-maker Quality on Decision Quality: a Theoretical Model and Simulation Analysis." *Decision Support Systems* 26 (4) (October): 275–286. doi:10.1016/S0167-9236(99)00060-3.
- Redman, T.C. 1998. "The Impact of Poor Data Quality on the Typical Enterprise." *Communications of the ACM* 41 (2): 79–82. doi:10.1145/269012.269025.
- Ren é van Dorp, J., and S. Kotz. 2002. "A Novel Extension of the Triangular Distribution and Its Parameter Estimation." *Journal of the Royal Statistical Society: Series D (The Statistician)* 51 (1): 63–79.
- Shankaranarayanan, G., and Y. Cai. 2006. "Supporting Data Quality Management in Decision-making." *Decision Support Systems* 42 (1): 302–317.
- Slone, J. P. 2006. "Information Quality Strategy: An Empirical Investigation of the Relationship Between Information Quality Improvements and Organizational Outcomes". Doctoral Thesis, Minneapolis, MN, USA: Capella University.
- Stein, W. E., and M. F. Keblis. 2009. "A New Method to Simulate the Triangular Distribution." *Mathematical and Computer Modelling* 49 (5-6): 1143–1147.
- Supatgiat, C., C. Kenyon, and L. Heusler. 2006. "Cause-to-effect Operational-risk Quantification and Management." *Risk Management*: 16–42.
- Wang, R. Y., and D. M Strong. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems* 12 (4): 33.