Appendix E
Quality Improvement and Data Quality

The best laid schemes o' mice and men
Gang aft a-gley;
And leave us naught but grief and pain
For promised joy
—Robert Burns, "To a Mouse" (1785)

Purpose

This appendix provides a context for approaching data quality strategy. It presents background on thought leaders in quality improvement and discusses how concepts associated with quality improvement have been used effectively to improve data quality. It also discusses the limitations of these concepts for data and asserts the need to approach data quality strategy from the perspective of knowledge management as well as product management.

A Brief History of Quality Improvement

Most approaches to data quality improvement start with a comparison between the production of data and the production of manufactured goods.[1] Data quality practices are rooted largely in product quality practices that originated in the manufacturing sector beginning in the United States the mid-1920s and blossoming more fully in Japan after the Second World War. These practices are well known in their modern forms as Total Quality Management and Six Sigma. Such approaches draw from the work of thought leaders in quality management, including Walter Shewhart (1891–1967), W. Edwards Deming (1900–1993), Joseph Juran (1904–2008), Kaoru Ishikawa (1915–1989), and Philip Crosby (1926–2001). Each of these leaders contributed specific tools and methodologies for improving manufacturing quality (See Table E.1). The intention of this brief history of quality improvement is to provide a context to those who may be unfamiliar with the background and to show how these ideas provide a foundation for data quality strategy.

Importantly for a discussion of strategy, these pioneers all recognized that quality products do not make themselves. They are created by people, using input of lesser or greater quality, through processes that can be executed to a greater or lesser degree of success, in order to produce output that satisfies customers to a lesser or greater degree. The quality of manufactured products depends on

[1] Wang (1998), English (1999), Redman (2001), Loshin (2001), and McGilvray (2008). See Pierce (2004) for an overview of literature related to the concept of data as a product.

Table E.1 Summary of Quality Pioneers and Their Contributions

Name	Major Contributions
Walter Shewhart	Control chart Statistical process control Process predictability Plan-do-check-act cycle *Economic Control of Quality of Manufactured Product* (1931) *Statistical Method from the Viewpoint of Quality Control* (1939)
W. Edwards Deming	System of Profound Knowledge Fourteen Points Achieving quality requires changing how organizations are managed *Out of the Crisis* (1986) *The New Economics for Industry, Government, Education* (1993)
Joseph Juran	Pareto principle (80/20 rule) Quality products must be free from defects and possess desired characteristics; Achieving quality requires overcoming cultural resistance to change *Quality Control Handbook* *Managerial Breakthrough* (1964) *Management of Quality Control* (1967) *Quality Planning and Analysis* (1970) *Upper Management and Quality* (1980) *Juran on Planning for Quality* (1988)
Kaoru Ishikawa	Cause/effect (fishbone) diagrams Quality circles Total quality management Six (now seven) M's: machine, method, manpower/mind power, materials, milieu, and measurement, management *QC Circle Koryo* (1970; translated to English 1980) *How to Operate QC Circle Activities* (1980) *What Is Total Quality Control? The Japanese Way* (1981; translated to English 1985)
Philip Crosby	Zero defects Do it right the first time The measurement of quality is the price of non-conformance *Quality Is Free* (1979) *Quality Without Tears* (1984) *The Eternally Successful Organization* (1988) *Quality Is Still Free: Making Quality Certain in Uncertain Times* (1996).

clearly defined and well-tested processes. These depend on the commitment of an enterprise to establishing quality processes and enforcing the standards on which they rely. The same can be said for the quality of data and information products.

Walter Shewhart

Dr. Walter Shewhart invented the control chart and introduced the concept of statistical process control and is usually recognized as the founder of manufacturing quality control practices. Shewhart's approach was to make processes more predictable through process analysis, measurement, the elimination of special cause variation, and the implementation of ongoing monitoring. Trained as a scientist, Shewhart also introduced the Plan-Do-Study-Act approach (PDSA), based on the scientific method (hypothesize, experiment, and evaluate), for process and product improvement. PDSA, known both as the Shewhart Cycle for its inventor and the Deming Cycle for its popularizer, has been adopted and has evolved through other methodologies. Among its descendants is Six Sigma's Define, Measure, Analyze, Improve, Control (DMAIC) approach.[2]

Shewhart's *Economic Control of Quality of Manufactured Product* (1931) and *Statistical Method from the Viewpoint of Quality Control* (1939) are founding documents of the quality control movement. Educated in both engineering and physics, Shewhart had a long and distinguished career as an engineer at Western Electric and Bell Telephone laboratories, and as an academic and consultant to the United Nations and several national governments.

W. Edwards Deming

Dr. Deming, a colleague of Shewhart's at Bell Laboratories, went on to apply his ideas about quality in post–World War II Japan and to develop a comprehensive approach to management of organizations—his System of Profound Knowledge, whose central concepts were articulated through his famous Fourteen Points. Deming's Fourteen Points have been translated for data and information quality by both Redman (1996, pp. 65–66) and English (1999, 337–399). Incredibly influential in Japan, Deming's approach was not recognized in the United States for several decades. His work with Ford Motor Company, which focused on improving management practices, turned the automobile manufacturer around in the mid-1980s.

Like Shewhart, Deming was an engineer and physicist and had a distinguished career in industry, government, academia, and consulting. He published several books on statistical process control in the 1940s–1960s. His two most famous works, *Out of the Crisis* (1986) and *The New Economics for Industry, Government, Education* (1993), were published late in his career.

Joseph Juran

Joseph Juran is best known for applying the work of Vilfredo Pareto and popularizing the Pareto principle (the 80/20 rule: for many events, 80% of the effects result from 20% of the causes), and for the "Juran Trilogy," an approach to quality that includes quality planning, quality control, and quality improvement. He formulated a simple but powerful description of quality: Quality products must be free from defects and possess desired characteristics.

Like Deming, though independently from him, Juran applied his ideas in post–World War II Japan. Having published his *Quality Control Handbook* in 1951, he began teaching management principles to Japanese businessmen in the mid-1950s. Influenced by anthropologist Margaret Mead, he

[2] The original formulation is: Plan, Do, Check, Act. I prefer "study" because of Deming's observation that Western audiences might misunderstand the meaning of "check" and because the purpose of the process is to build knowledge (hence the need for "study") (Moen and Norman, 2012). The American Society for Quality uses "study." Variations include Plan, Do, Check, Adjust and Standardize, Do, Study, Adjust.

focused on managing for quality and overcoming cultural resistance to change. He served as a liaison between Japan and the United States, promoting quality control in Japan and the Japanese concept of quality circles—developed by Kaoru Ishikawa—in the United States.

Active in the American Society for Quality Control (later the ASQ), Juran had a long career as a consultant in industry and government. In addition to the *Quality Control Handbook* (the fifth edition came out in 1995), he published *Managerial Breakthrough* (1964), *Management of Quality Control* (1967), *Quality Planning and Analysis* (1970), *Upper Management and Quality* (1980), and *Juran on Planning for Quality* (1988).

Kaoru Ishikawa

Dr. Kaoru Ishikawa, a Japanese engineer and scientist, was instrumental in both applying and expanding management and quality concepts from Deming and Juran. He also made significant original contributions to quality control, including the quality circle and the Ishikawa (or fishbone) diagram. Quality circles are groups of employees who are trained to look for, analyze, and propose solutions to work-related problems. Their purpose is to seek out opportunities for improvement—not only in product quality, but also in other areas, such as health and safety—and also to humanize the work environment. Ishikawa was critical in integrating quality circles into Nippon Telephone and Telegraph's Total Quality Management program.

Developed in the early 1960s, Ishikawa's cause and effect diagram is a standard tool of quality control. The diagram classifies factors that contribute to general problems in order to assess how they affect a specific problem. Depending on the variation of the fishbone used, these general factors go by different names. In manufacturing they are referred to as the six M's: machine (equipment, technology), method (process), manpower/mind power (people, working either physically or mentally), materials (raw materials or information), milieu (environment, surroundings), and measurement (inspection). Most analysts also add a seventh: management. For each category, a set of questions enables analysts to drill in to identify factors contributing to problems.

Ishikawa's career also spanned industry and academia. He wrote two books on quality circles *QC Circle Koryo* (1970; translated to English 1980) and *How to Operate QC Circle Activities* (1980). A thought leader in total quality management, he also published *What Is Total Quality Control? The Japanese Way* (1981; translated to English 1985). He was also recognized by the Japanese government for his contributions to the improvement of Japanese industry.

Philip Crosby

Philip Crosby, an influential American business consultant, introduced the concepts of zero defects and doing it right the first time. He simplified the language of quality by asserting that quality is defined by conformance to requirements. With clear statement requirements, he said, defects could be prevented—all defects. He promoted the concept of zero defects. His most famous book, *Quality Is Free* (1979), asserted that the measurement of quality was the price of nonconformance, and he helped launch a "quality revolution" in the 1980s. Crosby published more than a dozen books, including *Quality Without Tears* (1984), *The Eternally Successful Organization* (1988), and *Quality Is Still Free: Making Quality Certain in Uncertain Times* (1996).

Process Improvement Tools

Methods to produce quality goods have resulted in a range of tools and techniques that can be applied to data quality. *The Quality Toolbox*, a reference book published by ASQ, describes more than 140 tools and techniques that can be applied to all phases of the improvement process. A small subset presented here is essential for data quality improvement because they enable detailed process analysis.

Process Flowcharts

A flowchart is a picture of the steps in a process that allows you to see their sequence and the relationships between them. At its most basic, a flowchart consists of boxes that represent steps in the process and arrows that show how they are connected. Detailed flowcharts should also contain other signifiers, such as decision points in the process, inputs and outputs, delays, and stop points.[3]

It is possible (and probably also necessary) to describe processes in words, but it is also helpful to represent them in a picture because doing so surfaces characteristics of the process in a way that many people can follow. For example, a process flow can be used to depict process steps at a similar level of detail, so by laying out a process flow, teams can often identify missing or overlooked steps. Often the process of creating a flowchart will itself take several steps. It helps to depict the process first at a high level and then drill into the individual steps.

Creating a process flow is also a way of clarifying problems to be addressed (see Figure E.1). A finished process flow can be used to define the scope of the improvements. It is also necessary

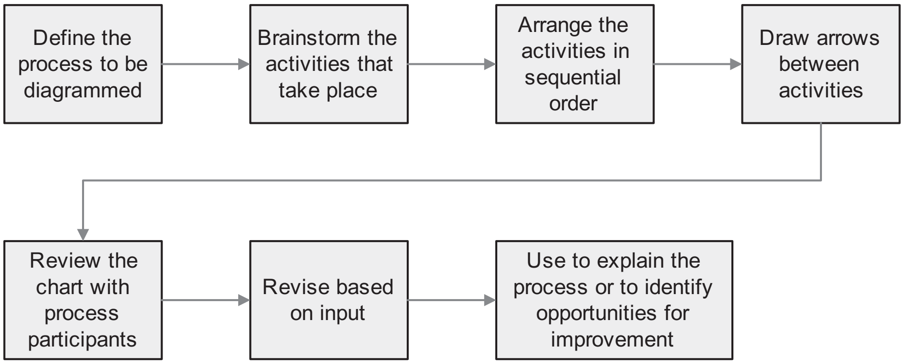

FIGURE E.1 How to Create a Flowchart

The figure outlines the steps required to produce a flowchart. Flowcharts enable an understanding of existing processes. They also can be used to define process improvements. Steps include defining the process to be diagrammed, brainstorming about the activities involved, and arranging them in sequential order. Once a chart is drafted, it can be used to get feedback and then revised. After the organization that needs a process documented has come to a consensus about the chart, it can be used for training or process improvement, or both.

[3] A visual vocabulary has been developed to create flowcharts. This vocabulary has the benefit of packing in a lot of information, but, unfortunately, since many people are not familiar with it, some of the significance of particular shapes can be missed. In most cases, especially in the early phases of understanding a process, simpler is better.

input for other types of analysis (e.g., cost of quality analysis). For data quality purposes, process flows are a form of metadata. They can be used to depict data lineage or the information life cycle. In this respect, they are very useful for people trying to make decisions about which data they should use for which purposes.

SIPOC: Supplier, Input, Process, Output, Customer

A SIPOC diagram is a specialized flowchart that focuses on the suppliers, inputs, process, outputs, and customers for the process being evaluated. One benefit of the SIPOC diagram is that it forces teams to think about the overall process. Without it, it is easy to forget about the supplier and the customer—both of which are essential to understanding quality. Because there are legitimate similarities between the manufacture of physical products and the production of information, the SIPOC model can be adapted to produce analyses that enable a clear understanding of how data and information products are made.

Fishbone Diagrams

Fishbone diagrams enable a different perspective on process analysis. Their structure is simple. They contain a center line with an identified effect (usually a problem that needs to be solved). Extending from this line are the fish bones that represent the categories to be analyzed. Traditional manufacturing fishbone diagrams include the six M's: machines, manpower, materials, methods, milieu and measurement. Onto these, possible causes can be mapped. Fishbone diagrams provide useful analysis categories related to what can go wrong with a process. They can also be adapted in several ways to draw out details about a process that might otherwise be overlooked. From a data quality perspective, they can be used to drill into issues. They can also be used in a forward-thinking manner to identify risks within a process. Figure E.2 is a fishbone diagram illustrating generic causes of data quality issues.

Plan, Do, Study, Act

One of the most critical processes in any effort at quality improvement is the Plan-Do-Study-Act cycle. This method is rooted in scientific reasoning, which requires a combination of hypotheses, observation (including measurement), inductive and deductive reasoning, and testing to reach conclusions. The PDSA cycle establishes an empirical foundation to understand and define quality problems and to address them in ways that can demonstrate improvement. But its most significant feature is that it is a circular rather than a linear process. By repeating the steps, one can make quality a continuous process. As Deming formulated it in the 1950s (quoted in Moen and Norman),

1. Design the product (with appropriate tests).
2. Make it; test it in the production line and in the laboratory.
3. Put it on the market.
4. Test it in service, through market research, find out what the user thinks of it, and why the nonuser has not bought it.
5. *Re*design the product, in the light of consumer reactions to quality and price.
   Continue around and around the cycle.

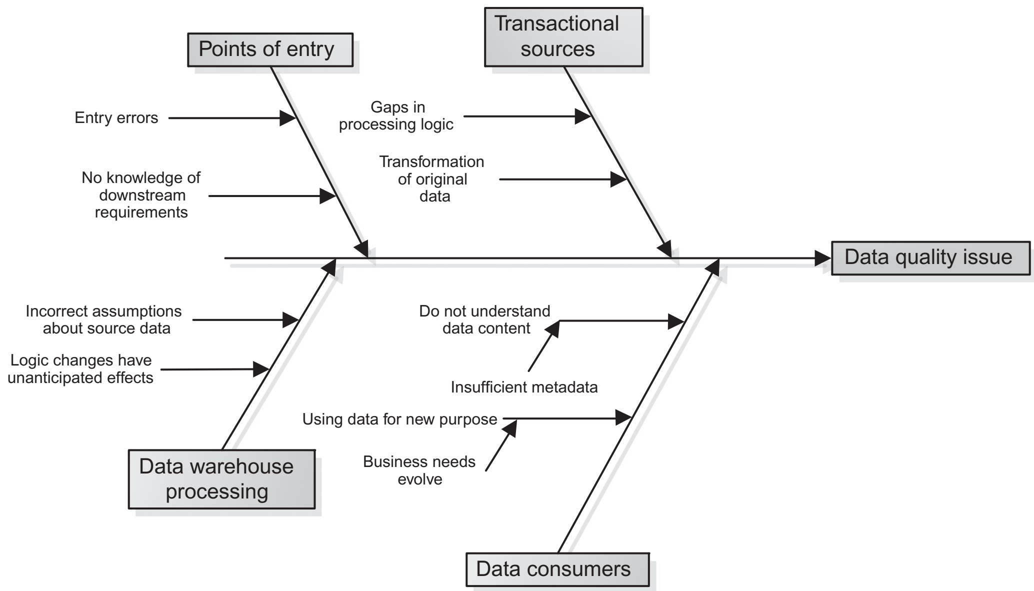

FIGURE E.2 A Cause-Effect Diagram for Data Quality Issues

A fishbone or cause-effect diagram enables analysis of factors that contribute to an effect or problem. The center line describes the effect. The fish bones extending from the center represent the categories to be analyzed. From these extend the primary and secondary causes. Analysis categories can be general (materials, methods, etc.) or specific to a type of problem. The figure illustrates generically factors that contribute to data quality issues in large data stores.

In a more generic form, PDSA includes the following:

- Plan: Recognize a problem or an opportunity for improvement, develop a hypothesis about how to solve it, and plan how to test it.
- Do: Conduct the tests of your hypothesis.
- Study: Compare the results of your test to your hypothesis. Determine where and how results differ from expectations. Define what you have learned from the tests.
- Act: Take action based on what you have learned. If your results demonstrate that the hypothesis was correct, then act on those results. If results show your hypothesis was not on target, then use the results to formulate a new hypothesis and re-test.

This process can be applied to any problem or improvement opportunity, from a subprocess in a service organization to a strategic plan intended to transform a large company. All four steps are necessary for overall success. They are tightly woven together. What the circular depictions of the PDSA cycle do not always show is its recursive elements. Planning includes not only recognizing a problem, but also understanding that problem in the context of an overall process in order to form a hypothesis about how to address it. Testing a hypothesis is itself a complex process that considers what can go wrong as well as what should go right. It requires all the input from the planning stage, and it looks forward to the study stage. Studying the results of your actions enables you to see in a new light both the original problem and the proposed solution. Acting is, in many respects, very similar to planning—though more compressed. It requires that you assess the new state and formulate a hypothesis as to how to improve it.

Define, Measure, Analyze, Improve, Control

The Six Sigma approach to problem solving—Define, Measure, Analyze, Improve, Control (DMAIC)—is a variation on the PDSA cycle developed by Motorola in the late 1980s. Six Sigma incorporates much that has been learned about quality control and assurance into a rigorous methodology. DMAIC steps and deliverables include the following:

- Define: Understand the problem to be solved (such as the defects to be addressed), and the scope of the project, including goals and estimated impact; define the customers and interview them to determine what qualities are critical to them. Deliverables from the define step include a project charter, initial current state description (including current process flow), stakeholder analysis, a business case, and a plan for measurement of the current state.
- Measure: Assess the current process performance (is it stable and in control?) and identify when and where problems occur and who is involved when they occur. Deliverables from the measure step include a detailed process analysis, measurement results and analysis, identified areas of opportunity, a validated business case, and a draft plan for the analyze step.
- Analyze: Determine what variables within the process impact process output. Conduct root cause analysis to determine which causes are primary. Assess which causes to address within the project. Deliverables from the analyze step include cause and effect analysis, validated root causes, scope for causes to be addressed, an updated business case, and a draft plan for the improve phase.
- Improve: Determine what actions will eliminate each root cause. Implement actions and measure their effects. Were improvements successful? Measure to define the new process capability. Define actions to mitigate the risk of changes to the process. Deliverables from the improve step include solutions to address root causes, results of the pilot, risk analysis on the improved process, an updated business case, and a plan for the control phase.
- Control: Determine who is responsible for maintaining improvements. Establish a monitoring approach, with clear controls points. Deliverables from the control step include the control plan, the transition plan, and the control audit plan.[4]

The basic contours of the DMAIC process are rooted in PDSA. What makes the DMAIC approach extremely useful for data quality improvement is its rigor around data itself. All parts of the process are interconnected, and all are dependent on measurement: from initial problem definition and upfront analysis of current state, through hypothesis testing and validation of improvements, to the implementation of standardization and controls to maintain quality.

[4] Adapted from several sources, including ASQ.org, Tague (2005) and notes from UnitedHealthGroup Foundational Quality course.

Implications for Data Quality

As noted at the beginning of the appendix, many data quality practices are rooted in product quality practices. Wang (1998) made clear the connection between the two: "Product manufacturing can be viewed as a processing system that acts on raw materials to produce physical products. Analogously, information manufacturing can be viewed as a processing system acting on raw data to produce information products." (p. 59) While recognizing the limits of the analogy (raw materials are not used up in the manufacture of an information product; some dimensions of data quality, such as timeliness and believability, do not pertain to manufacturing quality), Wang was still able to propose a Total Data Quality Management methodology that closely resembles those for manufacturing quality:

- Define the information product characteristics, requirements, and manufacturing system.
- Measure the quality of the information product.
- Analyze root causes of problems with the information product.
- Improve the information product by aligning it with business needs.

English (1999) approaches Total Quality Data Management directly from manufacturing quality improvement as well, drawing on Deming, Juran, Ishikawa, and Crosby, with their emphasis on building a culture and a mind-set focused on meeting customer needs through continuous improvement. His process improvement approach includes the following steps:

- Establish the information quality environment.
- Assess data definition and information architecture quality.
- Assess information quality.
- Measure the costs of nonquality information.
- Reengineer and cleanse in order to improve information products.
- Improve information process quality in order to prevent future defects.

English also advocates making direct improvements through small, manageable projects that replicate the pattern of definition, assessment, improvement, and control.

Likewise, Redman's (2001) Quality Improvement Cycle closely resembles the PDSA and DMAIC cycles (p. 132):

- Select a project (Plan, Define).
- Form and charter a project team (Plan, Define).
- Conduct root cause analysis (Measure, Analyze).
- Identify and test solution (Do, Study, Improve).
- Implement solution (Act, Improve).
- Hold the gains (Control).

Redman's approach recognizes that the best route to long-term improvement is through supply chain management, which itself requires a similar improvement cycle (p. 162):

- Establish management responsibilities.
- Describe the information chain.
- Understand customer needs.
- Establish a measurement system.

- Establish control and check conformance to requirements.
- Identify and select improvement opportunities.
- Make improvements and sustain gains (through understanding customer needs and measuring conformance to requirements).

McGilvray's formulation (2008), which includes a framework to understand data quality and the Ten Steps to Quality Data and Trusted Information™, is an extended application of PDSA and DMAIC methods adapted specifically to information quality improvement. Assessment is necessary for awareness; awareness leads to action; actions are verified by periodic assessments (p. 55, p. 277). This high-level approach provides the context for the Ten Steps™:

- Assessment (define, measure)
  - Define business need and approach.
  - Analyze the information environment.
  - Assess data quality.
  - Assess business impact.
- Awareness (analyze)
  - Identify root causes.
  - Develop improvement plans.
- Action (improve, control)
  - Prevent future errors.
  - Correct current data errors.
  - Implement controls.

These steps are further supported by ongoing communication of actions and success to the wider organization.

Limitations of the Data as Product Metaphor

Approaches to data quality improvement proposed by Wang, Redman, English, and McGilvray work because there are, in fact, important similarities between the production of physical objects and the production of data. Problem definition approaches, process flows, and other quality improvement tools can be adapted to improve how data and information products are made. One of the primary benefits of this approach is that it helps people understand that data is not just a by-product (an incidental or secondary result) of business processes. It is a usable result of these processes and necessary as input for subsequent processes.

However there are limitations to the product comparison, not so much because of what it asserts but because of what it leaves out. Recognition of the fact that data is not a physical product has important implications for how data is created, understood, and managed. Data is not only a product. It is also a representation of objects, concepts, or events. Using it requires interpretation. Data is a form of knowledge and using it requires knowledge. Managing knowledge is different from managing physical products.

Data as Representation

Data is an abstract representation of objects or concepts. Its primary function is semiotic: It serves as a sign of what it represents. As is the case with understanding any representation, using data involves a level of interpretation. This means it always presents the risk of misinterpretation. While some data is easy to interpret (we know what a person's name is, we know what a birth date is),[5] some is not. Data always requires context. In some cases, it requires a significant amount of context. For people to understand the quality of data they need to know what it represents; they also need to measure or make a qualitative judgment about how successfully it effects this representation.

Data as Knowledge

Because using data requires knowledge and context, managing data requires a level of knowledge management that physical products do not generally require. Seeing data management as a problem of knowledge management sheds a different light on what its purposes are. Because most organizations do not manage data as knowledge, much of the potential value of data is lost to the organizations. Many of the tools used in process analysis of data are aimed at recapturing knowledge that was the basis for choices in data processing but becomes obscured over time; or knowledge that was never made explicit but is nevertheless embedded in the assumptions governing data processing.[6]

Data is valuable not only for how it is used, but also for what it contains: Knowledge about an organization, including rules and assertions about how an organization operates, historical information about its interactions with customers, and routes back to the limitations of systems and other tools designed to make the organization work more effectively. Much knowledge about data can be captured in the form of metadata, but not all of it can be boiled down to metadata attributes. Effective data use requires other support structures, such as formal training and a consistent, usable system documentation.

Data as a Product Redux

The raw material for data is often other data. Any risks related to interpreting the information product that a particular data consumer is using are also associated with the input for that product.

Data is not consumed when it is used. The same data can be put to many uses with little or no cost to an organization. The risks associated with interpretation apply to any use of data. There is a risk that they increase as they move away from original intended purposes, because such movement often also means that data consumers have less knowledge about the data.

Not all data is created "for a purpose." Some data is created purposefully, but some is also a by-product of other processes (often those that involve creating the data that is needed for a specific purpose). It is not always possible to judge the "fitness for use" of data created as a by-product.

[5] I even hesitate to make the assertions about name and birth date, since different cultures have different conventions related to the representation of both these concepts.

[6] See Loshin (2001) for a full discussion on this critical aspect of data quality.

Instead, its quality must be understood based on some other criteria, such as how well it represents the concept that it is assumed to represent.[7]

Malcolm Chisholm (2012) neatly summarizes another facet of the "data as a product" problem in his argument that data quality should not be equated with "fitness for use." Data represents a thing or a concept, but it needs to be understood by someone; this someone is an interpretant —that thing or person to whom the sign represents the object. As Chisholm states, "The interpretant is independent of the data. It understands the data and can put it to use. But if the interpretant misunderstands the data, or puts it to an inappropriate use, that is hardly the fault of the data, and cannot constitute a data quality problem."

Data and Systems

In today's world, data is largely managed through technical systems. Choices in the design of these systems have direct effects on the ability of data consumers to understand and use data effectively (Ivanov, 1972). These systems can be very opaque.[8] A very significant component of analysis aimed at data quality improvement is an understanding of the technical systems through which data is created, processed, and stored. Without this analysis, it can be very hard to understand (interpret) what data you are looking at and to understand its condition or quality.

Concluding Thoughts: Building Quality in Means Building Knowledge in

Despite these limitations, it is clear that data can be managed in ways that produce higher quality results than would be produced without attention to the overall information chain, data definition, and intended use.

Drawing on the product metaphor, Wang's analysis (1998) recognizes the benefits of planning for better information right from the start: "[Information Quality] requirements can be designed into the new information manufacturing system, resulting in quality-information-by-design analogous to that of quality-by-design in product manufacturing" (p. 63). However, achieving this result requires a purposeful alignment between data producers and data consumers: "Information manufacturers as well as information suppliers need to expand their knowledge about how and why the consumers use the information. Conversely information consumers need to understand how information is produced and maintained so that the communication among the different roles can be effective" (p. 65). This kind of alignment requires planning and commitment on the part of the overall organization.

Understanding data quality from the perspective of knowledge management leads in the same direction. Many organizations have become information-based, as Peter Drucker predicted in the

[7]Thanks to Eric Infeld for conversation on this facet of data. Malcolm Chisholm has also written previously on this subject (2010).

[8]The problem of system opacity has been recognized for a long time (since the dawn of the Information Age), but it persists. Ackoff (1967) raised the question of whether managers need to understand their information systems, and based on the example of a costly and ineffective management information system, he recognizes that managers often feel incompetent to ask simple questions about technical systems. He asserts, "They would not have allowed a hand operated system to get so far out of their control. …No MIS should ever be installed unless the managers for whom it is intended are trained to evaluated and hence control it rather than be controlled by it."

1980s. Such organizations depend on individuals taking responsibility for the information they produce by asking: Who depends on me for what information? On whom do I depend? (Drucker, 1988, pp. 2, 10–11).

In an organization of knowledge specialists, one of the biggest challenges is knowledge sharing. Knowledge management literature identifies two types of knowledge: tacit knowledge, which is internal to a person, and explicit knowledge, which is formal and systematic and can therefore be studied and learned by individuals. Knowledge sharing enables tacit knowledge to be made explicit and explicit knowledge to be learned so that individuals internalize it (Nonaka, 1991).

When we think about knowledge related to data, it seems that it should be explicit. In order to create data, we assume a process of definition that should result in explicit knowledge. But this process is not always executed effectively, and the resulting metadata is not always managed. More importantly, because the uses of data evolve, there is need for ongoing documentation of knowledge. Very few organizations respond to this need. As a result, some of the most important knowledge about data resides in the heads of individuals rather than in documents or systems through which it can be shared and passed on.

Knowledge management and process management are rooted in the scientific method. Consequently, they share characteristics and methods for systematic problem solving, purposeful evaluation of experience for lessons learned, deliberate adoption and application of best practices, and the integration of these practices into the operations of the organization (Garvin, 1993).

Efforts to improve data quality will be more effective when seen from this twofold perspective: data as a product and data as knowledge. As Drucker points out, executive leaders in information-based organizations "need to think through what information is for them, what data they need: first, to know what they are doing; then to be able to decide what they should be doing; and finally, to appraise how well they are doing" (Drucker, 1988, p. 12). Such questions are focused on how to move an organization forward in its mission. Answers to them provide the foundation for data quality strategy.