

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,100

Open access books available

167,000

International authors and editors

185M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Quality 4.0: Data Quality and Integrity: A Computational Approach

Rob Christiaanse

Abstract

The use of modern techniques, such as IOT, AI, and machine learning, revolutionized the idea of quality and quality control. Auditors face a tidal wave of data. One of the key challenges is how to determine the quality of the data, systems and processes produce. We propose a computational model to learn the inherent uncertainty to data integrity subsumed in the claims actually done by stakeholders within and outside the organization. The decision procedure combines two strong forms of obtaining audit evidence. These two forms are external conformation and re-performance. The procedure fits in the current modern computational idea data-driven assurance, which is consistent with quality 4.0 concepts in quality control and quality audit practices.

Keywords: data integrity, measurement, uncertainty, quality control, quality audit, quality standards, assurance, data quality

1. Introduction

It is to be expected that emerging technologies will have a profound impact on how we produce products, grow our food, use resources, organize services, and so on. We call these systems cyber-physical systems (CPS). A key characteristic of CPS is that information is infused in physical infrastructures to improve performance, flexibility, up-time of machines, product quality, minimize rejection rates, and improve the perceived product and service quality by end users, such as customers, regulators, and other stakeholders coined as society at large [1].

The key idea buttressing these developments lies in the realm of measuring things. In our daily lives, the act of measuring things is very important. Think of the simple act of buying groceries, in our case tomatoes, at your local grocery store. You enter the store and ask for 2 kilograms of tomatoes shown on the counter display. The grocer will pick the tomatoes and put them on a scale. He or she reads from the device the weight in grams. After having done this, the grocer will calculate the amount due in the local currency. You pay the amount due to the grocer, and he or she will hand over the tomatoes you have just bought. The problem you might face is whether the measured quantity is correct, which is the physical quantity in metrology subject to measurement. In general, metrology is the science of measurement and its

applications. A measurement is said to be traceable when an unbroken chain of calibrations is established to a specific reference measurement standard, in particular realizations of the measurement units of the international system of units (SI) [2]. So, the grocer must be aware that the scale he or she uses does, in fact, measure grams in the correct way. Needless to say that this is far from simple to establish by him or herself. So, the grocer trusts the claim of the manufacturer of the scale. But there is an extra problem you might be aware of. The tomatoes you have just bought must be compliant with the regulation (EC) No 178/2002 of the European Parliament and of the council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety. The general food law regulation is the foundation of food and feed law. It sets out an overarching and coherent framework for the development of food and feed legislation both at union and national levels. To this end, it lays down general principles, requirements, and procedures that underpin decision-making in matters of food and feed safety, covering all stages of food and feed production and distribution. Food manufacturers, that is, tomato growers must make sure that their products are safe and do not endanger a human health. In North America, similar legislation is applicable, see Ref. [3]. Businesses involved with food must meet national safety and hygiene requirements in order to safeguard consumer health. So, the grocer again trusts the claim made by the tomato grower(s).

Clearly, there is a relationship between you buying tomatoes from the grocery store and the trust you must have in the measuring capabilities of the grocer using a scale and the quality of the tomatoes provided by the grocer trusting him or her that the tomatoes will not endanger your health. On a more abstract level, we face a network of interconnected entities like humans and objects like things interacting with each other. An ecosystem characterized as a communication and information exchange network in the world coined as social material systems (SMS) [4].

1.1 Research question and research approach

A quality control system (QCS) typically involves monitoring the performance of measuring instruments, standards of reference, and measurement information systems within the scope of the QCS. One might ask “How are ecosystems characterized as a communication and information exchange network, linked to a QCS so that we can trust the claim of suppliers making up the network?” In fact, this question is really about whether we can trust the data we use in the assertion of whether the claim of suppliers in a network is accurate. In this chapter, we propose a computational approach as a computational model to learn the inherent uncertainty to data integrity subsumed in the claim of stakeholders within and outside the organization. This research is in the realm of design science research and is to be characterized as design theory [5]. In this respect, this research coined as design relevant explanatory/predictive theory (DREPT) augments the “How” part or question with explanatory information on “Why” one should trust the proposed design will actually work. The key point is that the explanatory information is obtained using kernel theories. Kernel theories are established theories from social sciences, economics, mathematics, computer science, logic, and so on. We are interested in theory building on how to design effective and efficient governance and control systems, which may be interpreted as experimental scientific investigation. The ultimate unit of analysis is the individual coined as methodological individualism [6]. This chapter is outlined as follows. In chapter 2, we elaborate upon measurement, uncertainty, and how to model exchange

relationships mathematically. Next, we define what data integrity actually entails. In chapter 4, we elaborate on modeling decisions as evaluation procedures in auditing and quality control. The last two paragraphs put the procedure in the context of the strength of the audit evidence followed by conclusions.

2. Measurement and uncertainty

NIST defines measurement as an experimental or computational process that by comparing with a standard as a norm, produces an estimate of the true value of a property of a material, a (virtual) object, a collective of (virtual) objects, a process, an event, and a series of events, together with an evaluation of the uncertainty associated with that estimate and intended use in the support of decision making [7]. Measurement uncertainty concerns, that is, express the doubt about the true value of the measurand as the estimate of the true value of a property as defined after a measurement. The doubt relates to or is associated with the level of rigor to be determined on the level of uncertainty and what is needed to demonstrate its credibility that determines the adequacy to meet users' needs and wants. Most probably the adequacy is influenced by regulatory rules and regulations set by governmental bodies, such as governments, customers, demand, reputation of the company, ethical standards, and so on. This makes traceability to standards and assurance a complex endeavor to reach and maintain traceable performance standards.

Traceable to (SI) standards is not the same as counting objects. A claim that counts, as a result of a sample, are traceable to (SI) standards is not correct because it neglects the fact that counting inextricably involves the definition of what is being counted which definition is not a part of the (SI) standard, but when some characteristic of the object is measured then it might be possible that this particular measurement result is traceable to the (SI) standard. Making counting traceable to the (SI) standard is very important for economic life, one's health, one's security, and so on. The value lies in the precision of the measurement and therefore the measurement result. Put in other words: "Knowing the measurement uncertainty contributes to one's belief whether a measurement result as a count represents the quantity one has measured traced back to the (SI) standard."

Note that the noun phrase value like the noun clause traceability of measurement results has different meanings considering the situations in which the notions as concepts are used by humans interacting with artifacts, such as machines, software, and so on. This is why we must address the importance of measuring correctly when we evaluate the operational performance of some (organizational) entity that uses the measurement result for some purpose. Measuring correctly is inextricably related to data quality, but first, we have to model a standard that inextricably defines what is counted making it possible to make what is counted to be traceable to the (SI) standard.

2.1 Canonical model of exchange relationships

Exchanges are by definition reciprocal in nature and come in a large variety of what we coin as means like signed contracts, shaking hands, etc. For example, signing a contract by both parties is performative in nature; by the act of signing, we communicate that the exchange is done. Hence, a signed contract affords exchange. An affordance establishes the relationship between an object or an environment and an

organism here a (human) agent through a stimulus to perform an action. In our example, the stimulus is the signed contract and the detectable change in the external environment. We assume that the agent is sensitive and therefore able to respond to external (or internal) stimuli [8, 9].

Bilateral contracts are commonly used in business transactions. You buy 2 kilograms of tomatoes in our introductory example is a type of bilateral contract. The grocer promises to deliver the tomatoes and you promise to pay for the tomatoes by giving the grocer the indebted amount you have agreed upon when receiving the tomatoes. More formally, we can depict the canonical description of a value exchange cycle as in **Figure 1**. We use the following notation.

Notation 1.1: (Bilateral contract—canonical model), we will use the left and right harpoons exclusive for a bilateral contract between two agents $S \rightleftharpoons B$. Furthermore, actions are denoted as round-edged rectangles. Action nodes are connected via arrows that specify the control, that is, the information and communication flow. Together with the initial and the final node depicted as a solid circle and a solid circle surrounded by a hollow circle, we have a correct descriptive model of the value exchange cycle.

Note that money is exchanged for goods and or services. The exchange will actually occur in practice when parties agree upon a contract, that is, the transaction governance, the transaction structure, and the transaction contents, by the act of signing denoted by the initial node depicted as a solid circle. The contents reflect the objects of exchange. In our case tomatoes, now it is possible to extend the bilateral contract from a market point of view as depicted in **Figure 1** into a value exchange cycle from an organizational point of view. The final result is depicted in **Figure 2**.

Remark 1.2: (Bilateral contract—organizational view): Mark that the value exchange model described the sell-side of agent A and the buy-side of agent B. Now, it

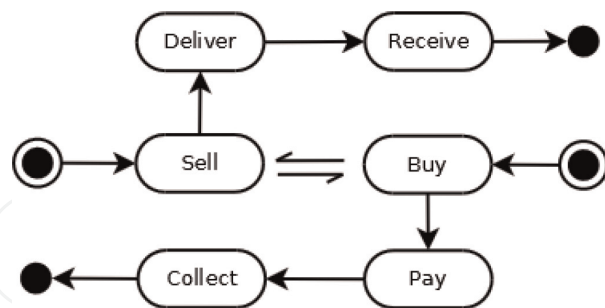


Figure 1.
Value cycle exchange.

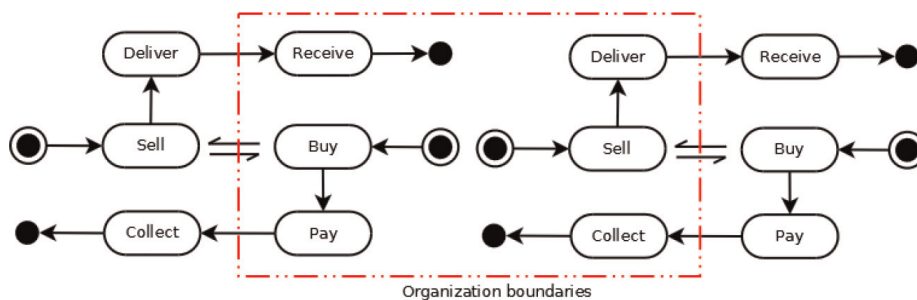


Figure 2.
Value exchange cycle double.

is easy to see that agent A as an organization must also have a buy-side otherwise he would not be able to deliver the ordered goods or services. The same type of reasoning does apply to buyer B who must have a sell-side otherwise or has enough budget to consume the goods or services. By simply doubling the model of the value exchange cycle (i.e., the bilateral contract—Marker view) we get the precise description of the value cycle of an organization in which organizational boundaries are denoted as the dashed line in red. For a more detailed exposition, we refer to Ref. [10].

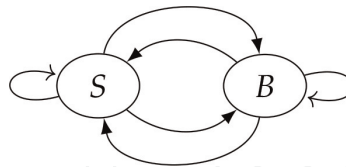
This concludes our informal description of bilateral contracts used in value exchange situations. We will see that under specific conditions the market view model is equivalent to the organizational view model. It is also easy to see that the organizational point of view is easily extended in a net(work) of contracts similar to supply chain models commonly used in logistics [11, 12].

2.1.1 Bilateral contract—Market graph

It is quite straightforward to translate the give-and-get relationship depicted in **Figure 1** in a mathematical graph. More specifically a give-and-get relationship is a directed graph. Formally a graph is defined as follows [13].

Definition 1.3: (Graph): A graph $G = (V, E)$ is a mathematical structure consisting of two finite sets V and E . The elements of V are vertices (or nodes), and the elements of E are the edges. Each edge has a set of one or two vertices associated with it, which are called endpoints. A formal specification of a general digraph $D = (V, E, \text{endpoints}, \text{head}, \text{and tail})$ is obtained from the formal specification of the underlying graph by adding the functions $\text{head}: E_G \rightarrow V_G$ and $\text{tail}: E_G \rightarrow V_G$, which designate the head vertex and the tail vertex of each arc.

Translating the value cycle exchange market view of the bilateral contract into a directed graph gives us the following result.



In a bilateral exchange relationship, money is exchanged for goods and or services. This is true from the buyers' perspective as well as from the sellers' point of view. We say that the proportion of goods and or services to money equals the proportion of money to the goods and or services. So, we get the following equality:

$$\frac{\text{Goods}}{\text{Money}} = \frac{\text{Money}}{\text{Goods}} \quad (1)$$

Let χ denote the goods and μ denote the money, so, we get:

$$\frac{\chi}{\mu} = \frac{\mu}{\chi} \quad (2)$$

Nodes S and B are in fact rationals, defined as follows [14]:

Definition 1.4 (Rational number): A rational number is an expression of the form $a//b$, where a and b are integers and $b \neq 0$; $a/0$ is not considered to be a rational number. Two rationals are considered to be equal, $a//b = c//d$, if and only if $ad = bc$.

Given the definition of a rational remark that money, goods and services are not equal objects, but that the exchange relationship itself is equal. We observe that

$$S = \frac{\chi}{\mu} \Rightarrow \frac{\chi}{\mu} \cdot \frac{\mu^2}{\chi^2} \Rightarrow \frac{\mu}{\chi} = B \quad (3)$$

and

$$B = \frac{\mu}{\chi} \Rightarrow \frac{\mu}{\chi} \cdot \frac{\chi^2}{\mu^2} \Rightarrow \frac{\chi}{\mu} = S \quad (4)$$

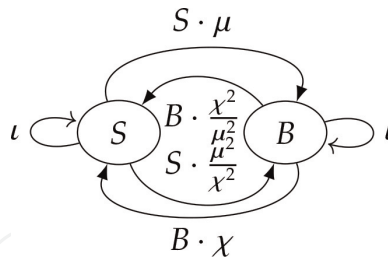
It follows that the following equality holds:

$$B \cdot S = \frac{\mu^2}{\chi^2} \cdot \frac{\chi^2}{\mu^2} \quad (5)$$

Remark 1.5 (Equality—bilinear): Equality (5) is not that easy to understand. For now, it suffices to state that the multiplication symbol as a connective is to be understood as a multiplicative $B \otimes S$ which is the bilinear version of and, dominated by the linear negation $(\cdot)^\perp$, which is a constructive and involutive negation defined in linear logic [15].

To be precise, the bilateral exchange relationship preserves the identity of the objects denoted as rationals. Consequently, S delivers χ , denoted as $S \cdot \mu$ and B pays the money μ , denoted as $B \cdot \chi$. Mark that ι denoted as a loop in the graph serves as an explicit precondition(s). Now, we can label the nodes and edges.

Remark 1.6 (Equality—linear): It is important to note that $S \cdot \mu$ and $B \cdot \chi$ are additives in linear logic, which is the linear version of and denoted as $S \& \mu$ and $B \& \chi$.



Returning to our example, suppose you have to pay 4 € for 2 kilograms of tomatoes. The grocer will hand you the 2 kilogram tomatoes expressed by $S \cdot \mu$. You pay 4 euros expressed by $B \cdot \chi$. Both actions will take place under the strict condition that you and the grocer agreed upon the contract denoted by ι and the unit tomatoes per euro preserving the identity expressed as: $S \cdot \frac{\mu^2}{\chi^2}$ and $B \cdot \frac{\chi^2}{\mu^2}$. Mind that S and B are rationals.

Up till now, our notions of goods, services, and money are in fact dimensionless. Parties will also have agreed upon the unit of measurement of the goods or services the seller will deliver and get paid for, respectively, the buyer will receive and is obliged to pay for the received goods or services from the seller. We will use the following notation.

Notation 1.7 (Units: measures and measurement): The quantity of the object O is measured in some standard unit expressed as a number and a reference denoted as superscript st and superscript m, the dimension quality denoted as (q) of object, and the dimension absolute frequency as a number of objects. Standard units expressed as

a number and a reference $Q_O^{st}Q_O^m$ can be denoted as $U_{(O_q)_s}$ for the sell-side and $U_{(O_q)_B}$ for the buy-side, where U denotes the standard unit expressed as a number and a reference. The quantity of the object O is measured in some standard unit U and the measurement is expressed as a product $Q \cdot U$, the dimension quality denoted as q of object, and the dimension absolute frequency as a number of objects.

We denoted χ for the goods and services and μ for money. For the sell-side we get:

$$Seller\chi := Q_{\chi_q}^S \cdot U_{\chi_q}^S \cdot U_{\chi}^S \quad (6)$$

$$Seller\mu := Q_{\mu_q}^S \cdot U_{\mu_q}^S \cdot U_{\mu}^S \quad (7)$$

For the buy-side we get:

$$Buyer\chi := Q_{\chi_q}^B \cdot U_{\chi_q}^B \cdot U_{\chi}^B \quad (8)$$

$$Buyer\mu := Q_{\mu_q}^B \cdot U_{\mu_q}^B \cdot U_{\mu}^B \quad (9)$$

We stated earlier that traceability to (SI) standards is not the same as counting objects. A claim that counts are traceable to (SI) standards is not correct in the case one neglects the fact that counting inextricably involves the definition of what is being counted which definition is not a part of the (SI) standard. The canonical model of the bilateral contract ensures that all characteristics of an object can be identified and thus be measured, so that the particular measurement results are by design traceable to the (SI) standards. Remark that money is considered as an abstract object alike goods and services. As we will see later on in this chapter it is this particular characteristic which is very convenient, that is, helpful, but first, we have to extend our model to fit the organizational view.

2.1.2 Bilateral contract: Organizational graph

From a business perspective, we have to translate the value cycle exchange market view of the bilateral contract into a directed graph representing the bilateral contract organizational view. To do so, we have to extend our definition for rational numbers for sum, product, negation, subtraction, and quotient.

Definition 1.8 (Rational number—sum, product, negation, subtraction, and quotient). If $a//b$ and $c//d$ are rational numbers, we define:

$$[sum](a//b) + (c//d) := (ad + bc) // (bd) \quad (10)$$

$$[Product](a//b) \cdot (c//d) := (ac) // (bd) \quad (11)$$

$$[Negation] - (a//b) := (-a) // b \quad (12)$$

$$[Subtraction](a//b) - (c//d) := (ad - bc) // (bd) \quad (13)$$

$$[Quotient]x/y := x \cdot y^{-1} \quad (14)$$

Next, there are basic properties of order on the rationals. Following Tao they are [14]:

Proposition 1.9 (Basic properties of order on the rationals): Let x, y, and z be rationals, then the following properties hold:

Laws 1.10 Order trichotomy. Exactly one of the three statements $x = y$, $x < y$ or $x > y$ is true.

Laws 1.11 Order is antisymmetric. One has $x < y$ if and only if $y > x$.

Laws 1.12 Order is transitive. If $x < y$ and $y < z$, then $x < z$.

Laws 1.13 Addition preserves order. If $x < y$, then $x + z < y + z$.

Laws 1.14 Positive multiplication preserves order. If $x < y$ and z is positive, then $xz < yz$.

Via law 1.10 order trichotomy, we know that it must be the case that exactly one of the three statements $x = y$, $x < y$ or $x > y$ is true. It follows that:

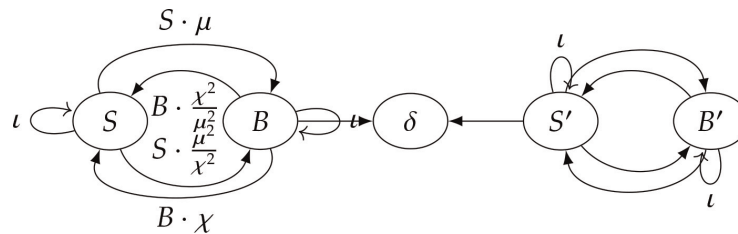
Laws 1.15: In the case $S=B$, then it must be the case that x equals y . In the case $S \neq B$, then it must be the case that $x < y$ or $x > y$.

Now, we introduce the notion of distance.

Definition 1.16 (Distance δ): Let x and y be rational numbers. The quantity $|x - y|$ is called the distance between x and y denoted as $d(x,y)$, thus $d(x,y) = |x-y|$.

It follows that $d(x,y) = 0$ if and only if $x = y$ and $d(x,y) \neq 0$ if and only if $x \neq y$.

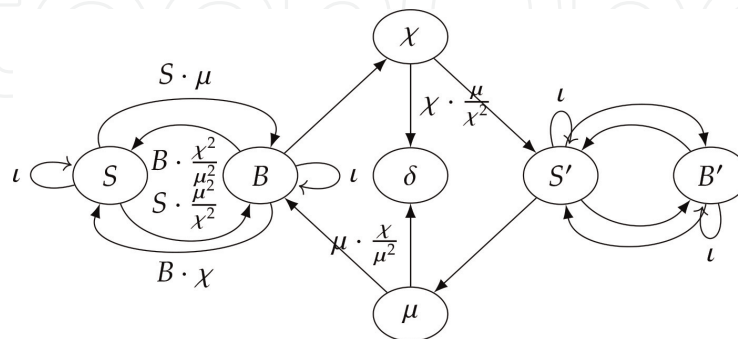
Translation of the value cycle exchange market view of the bilateral contract into a directed graph representing the bilateral contract organizational view we get the following result.



Subtraction of rationals is defined in eq. 13. When we apply subtraction of B and S' and take the absolute value, then we get the distance:

$$\left| \frac{\mu}{\chi} - \frac{\chi}{\mu} \right| = \left| \frac{\mu \cdot \mu - \chi \cdot \chi}{\chi \cdot \mu} \right| = \delta \tag{15}$$

Extending the graph gives us the following result.



Remark 1.17 (Equality - δ): To see that the extended graph—organizational view is equivalent to the canonical model of the bilateral contract—market view we take eq. 3, 4, and 15 into account. Eq. 15 gives the definition of δ :

$$\left| \frac{\mu}{\chi} - \frac{\chi}{\mu} \right| = \left| \frac{\mu \cdot \mu - \chi \cdot \chi}{\chi \cdot \mu} \right| = \delta \tag{16}$$

The formulas $\chi \cdot \frac{\mu}{\chi^2} \otimes \mu \cdot \frac{\chi}{\mu^2}$ can be rewritten by substituting χ by $S \cdot \mu$ and substituting μ by $B \cdot \chi$. We get:

$$S = S \cdot \mu \cdot \frac{\mu}{\chi^2} \Rightarrow S \cdot \frac{\mu^2}{\chi^2} \Rightarrow \frac{\chi}{\mu} \cdot \frac{\mu^2}{\chi^2} = \frac{\mu}{\chi} = B \quad (17)$$

and

$$B = B \cdot \chi \cdot \frac{\chi}{\mu^2} \Rightarrow B \cdot \frac{\chi^2}{\mu^2} \Rightarrow \frac{\mu}{\chi} \cdot \frac{\chi^2}{\mu^2} = \frac{\chi}{\mu} = S \quad (18)$$

Now, it is easy to see that both models—market vs. organizational view—are equivalent, that is, isomorphic.

When we interpret the graph, then it is easy to see that δ is only meaningful if and only if the units of measurement are identical. The following axioms must hold:

$$\text{Equality of units of measurement } \chi \quad U_{\chi_q}^S \cdot U_{\chi}^S = U_{\chi_q}^B \cdot U_{\chi}^B \quad (19)$$

$$\text{Equality of units of measurement } \mu \quad U_{\mu_q}^S \cdot U_{\mu}^S = U_{\mu_q}^B \cdot U_{\mu}^B \quad (20)$$

In the case, B and S' are the same agents as S and B', then $\delta = 0$. In the case, they are not the same agents then δ can have three values of which exactly one of the three statements $x = y$, $x < y$, or $x > y$ is true. It follows that when $x = y$ that the following laws hold:

$$\text{Equality} \quad B = S \quad (21)$$

$$\text{Equality} \quad S = B \quad (22)$$

If $x < y$ or $x > y$ is true, then the following equalities hold, respectively:

$$\text{Equality} \quad B + \delta = S \quad (23)$$

$$\text{Equality} \quad B = S + \delta \quad (24)$$

Remark that we are interested in the proportionality and not in the quotient arithmetically.

With this description, we are complete to elaborate on the notion and role of data quality in a rigorous way, but first, we have to elaborate on the notion of data integrity.

3. Data integrity

Making counting traceable to a (SI) standard contributes to the precision of the measurement itself and it determines whether the measurement result is acceptable to the adhered standards, implying that there exists a decision procedure as an evaluation procedure warranting that the standards are met.

A decision procedure as described assumes that the data quality is up to standards. To understand the concept of data quality, one needs to understand data

integrity. Data integrity in itself is defined as “the state that exists when data are unchanged from its source and has not been accidentally or maliciously modified, altered or destroyed” [16]. This view is consistent with the model proposed by Boritz in Ref. [17, 18] in which data integrity is subsumed in the notion of information integrity. Boritz defines information integrity as the representational faithfulness of information to the true state of the object that the information represents. His aim was to define and validate a general purpose framework that can be used for controlling and as well as for auditing purposes. In this way, information integrity impairments can be addressed in an organized and rigorous manner to guide management risk assessments and control deployment on the criteria to be addressed to attain reasonable assurance of whether information integrity objectives are met. Information integrity really concerns the validity and completeness aspects of the representation itself. Indeed, the object actually measured.

Boritz distinguishes (core) attributes from enablers, helping realize representational faithfulness. In his view, representational faithfulness is viewed as a degree of achievement of it rather than absolute quality. Practically it is all about accuracy/correctness, which has two dimensions viz. completeness on one side and validity on the other side. In the case, these dimensions are flawed, then it has negative consequences for the accuracy/correctness assertion. Obviously, there is a trade-off. Consequently, representational faithfulness is subject to some degree of imperfection, with the tolerable degree of imperfection being defined differently in different domains and contexts. In **Figure 3**, this trade-off relationship is depicted by the pointed arrows.

Now, it is quite logical how these core attributes help in realizing the representational faithfulness of information to the true state of the object that the information represents. From an user perspective, granularity enables understandability and relevance buttressing the decision-useful approach in decision-making. From a systems view is it essential that all data are available and accessible as enablers helping to warrant that the data are complete, current, and timely. From a data integrity perspective, security warrants as an enabler that the proper authorization is realized subsumed in validity. The attributes predictability, consistency, and neutrality preserve the informational quality as measurement. Neutrality warrants from this point of view that the information is free from biases, that is, neutrality preserves that objective standards are met. Verifiability as an enabler warrant the ability that independent observers, applying the same processes and tolerances for completeness, currency, timeliness, and validity that are used to produce the information, to replicate substantially the same result. Where auditability refers to the possibility to trace information back to its source and confirms the representational faithfulness of the information. It applies to all enablers that we design and implement controls to assure that the core attributes are fulfilled and therewith the representational faithfulness is attained. In **Figure 4**, we have extended **Figure 3** with the attributes which determine and influences the accuracy of the data.

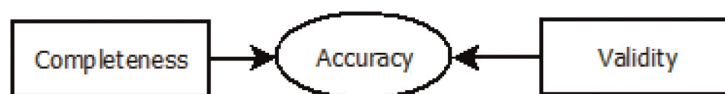


Figure 3.
Accuracy data.

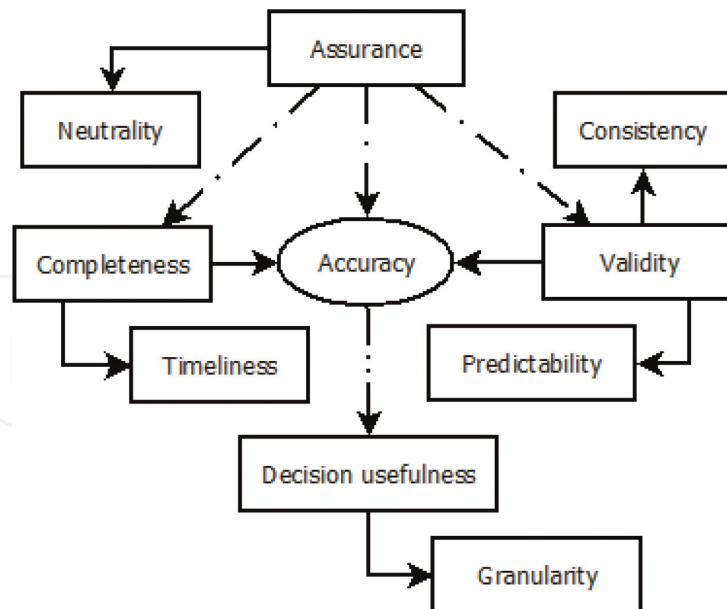


Figure 4.
Accuracy data and their properties.

4. Modeling the decision procedure as in auditing and quality control

In describing the decision procedure we use a pseudo-code format as in Ref. [19]. Our motivation is to reveal the algorithmic idea behind quality auditing and quality control practices within a business context but not limited to so. We happily leave it to the reader to make a final choice on how to implement algorithms using the program technology of his or her choice. It is important to be aware of what is to be considered as the object language. An object language is used to denote the language talked about for example formal expressions of propositional logic, linear logic, and so on. A metalanguage denotes the language in which we are talking about the object language, for example, a natural language augmented by a variety of common mathematical symbols. We think for the purpose of this chapter comprehensibility is preferred to the level of mathematical, that is, computational rigor required for implementing algorithms in some program technology running on some (preferred) hardware configuration part of a technological infrastructure coined as an information and communication network.

As we have stated in paragraph 2, measurement uncertainty expresses doubt about the true value of the measurement, as the estimate of the true value of a property as defined after a measurement. The doubt about the true value concerns the result of an evaluation of the uncertainty associated with the actual measurements compared with the estimated uncertainty and the intended use in the support of decision-making. In paragraph 2.2.1, we defined the notion of distance δ giving us the means to quantify the expected measurement result and to quantify simultaneously the actual impact of noncompliance by comparing the actual business outputs with expectations. This type of analysis is commonly known as variance analysis. Expectations are thought of as norms and predict normative behavior [20]. Norms in itself can be thought of as preconditions or postconditions that serve as conditionals in the determination of whether to accept the input conditions or to accept the output conditions. In the next sub-paragraphs we will elaborate on how to model an evaluation procedure to assess the data integrity. We define a process modeling language to translate the value exchange cycle as depicted in **Figure 2** to model the evaluation procedure to assess the data integrity of given data sets.

4.1 Modeling evaluation procedure data integrity

Accepting conditions presumes a decision procedure where inputs are compared with the norm(s) applicable to the input(s). So, we assume that there exists for any set S of formulas a valuation of S , which is a function v from S into the set t, f , where t denotes true and f denotes false coined as truth values. We say that X is true under v if $v(X) = t$, and false under v if $v(X) = f$. So, accepted inputs or outputs make up the truth set [21]. We give the following definition:

Definition 1.18 (Boolean valuation):

- B_1 : The formula $\neg X$ receives the value t if X receives the value f and f if X receives the value t .
- B_2 : The formula $X \wedge Y$ receives the value t if X, Y both receive the value t , otherwise $X \wedge Y$ receives the value f .
- B_3 : The formula $X \vee Y$ receives the value t if at least one of X, Y receives the value t , otherwise $X \vee Y$ receives the value f .
- B_4 : The formula $X \supset Y$ receives the value f if X, Y receives the respective values t, f otherwise $X \supset Y$ receives the value t .

By an interpretation of a formula X , we mean an assignment of truth values to all the variables which occur in X .

Proposition 1.19 (Accuracy): The data is accurate and is TRUE if and only if:

- (1) the data is VALID is TRUE is TRUE \wedge the data is COMPLETE is TRUE is TRUE.

Algorithm 1: Decision procedure interpretation

```

Input: Accuracy data model
1  ▷ see figure 3;
   Output: TRUTH set
   Data: data file
2  ▷ collection of attributes representing the object and its properties;
   Result: Assurance accuracy data
3  ▷ result of the decision procedure;
4  initialization: read data file ▷ one or more data files;
5  if the data is  $\neg$  VALID is TRUE  $\wedge$  the data is  $\neg$  COMPLETE is TRUE then
6  |   the data is  $\neg$  ACCURATE;
7  else
8  |   if the data is VALID is TRUE  $\wedge$  the data is COMPLETE is TRUE then
9  | |   the data is ACCURATE;
10 |   else
11 | |   the data is VALID is TRUE  $\vee$  the data is COMPLETE is TRUE;
12 |   end
13 end

```

This completes our description of the decision procedure. Now, we have to address the data. Data integrity are subsumed in the notion of information integrity coined as

the representational faithfulness of information to the true state of the object, measured and registered in an information system, that the data as information represents. Following Clark and Wilson [22], we recognize the notions of internal consistency and external consistency of the data produced by a system. The distinction is similar to the distinction between internal and external validity made in research methods. Suppose we have a well-managed computer system. Its specifications have been verified to be correct and the system itself has been tested and behaves according to its specifications. That means that when we enter data into the system that is valid, valid data will ensue (internal consistency). However, even in such a near-perfect system, there is nothing to ensure correspondence with reality (external consistency). In general, external consistency can only be ensured by a combination of organizational measures (segregation of duties, policies, and so on), procedural measures (e.g., processing controls and supervision), and physical measures (e.g., gates, fences, and use of IDs). These measures are basic and some authors, therefore, call these measures indispensable controls, because they must ensure external validity of the (quality) control and (quality) audit evidence. From a design point of view, the key questions we have to address are whether we can trust the data and can use the data. The question can we use the data really concerns the question of whether the data actually registered in the information system itself actually fits our information needs. This is the first step we have to consider and is directly related to the data file at the start of our decision procedure. In the case, the data file actually represents the data as information for decision purposes then it is useful to check whether the data file contains data that are valid and complete. To answer these questions, we extend our bilateral contract—organizational view, as depicted in **Figure 2** by introducing our process-model language.

4.2 Process-model language: Definition and meaning

For our purposes, we need a language to make sure that our reasoning is precise and most of all easy to use. There are numerous ways to model processes, techniques to choose from and methodologies to apply. For our purposes, it suffices to use UML (unified modeling language) because UML provides a common meta-model that formally defines the abstract syntax of all sorts of diagrams for modeling process behavior. The declarative meta-model is a very good alternative to grammar used to define formal languages. As we will see, this feature characteristic provides the possibility to reason in a correct way. In our exposition, we use activity diagrams to model process behavior. The next section is based on Ref. [23].

Actions describe the tasks that have to be performed in realizing a primary function to be viable [24]. An action stands for some transformation in the modeled system to be performed. The sequence in which the actions must be executed is the most fundamental control structure. As we have seen actions in our language are denoted as round-edged rectangles. The arrows between the action nodes are the activity edges which specify the control flow. Together with the initial and the final node depicted as a solid circle and a solid circle surrounded by a hollow circle we have a correct specification of the control flow (**Figure 5**).

The semantics is defined as a token flow that can also be used to refer to data and physical objects. The tokens are referred to as control tokens and as object tokens. Mind that actions can only start when tokens are available from the proceeding action or actions along the incoming edges. We say that tokens are consumed when action starts. Consequently, tokens are produced, that is, offered to the outgoing edges when



Figure 5.
Control flow.

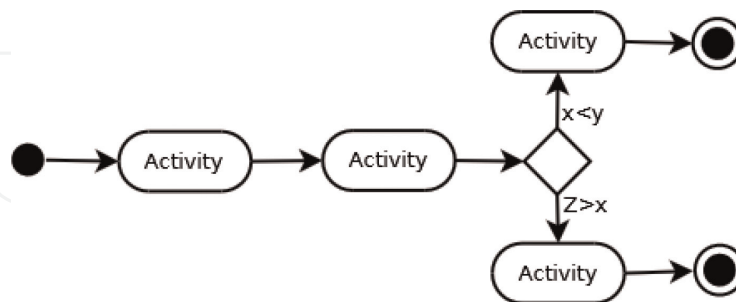


Figure 6.
Decision nodes and guards.

completed. In some circumstances, decisions have to be made for the choice of alternative control flows. Decision nodes are denoted as diamonds annotated by guards. The extended control flow can be depicted in **Figure 6**. Guards are logical expressions ending up to be true or false. Either we can state them in natural language, programming language constructs, or in formal mathematical logic. Guards can be refined as being pre and postconditions. When needed we will introduce them. The control logic remains the same. There are many more types of nodes used in modeling control flows, such as fork nodes, merge nodes, and join nodes. These types of nodes can be useful.

Finally, we have two types of nodes that are essential for our purposes. These are object nodes and data store nodes. Object nodes are needed to model the occurrence of objects at a particular moment or point in the process. Objects can be typed. We will extend this formalism extensively for our theory. To capture the object flow, the token flow semantics of activity diagrams is extended with object tokens. An object token behaves like a control token but it carries additionally a reference to a certain object type. Remark that we have to consider object type compatibility. This requirement is of utmost importance for our theory which we will see later in this chapter. A very convenient modeling notion is to use input pins and output pins which enables us to know which input and output parameters are assigned to various actions in the process. Pins are depicted as small hollow squares with their type written next to the square. In the case, we want to store information about orders, for example, than we can model such an action as a data store using data store nodes denoted as a rectangle. A data store node keep all tokens that enter it, copying them when they are chosen to move downward. See **Figure 7** for an example.

4.3 Modeling evaluation procedure data integrity extended

Now, we can extend our bilateral contract—organizational view to get a clear view about the informational needs and therewith next to it the requirements to meet a company's control and auditing objectives. The result is depicted in **Figure 8**.

Our objective is to assert whether the data stored as depicted in **Figure 8** can be considered to be accurate. More specifically these data stores enable us to extract one or more data files we need as input data in our decision, that is, evaluation procedure

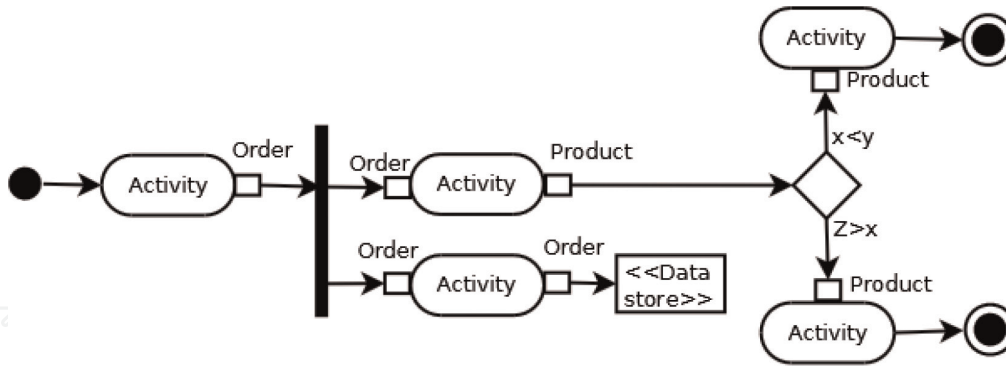


Figure 7.
 Pins and data store.

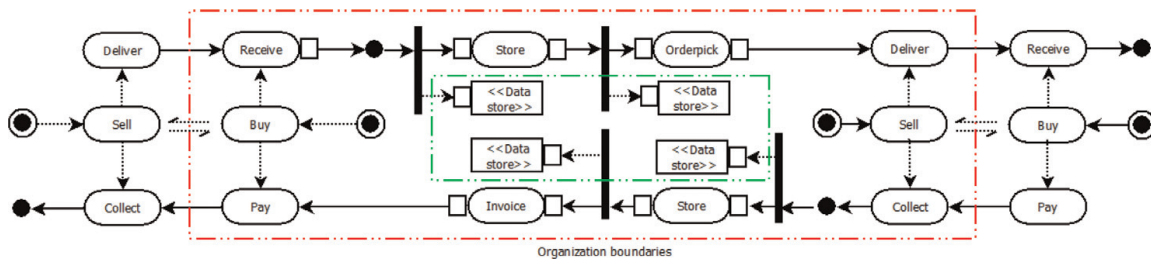


Figure 8.
 Control flow extended.

to assert the accuracy of the extracted data set(s) and its acceptability, that is, adequacy for quality control and quality audit purposes. In our example, we have identified data about stored goods, data about order-picked goods (to be) delivered, data about the collected revenues of the goods sold, and data about the actual payments of invoices received from suppliers for the goods we have received and stored in the warehouse. In general, a process stands for the behavioral pattern of an object, as far as it can be described in terms of the given named activities selected as its alphabet [25]. An alphabet denotes a permanent predefined property of an object. Remark that the name of an activity denotes an event class. There may be many events in a single event class named as an activity. Choosing an alphabet involves careful deliberation to decide which properties should be considered. A trace of the behavior of a process is defined as a finite sequence of symbols recording the event in which the process is engaged up to some moment in time. More formally:

Definition 1.20 (a trace is the sequence of symbols separated by commas closed by angular brackets):

- $\langle x, y \rangle$ denoting two events, x followed by y ,
- $\langle x \rangle$ denoting one sequent, containing only the event x ,
- $\langle \rangle$ denoting an empty sequence.

So, our extracted data set(s) from the data store(s) must contain all traces of the goods received from suppliers, all traces of the order picked goods (to be) delivered, all traces of the collected revenues goods sold to customers, and all traces of the paid invoices for the received goods from suppliers. It follows from the definition of a process that an alphabet defines the dimensions as column attributes giving us the names of all attributes making up the first row of the data set extracted from the data store(s) stored in the data file. The alphabet also gives us precise definitions of the object types and their properties. How do we proceed from here? Let us extend our running example.

4.3.1 An example

Suppose the organization we focus on is a trading organization specialized in tomatoes. On a daily basis, the organization buys the needed tomatoes at a local vegetable auction. The clients of the organization are retail organizations serving end customers. It is important to point out the fact that the organization has to comply with strict food safety regulations like Ref. [3]. For tomatoes quality indicators have been well established by total soluble solids measured by Brix-scale, dry matter, and acid contents. A Brix rating is important because it informs us about the quality of the tomato. The measurement is worked out on a scale based on 1° Brix denoted as °Bx, which is 1 g of sucrose per 100 g of solution. A low Brix rating indicates a nutrient deficiency. The Brix rating is used to measure the sweetness of tomatoes, but the rating is also linked to the acidity or PH level of the tomato. Tomatoes have on average a PH level between 4.3 PH and 4.9 PH on a scale of 0–14 PH. It is the combination of sweetness vs. acidity that gives the tomato its unique flavor. The Brix rating can be measured by using techniques labeled as NIR-spectroscopy, see Ref. [26]. From quality control and quality audit perspective, we need to know the unit(s) of measurement to determine whether the procured and sold tomatoes comply with quality standards to adhere for tomatoes. Relative density or specific gravity is defined as the ratio of the density (mass of a unit volume) of a substance to the density of a given reference material (substance). More formally:

$$RD = \frac{\rho_{substance}}{\rho_{reference}} \quad (25)$$

where RD denotes the relative density and ρ denotes density. So, a reference material is indicated as $RD_{substance/reference}$ which means the relative density of substance with respect to the reference. This description is equivalent to the notation and definitions introduced in paragraph. 2.1.1. Mind that mass and weight are separate quantities, they have different units of measure.

Let us assume that the organization bought 4000 kg of tomatoes and sold the 4000 kg to clients of the company. The company trades in one type of a large variety of tomatoes is the assortment. The buying price was €2,51 kg. The selling price was €2,63 kg. The agreed-upon contracts stipulates all sorts of requirements, including quality standards, applicable to the tomatoes. Parties agreed upon the acidity of the tomatoes must have a PH level between 4.3 PH and 4.5 PH on a scale of 0–14 PH and a sucrose RD of 9,993,325 °Bx.

4.3.2 Data integrity revisited

To assert whether the data stored as depicted in **Figure 8** can be considered to be accurate one needs to understand the objective of the evaluation procedure modeled in Algorithm 1: Decision procedure interpretation. The proposition is:

Proposition 1.21 (Accuracy): The data is accurate is TRUE denoted as T if and only if:

The data is VALID is T is $T \wedge$ the data is COMPLETE is T is T.

In paragraph 3, we elaborated on the concept of data integrity. In **Figure 4**, we depicted the accuracy data model and its key aspects that determine the accuracy of the data. There are three major aspects that determine the accuracy of the data and therefore its data integrity. These are:

- Consistency
- Predictability
- Timeliness

All other aspects are derived from notions necessary to trust the data and to strengthen one's belief that the information integrity is assured. Consistency has a variety of meanings, such as coherent, consistent, cohesive, connected, connective, sequacious, and so on. So, it is important to be specific about what is to be understood in the context of data accuracy. In this chapter, we choose a mathematical logical definition, which fits its purpose [21]. On the other aspects, we will elaborate in due course.

Definition 1.22 Consistency: A set X is called consistent if and only if for no finite subset Y of X at most one of A and A belongs to X , but not both. Meaning A cannot be both true and false.

As we can see, there is a strong relationship between the contract with the supplier and the purchase order of the goods here tomatoes. The contract specifies the conditions the organization and the supplier agreed upon. So, we have data about the price, quantity ordered, and quality norms applicable to the tomatoes. The same is true for the contract agreed upon with the customer and the sales order. Remark that next there is a strong relationship between ordering goods and money outflow due to paying the invoice. The same is true with respect to the sales of tomatoes and receiving the money. The type of controls to re-perform the relations are called reconciliation controls [27]. These types of controls follow directly from paragraph 2.1.2—extended graph bilateral contract—organizational view and **Figure 8** control flow extended.

4.3.3 The nature of controls: A classification

Before we extend our evaluation data integrity procedure, we have to elaborate on the nature of internal controls to be distinguished from processing controls, such as quality controls and quality audits. Internal controls are subsumed in processing controls as data integrity is subsumed in information integrity (**Table 1**).

As we see, there are two categories of controls making up five types of controls.

A. Access controls are what we coin as identity access controls (IAC) also known as segregation of duties controls. We prefer the term IAC. There are three elements

Type controls	Data integrity	Information integrity
Internal controls	Access controls	Accessibility controls
	Application controls	
	Reconciliation controls	
Processing controls	Availability controls	Usability
	Process logic controls	

Table 1.
Typing controls.

that buttress IAC. First identity control think of your user-ID. Secondly, there is authentication control, think of your password or passport. Thirdly, there is access control, think of authorization entering a theater or some office building where the porter lets you in. IAC enables an organization to safeguard assets or data of an organization. Remember, data integrity in itself is defined as “the state that exists when data are unchanged from its source and has not been accidentally or maliciously modified, altered or destroyed” [16].

- B. Application controls come in a large variety. The there main purpose is to enforce that data is entered in the correct way. Well-known examples are field check, sign check, limit check, range check, size check, completeness check, validity check, and closed loop verification. Application controls make sure that the right syntax is used and make it possible to implement business rules and constraints that fit the authorization level of an employee.
- C. Reconciliation controls are simply a comparison of the amounts that appear on the company’s balance sheet general ledger accounts to the details that make up those balances, while also ensuring that any differences between the two are adequately and reasonably explained.
- D. Availability controls are part of what is coined as information technology general controls (ITGC), which are the basic controls that can be applied to IT systems, such as applications, operating systems, databases, and supporting IT infrastructure. The objectives of ITGCs are to ensure the integrity of the data and processes that the systems support.
- E. Process logic controls are controls that determine whether the process is executed in the sequence that must be executed. For example, a procurement activity can not start in the case the contract with the supplier is not signed by an authorized employee.

When we map the data integrity controls onto the properties of data accuracy, we get the following result (**Table 2**).

Remark that the listed data integrity controls instantiate of what we have addressed as guards in our process-model language to be considered as preconditions

Type controls	Data integrity	Property	Completeness	Validity
Internal controls	Access controls	Consistency	X	X
	Application controls	Consistency		X
		Completeness	X	
	Reconciliation controls	Completeness	X	
Processing controls	Availability controls	Techn. data integrity	X	X
	Process logic controls	Timeliness	X	
	Process logic controls	Predictability		X

Table 2.
Typing controls.

and postconditions. Now, we are able to extend our evaluation data integrity procedure computationally.

4.4 Evaluation procedure assessment data integrity

We have seen that by typing controls in terms of internal controls and processing controls, we are able to clarify the property of the control subsumed in completeness and validity. We see that access controls and availability controls both underpin completeness and validity. Availability controls can be characterized as technical preconditions defining the types of attributes enabling us (human or machine) to register data in the preferred format, ensuring data integrity and data processing integrity, so no data get lost (in the information system). Considering the control flow depicted in **Figure 8**, and we analyze the data given in our example as given in paragraph 4.3.1, we come up with a specification of the attribute types specified in **Table 3**. We have listed the attributes and definition of its syntax, making up the alphabet as described earlier in this chapter. This completes our description of the alphabet we need.

The combination of unique number of ContractID, ActivityID, EmployeeID, RoleID, ProductID, and MachineID with the units °Bx, kg, €, and PH preserves the

Attribute definition	Description	Attribute definition	Description
ContractID ::= <integer>	Unique number	NetTotalInvoice ::= <00000,00>	Total invoice excl. VAT
AgentType ::= <integer>	Unique number	€	valuta r
AgentTypeDescription ::= <text>	Supplier, Buyer	TotalInvoice ::= < 00000,00>	Total invoice incl. VAT
NameAgentType ::= <text>	Description	MinAcid ::= <00,00>	Minimum PH
ActivityID ::= <integer>	Unique number	MaxAcid ::= <00,00>	Maximum PH
ActivityDescription ::= <text>	Buy, Receive, Sell, Deliver, Collect, Pay	PH	Unit PH
ActivityDate ::= <dd-mm-yyyy>	Date activity	Brix ::= < 000000000,00000 >	Brix ratio
SigDate ::= <dd-mm-yyyy>	Date signature	°Bx	Unit Brix
EmployeeID ::= <integer>	Unique number	QuantityWeighted ::= <00000,00>	Weighted kilos
NameEmployee ::= <text>	Name	Kg	Unit kilogram
RoleID ::= <integer>	Unique numberr	MachineID ::= <integer>	Unique number
RoleDescription ::= <text>	Description Role	MeasuredAcid ::= <00,00>	Real measured PH
ContractPrice ::= <00,00>	Decimal price	MeasuredBrix ::= < 0000,00000 >	Brix ratio real measure

Attribute definition	Description	Attribute definition	Description
€	Valuta	ExpAcitviityDate =:: <dd-mm-yyyy>	Expectation
VAT =:: <00,00>	Perinuage		
ProductID =:: <integer>	Unique number		
ProductDescription =:: <text>	Description		
Quantity=:: < 000000000,00>	Quantity		
Kg	Unit kilogram		

Table 3.
Attribute types description.

identity, which is elementary for data integrity in itself and the processing of data, so no data are lost. Put in other words. It is expected that the system is consistent. This notion as concept is fundamental to understand from a mathematical logical point of view but also to understand the notion of uncertainty.

Earlier we addressed that we make a distinction between the object language and metalanguage. In our pseudo code, the metalanguage is expressed as comments on the algorithm for its purpose. The logic is we have input data; we get output data for some purpose to be interpreted by a machine, human, or both. The algorithm specifies the rules fulfilling some computational task realizing the goal function [28].

4.4.1 Evaluation procedure: Availability

The first step is to create our alphabet in the database. The procedure describes the creation of the attributes in the reference model attribute database which serves as a reference to asses the data integrity of external data sets.

Algorithm 2: Create definition attribute types

```

[1]   Input: input data: definition attribute types table attribute types description
[3]   Output: output data: reference model definition attribute types
[5]   /* This procedure implements the necessary conditions data
      integrity definitions. Function assign variables as
      attributes write to database ▶ Reference model attributes
      database */
[6]   forall elements of Data file attributes do
[7]     | Create attribute variable in data base: write attribute variable;
[8]     | Define attribute type in data base: write attributes variable type;
[9]   end

```

The procedure describes the creation of the attributes in the reference model attribute database which serves as a reference to asses the data integrity of external data sets.

4.4.2 Evaluation procedure: Application controls

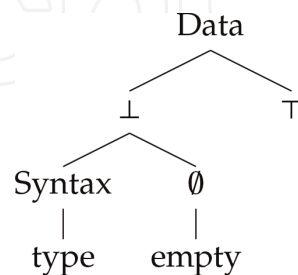
Next, we give the upload procedure data files for assessment reference attribute syntax in data file with reference to reference model definition attribute types.

Algorithm 3: Syntax data quality attributes of data files

```

[1]   Input: input data: definition attribute types table attribute types description
[3]   Output: output data: database contents data file input including interpretation
        syntax
[5]   /* This procedure implements the assessment of the data
        syntax quality in external data files write to database ▶
        Reference model attributes syntax - data file T or F    */
[6]   forall elements of external data file attributes do
[7]   |   Select first row and create header attribute variable in data base: write attribute
        variable;
[8]   |   Select second and write value of attribute to data base: write value variable
        type;
[10]  |   repeat
[11]  |   |   value of attributes of rows to data base: write value variable type ;
[12]  |   until last row;
[13]  |   if the attribute syntax data file = reference attribute syntax of reference is
        TRUE then
[14]  |   |   the data syntax of data file TRUE: write T in database;
[15]  |   else
[16]  |   |   the data syntax of data file  $\neg$  TRUE is T write F in database ;
[17]  |   end
[19]  |   return done
[20]  end
    
```

The result can be presented as a tree. On top, coined as the root, we see the data. Our algorithm checked the syntax of the data with the reference attributes as defined. The result is a clear insight per attribute into whether the syntax is correct or not. Hence, that empty attributes are distinguished from wrong types.



Remark 23 (Granularity): Remember that syntax type information is stored, so the syntax type can be specified in the details of the reference attribute types.

4.4.3 Evaluation procedure process logic controls: Some examples

The next step is to verify whether the process is executed as expected.

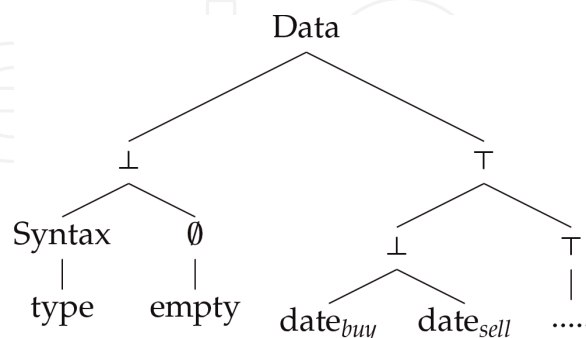
Algorithm 4: Process logic quality timeliness and predictability of processes

```

[1]   Input: input data: database contents data file input including interpretation syntax
[3]   Output: output data: database contents data file + interpretation syntax +
        interpretation process logic rules
[5]   /* This procedure assesses whether the process is executed in
        the sequence that should be executed. ▶ Reference model
        attributes values - data file T or F */
[6]   forall elements of external data files attributes do
[7]     Bundle data files: write values to data base bundle;
[8]     Select second row of all data form data bundle;
[10]    repeat
[11]      | write value interpretation variable type constraints ;
[12]    until last row;
[13]    if Date buy < Date receive is TRUE then
[14]      | Date buy of data file is TRUE: write T in database;
[15]    else
[16]      | Date buy of data file  $\neg$  TRUE is T write F in database ;
[17]    end
[18]    if Date sell < date deliver then
[19]      | Date sell of data file is TRUE: write T in database;
[20]    else
[21]      | Date sell of data file  $\neg$  TRUE is T write F in database ;
[22]    end
[24]    return done
[25]  end

```

As we have seen, we can present the result as a tree. On top, coined as the root, we see the data. Our algorithm checked the syntax of the data with the reference attributes as defined. Now, you see that on the right-hand side, some dates of buying and selling transaction are not timely.



Remark 24 (Granularity): Remember that the date buy and the date sell can be specified in the details of the reference attribute types.

4.4.4 Evaluation procedure application and reconciliation

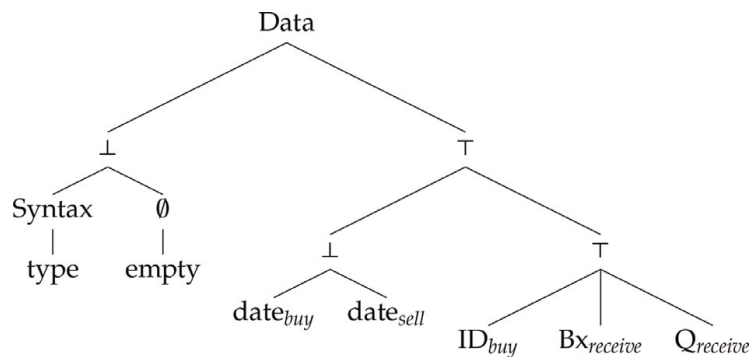
Now, we start to look at the content matter of the data.

Algorithm 5: Contents of the process is complete and consistent

```

[1]      Input: input data:database contents data file + interpretation syntax +
           interpretation process logic rules
[3]      Output: output data: database contents data file + interpretation syntax +
           interpretation process logic rules + reconciliation
[5]      /* This procedure assesses whether the contents of the
           process is complete and consistent. ▶ Reference model
           attributes values - data file T or F */
[6]      forall elements of external data files attributes do
[7]      Bundle data files: write values to data base bundle;
[8]      Select second row of all data form data bundle;
[10]     repeat
[11]     | write value interpretation variable type constraints ;
[12]     until last row;
[13]     if ContractID type buy = ContractID type receive is TRUE then
[14]     | ContractID of type buy of data file is TRUE: write T in database;
[15]     else
[16]     | ContractID of type buy of data file  $\neg$  TRUE is T write F in database ;
[17]     end
[18]     if MeasuredBrix type receive = Brix type buy then
[19]     | MeasuredBrix type receive of data file is TRUE: write T in database;
[20]     else
[21]     | MeasuredBrix type receive of data file  $\neg$  TRUE is T write F in database ;
[22]     end
[23]     if Quantity Unit type kg Type received = Quantity Unit type kg Type buy then
[24]     | Quantity Unit type kg Type received of data file is TRUE: write T in
           database;
[25]     else
[26]     | Quantity Unit type kg Type received of data file  $\neg$  TRUE is T write F in
           database ;
[27]     end
[29]     return done
[30]     end
    
```

When we look at tree result, then we see that the interpretation of the algorithm result gives us, on the right-hand side, the truth conditions about the unique ID, the unique measure of the quality Bx and the quantity received from the supplier.



Remark 25 (Model consistency): Remember that all transactions in the database which do not have the interpretation of being TRUE, under the conditions specified in our algorithm, are \neg TRUE registered in the database as F. So, we can switch if we are interested in the counterpart of the data set under consideration. This can be understood as a direct result of the compactness Theorem for ordered trees using König's Lemma see Ref. [21]. Mind that for unordered trees we need the axiom of choice.

5. Evaluation procedure data integrity

As we stated in paragraph 4.1, the accuracy of a data file under consideration is said to be accurate when the following proposition hold:

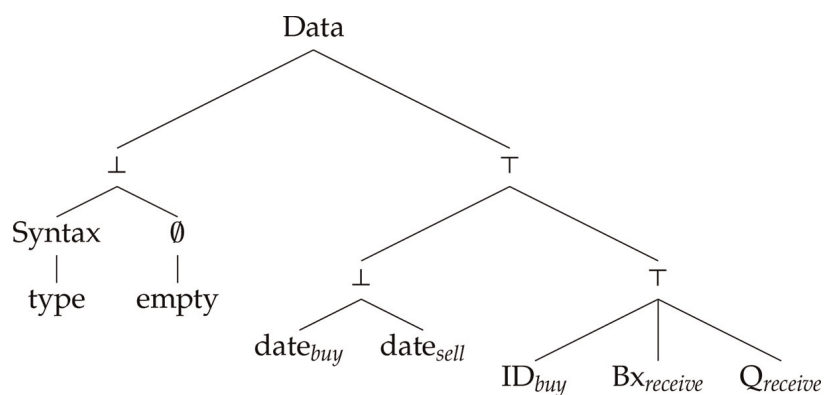
Proposition 26 (Accuracy): The data is accurate is TRUE if and only if:

(1) the data is VALID is TRUE is TRUE \wedge the data is COMPLETE is TRUE is TRUE.

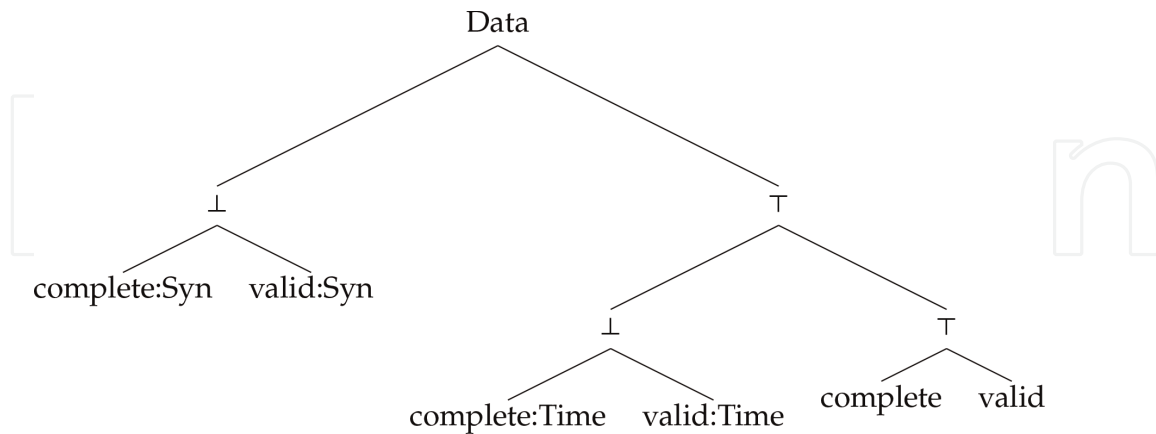
Algorithm 6: Decision procedure interpretation

Input: Accuracy data model
 1 \triangleright see figure 3;
Output: TRUTH set
Data: data file
 2 \triangleright collection of attributes representing the object and its properties;
Result: Assurance accuracy data
 3 \triangleright result of the decision procedure;
 4 initialization: read data file \triangleright one or more data files;
 5 **if** the data is \neg VALID is TRUE \wedge the data is \neg COMPLETE is TRUE **then**
 6 | the data is \neg ACCURATE;
 7 **else**
 8 | **if** the data is VALID is TRUE \wedge the data is COMPLETE is TRUE **then**
 9 | | the data is ACCURATE;
 10 | **else**
 11 | | the data is VALID is TRUE \vee the data is COMPLETE is TRUE;
 12 | **end**
 13 **end**

The result of our algorithm can be depicted as:



The types of controls give us the information about the property of some control preserving the data integrity. So, the outcome of our algorithms can be mapped onto our decision procedure. The result can be depicted as a tree:



The proof of our proposition is to be found in the mathematical theory as developed in chapter 2, the application of propositional Linear Logic [15] and analytical tableaux [21]. For now, we have decided to leave this formal proof out of this chapter for clarity reasons and left it for future research.

6. Re-performance as evidence in quality control audits

In general, there are six ways of obtaining audit evidence: (1) inspection, (2) external confirmation, (3) observation, (4) re-performance, (5) analytical procedures and (6) inquiry [29], see also the ISA 500 standard on audit evidence [30]. Re-performance refers to the practice where the auditor makes essential calculations and verification are again based on raw evidence. Automated forms of control, such as controls built into business processes, are more difficult to manipulate and can in principle cover the whole relevant population, not just a sample. These various ways of obtaining audit evidence can be ranked in a kind of hierarchy of evidence reliability. Inspection (1), external confirmation (2), and re-performance (4) are considered stronger because they produce relatively direct forms of evidence without the interference of the auditee, whereas observation (3), analysis (5), and inquiry (6) are considered relatively weaker depending on the sources (human or automated), expectations, procedures, and audit planning. Note, moreover, that evidence collection types (1), (2), and (4) are also the most time-consuming for the auditor and therefore the most expensive for the client. Audit fees are born by the company being audited and make up a large part of the costs of control [31]. Our decision procedure combines two strong forms of obtaining audit evidence. These two forms are external confirmation and re-performance. The procedure fits in the current modern computational idea data-driven assurance, which is consistent with quality 4.0 concepts in quality control and quality audit practices. The computational approach as developed in this chapter combines the logic of product and process audits, which ensures that the uncertainty inherent to data integrity can be known as a distribution. It follows from the computational approach that auditors can apply dual-purpose testing which

fits a data-driven, that is, fact-driven approach to decision-making of management and the stakeholders of organizations.

7. Conclusions

In this chapter, we have proposed a computational approach as a computational model to learn the inherent uncertainty to data integrity subsumed in a claim or claims made by stakeholders inside or outside the organization. Knowing the measurement uncertainty contributes to one's belief whether the measurement result as a count represents the quantity one has measured traced back to (SI) standards. Our computational model makes counting objects, persons, buildings, and so on traceable to the (SI) standard. The novelty in our approach is that the notion of equality has two different properties as being bilinear and linear. Our canonical model of the bilateral contract ensures that all characteristics of an object can be uniquely identified and thus be measured so that the particular measurement results are by design tractable to the (SI) standards. The result of our evaluation procedure of the data integrity is in fact an ordered dyadic tree whose presentation is understandable by humans and gives good insights into where to start the audit investigations in QCS.

Author details


Rob Christiaanse^{1,2}

1 Technical University, Delft, Netherlands

2 Free University, Amsterdam, Netherlands

*Address all correspondence to: r.christiaanse@efco-solutions.nl

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Tambare M, Lee R, Imoize. Performance measurement system and quality Management in Data-Driven Industry 4.0: A review. *Sensors*. 2021;**22**: 24. DOI: 10.3390/s22010224
- [2] Possolo B, Watters jr. Metrological traceability. Frequently Asked Questions and NIST Policy. National Institute of Standards and Technology. Technical Note. 2021;**2156**:42. DOI: 10.6028/NIST.TN.2156
- [3] Commodity Specific Food Safety Guidelines for the Fresh Tomato Supply Chain. 3rd Ed. NATTWG. United Fresh Produce Association, 2018. p. 50
- [4] Contractor N, Monge P, Leonardi PM. Network theory, multidimensional networks and the dynamics of Sociomateriality: Bringing technology inside the network. *International Journal of Communication*. 2011;**5**:39
- [5] Kuechler W, Vaishnavi V. A framework for theory development in design science research: Multiple perspective. *Journal of the Association for Information System*. 2012;**13**:6
- [6] Arrow KJ. Methodological individualism and social knowledge. *The American Economic Review*. 1994; **84**(2):1-9
- [7] National Institute of Standards and Technology (NIST) Quality Manual for Measurement Services. NIST-QM-I. version 11. 2019. p. 85
- [8] Rob Christiaanse. Modeling norms embedded in society: Ethics and sensitive design. In: *IfCoLoG Journal of Logics and their Applications* 5, 2 (April 2018). 2018. pp. 591–628
- [9] Christiaanse R. Human centric Design in Smartcity Technologies: Implications for the governance, control and performance evaluation of mobility ecosystems. In: *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion)*, April 25–29, 2022, Virtual Event, Lyon, France. New York, NY, USA: ACM; 2022. p. 10. DOI: 10.1145/3487553.3524847
- [10] Christiaanse R. Mobility as a service: A valuecentric approach to design. In: *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion)*, May 13–17, 2019, San Francisco, CA, USA. New York, NY, USA: ACM; 2019. p. 10. DOI: 10.1145/3308560.3317050
- [11] Hulstijn J, Overbeek S, Aldewereld H, Christiaanse R. Integrity of supply chain visibility: Linking information to the physical world. In: Bajec M, Eder J, editors. *Advanced Information Systems Engineering Workshops. CAiSE 2012. Lecture Notes in Business Information Processing*. Vol. 112. Berlin, Heidelberg: Springer; 2012. DOI: 10.1007/978-3-642-31069-029
- [12] Veenstra AW, Hulstijn J, Christiaanse R, Tan YH. Information exchange in global logistics chains: an application for model-based auditing. In: *WCO, Proceedings of the 8th annual PICARD Conference*. St. Petersburg, Russia. 2013. p. 13
- [13] Jonathan L. Gross and Jay Yellen. *Graph Theory and its Applications*. New York: CRC press, Taylor & Francis Group; 2005. p. 567
- [14] Tao T. *Analysis 1*. 3rd. Vol. 1. New Delhi: Hindustan Book Agency; 2017

- [15] Girard J-Y. Linear logic. *Theoretical Computer Science*. 1987;**50**(1):1-101. ISSN 0304-3975. DOI: 10.1016/0304-3975(87)90045-4
- [16] Welke SM, Mayfield WT, Roskos JE. Integrity and Information Protection, Report of the International Workshop on Data Integrity, NIST Special Publication 500-168. Gaithersburg: National Institute of Standards and Technology; 1989
- [17] Boritz JE. IS practitioners' views on core concepts of information integrity. *International Journal of Accounting Information Systems*. 2005;**6**(4):260-279
- [18] ITGI. *Managing Enterprise Information Integrity: Security, Control and Audit Issues*. Schaumburg: IT Governance Institute; 2004
- [19] Smed J, Hakonen H. *Algorithms and Networking for Computer Games*. 2nd ed. Hoboken, USA: Wiley Publishing; 2017
- [20] Griffioen P, Christiaanse R, Hulstijn J. Controlling production variances in complex business processes. In: Cerone A, Roveri M, editors. *Software Engineering and Formal Methods*. SEFM 2017. Lecture Notes in Computer Science. Vol. 10729. Cham: Springer; 2018. DOI: 10.1007/978-3-319-74781-1-6
- [21] Smullyan RM. *First-Order Logic*. New York [Etc.]: Dover publications; 1995
- [22] Clark, Wilson. A comparison of commercial and military computer security policies. In: *Proceedings of the 1987 IEEE Symposium on Research in Security and Privacy (SP'87)*, Oakland, CA. 1987. pp. 184-194
- [23] Dumas, Aalst van der, Hofstede. *Process-Aware Information Systems*. Hoboken, USA: John Wiley an Sons; 2005. 13 978-0-471-66306-5
- [24] Achterberg J, Vriens D. *Organizations Social Systems Conducting Experiments*. 2nd ed. Berlin: Springer Heidelberg; 2010. 978-3-642-14315-1
- [25] Hoare CAR. *Communicating Sequential Processes*. Series in Computer Science. Hoboken, USA: Prentice Hall publishers; 1985
- [26] NIR Calibration Model. <https://calibrationmodel.com/potential-usage-of-nir-analysis-and-its-industry-fields-of-applications-386/>
- [27] Romney MB, Steinbart PJ. *Accounting Information Systems*. Ed. 14 ed. London: Pearson; 2016
- [28] Hurwicz L, Reiter S. *Designing Economic Mechanisms*. New York: Cambridge University Press; 2006
- [29] Knechel W, Salterio S, Ballou B. *Auditing: Assurance and Risk*. 3rd ed. Cincinnati: Thomson Learning; 2007
- [30] International Standard on auditing. *IAASB Handbook*. 500 Audit evidence. 2020. p. 16
- [31] Christiaanse R, Hulstijn J. Control automation to reduce costs of control. *IJISMD*. 2013;**4**(4):27-47. DOI: 10.4018/ijismd.2013100102