Crafting Rules: Context-Reflective Data Quality Problem Solving

YANG W. LEE

YANG W. LEE is an Assistant Professor and Joseph G. Reisman Research Professor in the College of Business Administration at Northeastern University. Dr. Lee's publications have appeared in leading journals such as *Communications of the ACM, Journal of Management Information Systems, Sloan Management Review, Information & Management,* and *IEEE Computer.* She coauthored *Quality Information and Knowledge* (Prentice Hall, 1999), *Data Quality* (Kluwer Academic, 2000), and *Journey to Data Quality* (MIT Press, forthcoming). Her research interests include data quality, IT-mediated institutional learning, and systems integration. She was a visiting professor at MIT Sloan School of Management, where she taught e-systems integration and conducted research on data quality. Professor Lee cofounded Cambridge Research Group, a firm specializing in information quality. She was also a Conference Co-chair for International Conference on Information Quality (ICIQ) in Cambridge. She received her Ph.D. from MIT.

ABSTRACT: Motivated by the growing importance of data quality in data-intensive, global business environments and by burgeoning data quality activities, this study builds a conceptual model of data quality problem solving. The study analyzes data quality activities at five organizations via a five-year longitudinal study.

The study finds that experienced practitioners solve data quality problems by reflecting on and explicating knowledge about contexts embedded in, or missing from, data. Specifically, these individuals investigate how data problems are framed, analyzed, and resolved throughout the entire information discourse. Their discourse on contexts of data, therefore, connects otherwise separately managed data processes, that is, collection, storage, and use. Practitioners' context-reflective mode of problem solving plays a pivotal role in crafting data quality rules. These practitioners break old rules and revise actionable dominant logic embedded in work routines as a strategy for crafting rules in data quality problem solving.

KEY WORDS AND PHRASES: context-reflective problem solving, data quality, data quality rules, information quality, problem solving, reflection-in-action, situated practice.

BEFORE BASEBALL RULES WERE CREATED, the notion of "striking out" did not exist. A rule defines a practice, and a practice requires a rule. Rules and practices have a reciprocal relationship [13, 23, 40, 79, 81]. Like baseball, other professions espouse and use rules of their practices. Some cumulated and repeated rules form stable sanctioned routines in practice. Unlike baseball rules, data quality rules have changed dramatically as data are increasingly used in networked, data-intensive, and distributed business

environments. Despite the short history of data quality as a distinct field, data quality professionals have been developing, revising, and crafting rules for data quality problem solving for a long time, and are often forced by urgency to do so in unconventional ways that bypass standard organizational procedures. Some new rules conflict with, and thus break, stable conventional routines and rules in practice.

For example, a data quality administrator in a hospital retraced missing contexts of a data set [78], an unconventional practice in the hospital. A global manufacturing company strived to reintegrate "official" procurement data across distributed and autonomous business units [30], a new practice in the company. A managed care organization (MCO) struggled to unravel the root causes of overpaid claims [41], a first attempt in the organization. This new, untested, and sporadic problem solving by data quality practitioners in the field has not been studied in a formal research context. Often, these practical trials have been demoted as temporary fixes or workaround resolutions for exceptional and emergency cases.

Some might argue that this study simply advocates such exceptional activities instead of promoting a standard practice. In fact, the standard practice of data quality is to understand the rules for exceptions. These exceptional activities offer an opportunity to review and learn underlying reasons for the disruptive manner in which organizations manage data, and new ways and rules in which data quality problems can be analyzed and solved. As high-quality data has emerged as a new basis for competition, analyzing leading data quality practice is a must. As today's business strives to offer agile, highly customized, and globally sustainable solutions, producing high-quality data has become a baseline for managing strategic corporate capacity and assets beyond operational necessity.

On the research front, the data and information quality field has produced research in areas such as modeling information manufacturing processes [7], defining data quality [78, 85], measuring data quality [64], and managing data quality [49, 50, 78, 86]. This stream of research has provided a solid foundation, which data quality researchers and practitioners can apply to various settings and problems.

To move research forward, data quality researchers need to build a deeper and broader theoretical foundation that incorporates multiple theoretical underpinnings in complex data quality practice. It is critical to establish further theoretical foundations in order to understand how different theoretical principles in technical and managerial areas are integrated and applied to solve data quality problems in actual organizational settings. This study, which raises an inquiry into context-reflective data quality problem solving, aims to provide a step toward building such a foundation, the current foundation underlying the study of data quality.

Consider an example. A data quality manager and his team traced the divergent and conflicting principles used by two different disciplinary fields, accounting and database, in recording the same event or business process, that is, sample shipping. Understanding the "differences," not as violations of rules or wrongdoings, the team resolved the problem and revised its data quality practice. This situated practice by the data quality team exemplifies how practitioners resolve the contested contexts presented in various situations [13, 46, 79].

Thus, a general question of this paper is: How do practitioners in leading organizations solve information quality problems? To understand how data quality professionals solve information problems, this study operationalizes the question as follows: How do these practitioners identify, analyze, and resolve information quality problems within the information production discourse, specifically, the information collection stages, the storage protocols, and the general use of information?

The study findings suggest that professionals solve problems by crafting rules to integrate business process and data process. They approach this integration by explicating data quality contexts embedded in data or reestablishing the missing contexts in data. Reflecting on and explicating contexts dictates how a data quality problem is framed, analyzed, and solved. The study findings further suggest that context-reflective knowledge about solutions must be recorded and shared. Reflecting context explicates knowledge about why a company collects particular data, how the data were stored, what constraints were imposed, and how information was used. This study also introduces different contexts, such as paradigms, goals, roles, time, and space, which form salient contexts for data quality problem solving.

Theoretical Grounding of Key Concepts

Rule

A RULE IS DEFINED AND USED IN VARIOUS AREAS. A rule defines processes and activities of a practice [40]. Rules indicate dominant logic of a company's business strategy [67], an example area of business practice. Rules are also used to elaborate and frame conflicts embedded in a policy as different rules compete and shape the design of the policy discourse [69], an example process of rule-making in practice.

A rule embeds actionable dominant logic, sanctioned by collective decisions, customs, and culture of an entity. Stable rules are institutionalized and embedded in structured work, such as routine work procedures, software codes, and automated information systems processes. The institutionalized rules are embedded in work routines and in resulting information. Some rules are followed habitually in accordance with their intent, and thus escape reviews. Some rules, therefore, can be hidden and forgotten until a form of inquiry is raised [3, 4, 10, 12, 23, 65, 73, 87, 88]. Explicating obliviously practiced rules in work routines often involves practitioners' reflection-in-action on the contexts of the rules.

Context

Context is commonly used, first, to specify a scope or a boundary of a study area or a discourse. Demographic or geographical contexts are examples [2]. Second, context is often referred to as an external environment. Studies in the systems development area treat a context as a boundary to focus the area of study, and tend to focus on the system developer and producer contexts [14, 28]. Recent works in distributed computing recognizes context as an important factor for interoperability of computing.

Chiklin and Lave [15] point out the limitations of studying phenomenon without exploring the socially constructed context. Despite various interpretations of context, the commonly implied thread among all uses of context stems from the notion of fundamental environment that contains activities and phenomena in study.

This study offers a third, alternative view. Context is a differentiator and relationship builder, specifying the relationship between contents and environments. It connects and shapes contents and the environment, which in turn structures and influences the activities via the context. Context is the differentiating and connecting relationship entity that inherits the imprints for further processes and activities.

In practice, contexts in data quality management have been implicit, yet they have been a critical part of resolving data quality problems. Naturally, exploring context is a critical direction for practitioners in devising their solutions. Despite the fact that many conventional data quality practices focus inevitably on cleansing data in stored databases, many engage in solving problems in the broader scope of data collection, storage, and use over time.

Once the contexts of data are known, recorded, and stable, solutions, such as context interchange technology, can be applied for mediating differences in data and systems integration between sending and receiving entities [54, 55]. This solution approach is particularly useful for disparate heterogeneous databases that require data aggregation or integration across space and time.

This paper defines a context that includes paradigm, role, goal, time, and space. Each of these contexts are further defined and elaborated in the fourth section.

Reflection-in-Action in Problem Solving

This study uses Schön's concept of reflection-in-action as a guide to understanding and analyzing data quality practice. Schön [76] identified reflection-in-action as a lens for professionals to reflect theory-in-action in their work contexts. Reflection-in-action refers to a *reflective conversation with the materials of the situation* and is based on a *sense of the situation as a whole* [74, 76]. It is a dynamic process. As a practitioner *acts,* his sense of the whole is changed. And as he makes new sense of the whole, he will change his actions accordingly. Schön calls the situation's *back talk,* as the situation's "answer" to the acts of the practitioner. A reflective practitioner is able to handle situations characterized by uniqueness, complexity, and conflicts because of his ability to reflect-in-action. He is, thus, capable of making sense of situations he has never met before [4].

Problem solving is a particular kind of a decision-making process [59, 60, 66]. More complex factors become relevant as the problem solving level raises from an individual to a group and to an organization. Organizational problem solving involves a wide range of resources, potential for conflicts, an increased number of stakeholders, and consequences that might expand to broader environments.

Data quality problem solving involves organizational dynamics and various natures and implications of information technologies, as data involves both business processes and information technology. Practitioners' reflection-in-action during prob-

lem solving activities involves review, evaluation, learning, and revising decisions on various issues at different levels of an organization.

Revising Dominant Logic as a Mechanism for Revising Practice

Organizational change and strategy theories attempt to explain how organizations move from one stable state to another. Prahalad and Bettis [67] offers dominant logic as a pivotal mechanism for strategic choice and performance. Bacharach et al. [5] offers logics of action as a useful tool for understanding the mechanism for change. Quoting Goffman [33], Collins [19], and DiMaggio [24, 25, 26], Bardach et al. states that the logic of action allows actors to frame the specific means and ends that they bring to an exchange relationship. Bardach et al. further explains that logic of action is, for the most part, taken for granted. It becomes manifest when parties try to explain to themselves or justify to others the selection for specific means, ends, and the linkage between the two. Similarly, Giddens [32] called this *rationalization of action.* Stokes and Hewett [77] used the term *aligning actions,* and Miles and Snow [58; see also 57] used the term *frame alignment processes.*

When challenged by a problematic situation, practitioners reflect on how the problem is set and examine the assumptions made, in order to analyze and resolve the problem. Their dominant logic of action will be revised based on their reflection-in-action, while they review the entire information discourse. Eventually, data quality professionals provide insights into the transformation from one state of consistency to another. This transformation is the lever used by this study to explicate the process by which dominant logics of improving data quality in organizations are transformed by analyzing how problem contexts are changed over time and how professionals revise rules to overcome the incongruities between different organizational and information rules embedded in their work routines.

Research Methods and Sites

Research Methods

THIS STUDY USES A COMBINATION OF embedded case analysis and longitudinal action research. This combination is often suggested for theory-building in a new research area involving complex phenomena [8, 29, 47, 80, 89]. Embedded case design reinforces internal rigor, as case data are analyzed within the broader context of the organization. In this study, project episodes are further analyzed within the context of data quality initiatives, and within the broader organizational context. Multiple cases additionally add external validity.

The focus of this study, data quality problem solving and practitioners' reflections on their work demand (1) intimate knowledge of data quality and (2) long-term observation of practitioners at work. A longitudinal action-research method provides meaningful access to organizations and allows a researcher to closely examine and reflect on initial findings, avoiding modeling and reporting temporary

phenomena. This method is also in line with Daft and Weick [21], who calls for intensive researching and experiencing in the study environment [82] for studying new and complex phenomena.

For analysis, this study adapted Schön's approach emphasizing theory-in-use along with espoused theory of agent's actions and their actual meanings in addition to traditional case analysis [74, 76]. To further record and analyze data quality problem solving activities, this study adapted a classical coding protocol, problem behavior graph (PBG), and the approach for studying "episodes," originated by Newell and Simon [60] in their cognitive study of human problem solving. Using a project or an episode as a unit of analysis enables one to capture detailed modalities within the action taken, and, thus, record the detailed changes and activities within a process of problem solving. A project/episode has a goal in mind that opens and closes an action, which cumulatively shapes behaviors of the problem in study. Moreover, this detailed analysis enables researchers to understand contexts both before and after the project, so as to understand how rules are changed over time.

The unit of analysis is a data quality project within a data quality initiative. The data quality initiative includes a variety of projects. A project is treated as an "episode" in this study. Episode is defined as a sequence of moves, each of which is a succinctly describable segment of behavior associated with attaining a goal. Newell and Simon [60] treat each episode as a unit, paying attention primarily to what determines which episodes are initiated and how the episodes are terminated. The use of episode in Newell and Simon is at the individual and cognitive level: their logic and organization of "episodes" of behavior in a scene of a game is applicable in data quality activities in organizations. Episodes, since they are tied to goals, can be hierarchical, with one episode embedded in another. Within the episode, a subject may initiate subepisodes. As such, a project may have subprojects. This study used a data quality project as a unit of analysis, an episode within an overall data quality initiative in an organization.

Research Sites

Early adopters of data quality practice are ideal candidate organizations for this type of study because of their intense reflection on the problems and their activities. Problem solving involves focal agents. They are data quality managers and executives with various job titles, such as Data Quality Manager, Information Product Manager, Senior VP of Information Quality, Data Quality Administrator, Global Information Architect, Data Analyst, and Data Service Manager. They manage organizational data, provide an analysis of problems and important decisions, and, therefore, are key agents of making and revising rules about data quality improvement. The five companies involved in this study are: a global manufacturing company, a hospital, an agricultural company, a global financial institution, and a data service company. I have had access to these organizations, for this research, for five to six years as of this writing. I had longer access to some of these companies for observing their overall data quality work and progress.

Each organization had a key informant who led the organization's data quality initiatives. I have been in contact with them periodically, visited the sites, and discussed projects periodically face-to-face, by e-mail, and on the phone. When on-site, I observed data quality–related meetings, collected reports, and participated in work seminars and meetings to gain greater access and understanding. Each site's characteristics and its data quality work are summarized in Table 1.

Context-Reflective Problem Solving

Contested Contexts Frame Problems

THIS STUDY STARTED WITH A NOTION that data activities, as with other activities, are influenced by contexts [15]. Based on previous research and a field study, this study identifies five key related contexts for data quality problem solving: paradigm, role (information role), goal, time, and place. Understanding these contexts is pivotal in setting and analyzing problems and devising solutions [69]. Multiple contexts are contested over the entire information discourse, which frames problems and thus directs problem analysis and solutions. Data flow across multiple places, where different customs, policies, and procedures[1] shape data and data-related problem solving efforts. Commonly identified contexts are *place* and *time,* as they typically specify what data mean and how data are represented. *Place* dictates the socially constructed meaning of data. *Time* further determines how certain data are represented and used. A common experience of traveling to a foreign country provides many everyday examples of place and time context, such as, calculating and converting currency and time. Dramatic examples are often reported in newspapers. NASA lost a Mars shuttle a few years back because highly trained engineers, working in geographically different places, routinely used local units of measuring systems for their globally collaborated work.

Currently, most data quality researchers agree on the meaning of high-quality data as that data which are fit for use by data consumers. *Goals* or purposes of data, therefore, is a basic and important context. This study also includes *role,* that is, information role, as a key context. Taking the view that data ought to be managed as a product [86], the entire information production system that includes the three distinctive information production areas and the corresponding roles are an integral part of data production and data quality problem solving. The three data roles include a data collector, data custodian, and a data consumer [78]. Data, though intangible, as with any tangible artifacts, are influenced by *paradigms* or principles that individuals hold and apply when designing, collecting, storing, and using data. These principles are collectively agreed upon, taught, and practiced in a particular field over time, and have a lasting influence on how individuals solve problems. As data quality problem solving involves various subject areas, it is critical to explicitly include *paradigms* as a relevant context. The five related contexts form primary contexts for data quality problem solving. These contexts play critical roles for practitioners in their reflection-in-action as they break old rules, and revise and create new rules. The rules

Table 1. Summary of Companies Involved in the Field Research

(1) Organization (pseudonym)	(2) Industry and business	(3) Organizational setting	(4) Information systems setting	(5) Data quality environment	(6) Focus data types	(7) Data quality mechanisms	(8) Example rule in question	(9) Study period
Matcom	Manufacturing: consumer products manufacturer.	A multidivisional global manufacturing company headquartered in the United States.	Data warehousing and enterprise resource planning (ERP) distributed computing architecture.	Data quality in ERP and data warehouse for global business.	Sales and shipping data. Product data.	Information quality manager, data quality software, systems development.	Data integrity rules for local and global business.	1997–2003
Healthcom	Health care: hospital.	A major hospital with a central data service in the United States.	Data warehouse, client-server architecture.	Data quality in their operational and R&D analysis databases and data warehouse.	Patient care and patient discharge data.	Information quality manager, total quality management, statistical measurement and control tools.	Data flow and multiple uses of single data attribute.	1997–2003
Agcom	Agriculture: agricultural products and meat processing company.	A major agricultural and meat processing company in the United States.	Data warehouse project to provide quality information.	Data tailored to production department did not make business sense for the enterprise.	Production and sales data.	Systems development, data quality tools.	Representing cross-functional data.	1996–2002

| Fincom | Financial: investment company. | A major international bank headquartered in the United States. | Distributed heterogeneous systems. | Data in their customer profile database for risk management. | Customer data. | Data quality tools, information managers. | Currency of customer data for global risk management. | 1995–2000 |
| Datcom | Information technology: data services company. | A major data product and service provider in the United States. | Client-server architecture with heterogeneous database system. | Data quality in their product service delivery. | Business operational data, product service data. | Data quality tools, quality process control, data quality managers. | Coordinating work routines across information production process. | 1996–2001 |

underlying these multiple contexts are contested, negotiated, and determined for creating new rules to solve a data quality problem. These new rules are likely to revise data quality practice. These five contexts are further elaborated as we examine case studies in the next subsection. Their conceptual definitions are provided below.

The *paradigm context* consists of disciplinary principles that form the rules and procedures for data quality problem setting and solving [45, 62, 63, 65, 71]. For example, data fields without missing values in a relational database are important data integrity principles in database technology. As a result, one might flag a field for missing values and add edit checks [16, 22, 70]. This study calls it a "book" rule as shorthand for this context. The *role context* refers to specific roles that individuals play in the information discourse and the entire information production system. This study calls it the "who" rule. Previous research has identified three key data quality roles: data collector, data custodian, and data consumer [39, 78, 86]. For example, data collection is part of the data collector's role. In such a role, one may not have complete knowledge of how data will be used by data consumers. The *goal context* includes objectives that individuals and organizations aim to achieve through creating, using, and processing data. For example, a marketing manager wants to use data to increase market share. This goal may be local or global, operational or strategic, internal or external. This study refers to the goal context as the "why" rule. The *time context* is the time frame during which, or for which, data is used and processed. For example, a financial manager using the same set of data may assess its quality differently based on short-term investment decisions versus long-term investment decisions. This study refers to the time context as the "when" rule for short. The *place context* refers to the locale for which data are generated, used, or processed. For example, locally different units, formulas, customs, and legal requirements can demand data of different place contexts. The "where" rule applies to this place context.

Table 2 summarizes example projects from all five companies that illustrate the contexts explained here. Excerpts of project episodes are included in Appendix A.

Reflection-in-Action and Problem Solving

A reflective practitioner [75, 76] relies on problem solving mostly in conversations with a situation. Their reflective conversations can be gleaned from analyzing contexts explicated, activities involved, and rules used and discussed in the entire information discourse. A Matcom case explains activities involved in each move that reveal different contexts and rules-in-use. For example, rules are applied and revised based on paradigms and principles from different disciplinary subject areas. Problem solving activities at Matcom with graphical emphasis on key moves is shown in Figure 1.

Newell and Simon [60] originated rules for PBGs by coding and analyzing problem solving activities—for a chess game. They traced each episode of chess moves. Because of the similarities of retracing and going back to the old moves to review and analyze the strategies of a move in the entire game context, data quality problem solving could be coded with extended PBGs. The extended PBG incorporates additional representations necessary for data-specific changes and requirements in the

Table 2. Context-Reflective Problem Solving: Examples in Data Quality Projects

(1) Organization (pseudonym)	(2) Data quality initiative/ objective	(3) Data quality project	(4) Problem identified	(5) Salient rules in question	(6) Contested contexts	(7) Reflection-in- action	(8) Revised rules
Matcom	Improving data quality for global enterprise use in conjunction with their data warehouse and enterprise resource planning.	Data integrity improvement for global data integration.	Upon auditing, stored data violate data integrity rules thus not acceptable for further use.	What do we mean by data integrity rules from local accounting and global business perspectives?	*Paradigm, role* contexts: database principles and accounting principles were in conflict.	Understandability of data for local business use and consistency of data for global business purposes.	Devise technical data design, global process, and product list to reflect new business needs.
Healthcom	Data quality improvement for external reporting and internal analysis.	Mapping data and process flow to streamline and find sources of data quality problems.	Data passed all conventional data auditing for proper storage but hard to interpret the meaning.	Same data attribute used for multiple purposes over time by different departments.	Different *goal, time*, and *place* contexts.	Conflicting uses of attributes by different departments and rigorous rules for database design.	Information product map needs to be reinforced by collaboration and communication policy.

(continues)

Table 2. (Continued)

| (1) Organization (pseudonym)	(2) Data quality initiative/ objective	(3) Data quality project	(4) Problem identified	(5) Salient rules in question	(6) Contested contexts	(7) Reflection-in-action	(8) Revised rules
Agcom	Quality information for different departmental needs and enterprise-wide analysis.	Data quality assessment by multiple stakeholders and data warehouse project.	Data from production and sales were accurate but did not make business sense for company-wide business purposes.	Accurate data from multiple departments does not add value for business understanding.	Representing different *goal* and *place* contexts.	Understanding cross-functional needs for data that can be shared.	Departmental data should be collected and stored to meet enterprise business purposes.
Fincom	Data quality in customer profile for analysis and customer management.	Data needs assessment for customer database.	Customers' complaints and legal complications are based on data quality.	Security and accessibility versus currency of customer data.	*Time, role* contexts for customer profile data; short versus long term, data collector versus consumer.	Time context was not understood and represented in data collection policy.	Data consumers need to inform data custodians about data needs.
Datcom	Data quality improvement in customer service and manufacturing database.	Process mapping and finding bottlenecks.	Complaints about delivery, after streamlined process reengineering.	Speeding up efficiency of work versus coordination of work.	*Goal* and *place* contexts: completion of work and delivery of product.	Cross-divisional data product delivery is not in synch. Working fast is not the only goal.	Data production needs to be collaborated and synchronized.
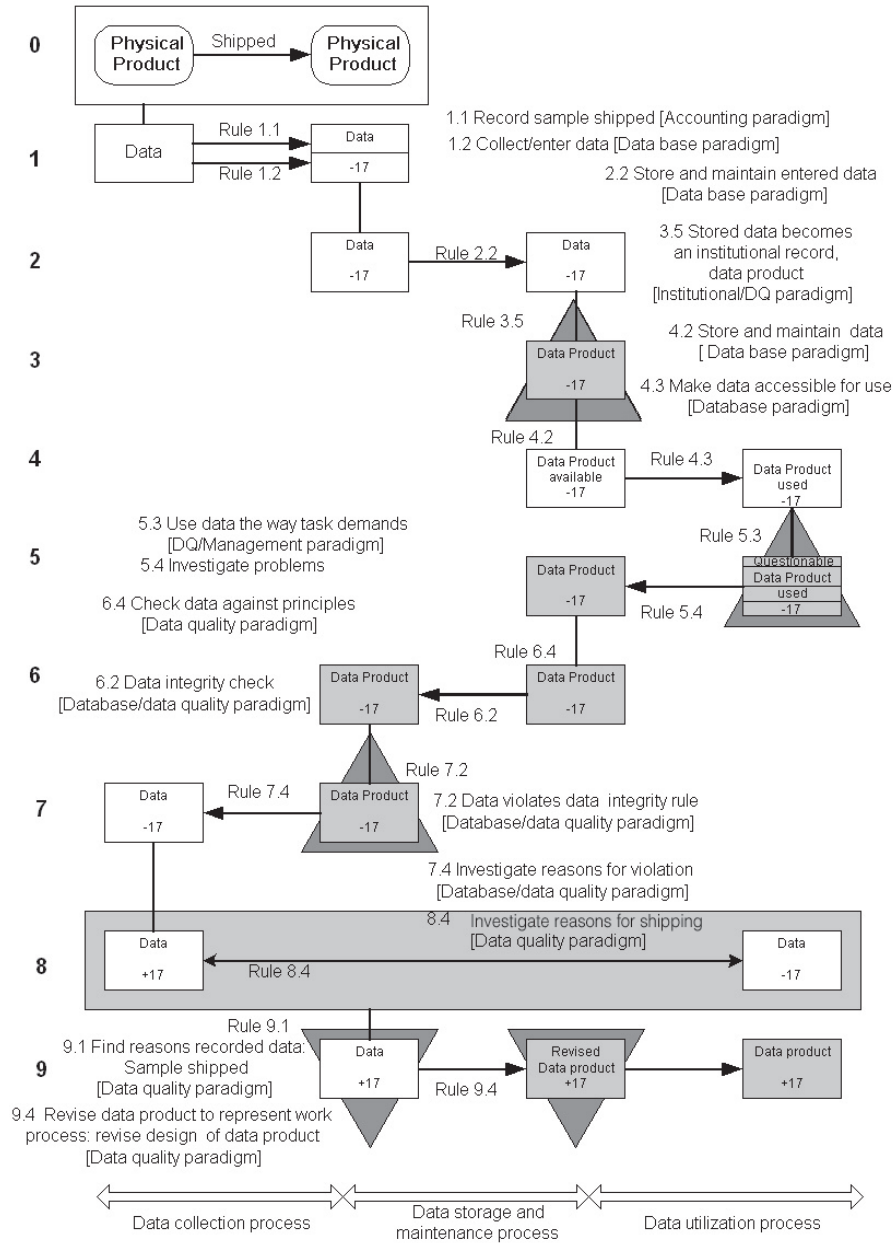
Figure 1. Rule Crafting: Problem Behavior Graph of Example Project Episode at Matcom

information discourse, such as data product changes and data production processes. Table 3 identifies legends for extended PBG, a data quality problem behavior graph (DQ-PBG), as used in the study.

Each move in the Matcom's episode on sample shipping is further elaborated below.

Table 3. Coding Details for Data Quality Problem Behavior Graph

1. [Data] A rectangle identifies each node of data or data product status.

2. ⟶ An arrow identifies application of a rule.

3. 1, 2   Numbers in vertical order identify moves of activities associated with data, which signifies an attempt of reflection-in-action by problem solvers.

4. ▲ A triangle behind the rectangle identifies a change of data status.

5. ⟺ A thick arrow identifies data quality problem space: data collection, storage, and utilization.

6. | A straight vertical line that connects nodes identifies different moves of node without changes in data.

7. Rule 1.1 ⟶ Application of rule ("operator" in Newell [59]) is specified such as 1 or 2 to record change/move in node (data) and 1.1, 1.2, 1.3, and 1.4 to further identify paradigms-in-use from different areas, such as accounting, database, management, and data quality.

8. Time runs to the right, down and left, following the arrows. When no change is made to a node (data), a straight line is used instead of arrows with directions.

9. ▭ Large rectangle behind data rectangle or a node identifies a major reflection-in-action and rule changes. Thus, a locus of pivotal reflection by a problem solver can be identified on the graph. The rules that applied, revised, and changed to a specific data and data product can be traced and recorded.

10. ⟺⟺⟺ The three processes (collection, storage, and use) are three forums of problem space for reflection and discussion for data quality problem solving.

Move 1: Rules 1.1 and 1.2 show rules using the paradigm context in accounting and database. Rules used by practitioners are not necessarily textbook principles; they are paradigms and rules as interpreted and actually used by practitioners in Matcom. In recording shipping data, accounting paradigms were upheld and detail rules were used. "Shipping samples is not a revenue-generating activity, so they should not be counted. We should differentiate from the real shipping. Therefore, we use a minus sign showing the non-revenue–generating shipping activities. We thought we were careful and quite clever on this" [a Matcom manager].

Move 2: Rule 1.2 took "–17" as an official value and stores it without changes or tempering.

Move 3: Now, with the proper storage and maintenance rules of database, the data on shipping became technically and authentically an institutionalized data product.

Move 4: The institutionalized data product on shipping was prepared for users (data consumers) to retrieve and use for their business purposes.

Move 5: Rule 5.4, investigating problems, set off-tract from the typical data process. Note the reverse arrow direction. The data showing minuses puzzled a group of data consumers. Reverse arrows signaled the starting point of activities for review of upstream information discourse.

Table 4. Moves in Example Episode: Summary of Activities

Move	Summary of activities
0	Episode involved in the discourse.
1	Produce data for the activity, shipping.
2	Enter data from Move 1.
3	Store data in database and make it accessible for users: data are technically ready for use. Data become institutional record.
4	Use data.
5	Find a puzzling data in use.
6	Local review in database storage.
7	Shift focus of discourse upward.
8	Comprehensive review and revise of rules includes all data productions processes.
9	Revise rules on data product and their processes.

Move 6: Reviews data against local principles, that is, database principles, data integrity rules in specific, the most fundamental principles with available tools that can be handled within information systems storage process, that is, database and information systems area of organizational functional reporting boundary. Move 6 produced a report of all data that violated data integrity rules, data storage domain's paradigm.

Move 7: The discourse shifted the focus to data collection process, reviewing Moves 1 and 2.

Move 8: Conducted a comprehensive all information discourse review, revised rules involved and discuss, share, and record learning. This move marked the climax of the Matcom's practitioners' reflection.

Move 9: Implemented the results from Move 8, and institutionalized changed rules on data product and related business processes. The company revised rules on data product and reinstituted the changes in samples shipped. New data attributes are created in their database. Old rules of recording in the collection process were revised and the review process in the storage process established, and the use process was ensured of new rules.

The discussion on the sample-shipped information discourse was shared, discussed, recorded, and reinstituted. All move activities are summarized in Table 4.

The five contexts represented in the study—paradigm, time, goal, role, and place—are further elaborated below. Note Move 1. Rules 1.1 and 1.2 are assumed and applied in parallel with no contest. Potential conflicts between two paradigms, accounting and database were not obvious at this stage of the information discourse.

Moves 5 through 7 demonstrate active reflections by participating practitioners. Major context in contest was paradigms and principles in different areas. It also involved *role* contexts contested and reviewed, as each area of information production, collection, storage, and use context (3Cs as three roles, as customarily called in the data quality industry: data or information collectors, custodians, and consumers). Reflections on *role* contexts become even more obvious in Move 8.

Move 8 reviews *time, place, goal, role,* and *paradigm* contexts as integral contexts to review for revising rules on how data and related processes should be managed and recorded. Local and global business *goals* and associated data *goals* are discussed along with cultural differences in different *places.* Different information *roles* are discussed. *Time* contexts for adjusting and reconciling data are also reviewed. Overriding contexts in this episode were *paradigm* and *role.*

Moves 2 through 4 illustrate local implementation of data discourse at the information systems area. Database principles were used without question, as a typical data storage and maintenance process. Moves 4 and 5 also illustrate typical use of data that is prepared for access and further use.

Because this study used a multiple context perspective, analyzing practitioners' reflections throughout the information discourse had revealed changes in contested contexts and rules used.

During the five-year observation, Matcom was often engaged in reflection of various contexts. *Place* contexts were often discussed due to their global business focus. Same products are named and marketed differently as a way of marketing a product that respects and exploits geographical and cultural nature of different places. This differences in *place* contexts discussion prompted Matcom to launch its data modeling and data architecture projects for their entire product, its materials, packages, and distribution channels. Their product categories involved *time* contexts as the categories were evolved and new products were produced. The product family trees got complicated. Data represented transactional *goals,* but did not satisfy strategic *goals,* which aimed at answering strategic business questions. For example, the Matcom's data were unable to answer: "How much business do we have globally with our best customer, Wal-Mart?" These discussions prompted their initiative to focus on the enterprise-wide view for answering strategic questions and launched enterprise resource planning with strategic questions and answers in mind; the *goal* context was discussed for managing this problem. All these contested contexts, reflected over the years, were the backbones of their data quality initiatives that were direct results of Matcom's practitioners' collective reflection-in-action of their entire data quality discourse.

Figure 2 depicts critical moves that overlay the details of problem solving activities. Table 5 summarizes the contexts involved in the critical moves in Matcom's example.

Other companies' episodes analysis also revealed that all these contexts were contested, reviewed, and reflected upon. Due to space limitation, Table 2 summarized the contexts involved.

As illustrated in Figures 1 and 2, a data quality problem is revisited over time as the problem moves along different problem areas. Each move and review of the problem offers an opportunity for the problem solver to reflect on the problem and devise a solution. At each move, the problem solver is applying rules based primarily on different paradigms, that is, principles used and sanctioned in different subject areas, such as accounting, database, management, or data quality. The same or different problem solvers may approach each move differently to form the company's infor-
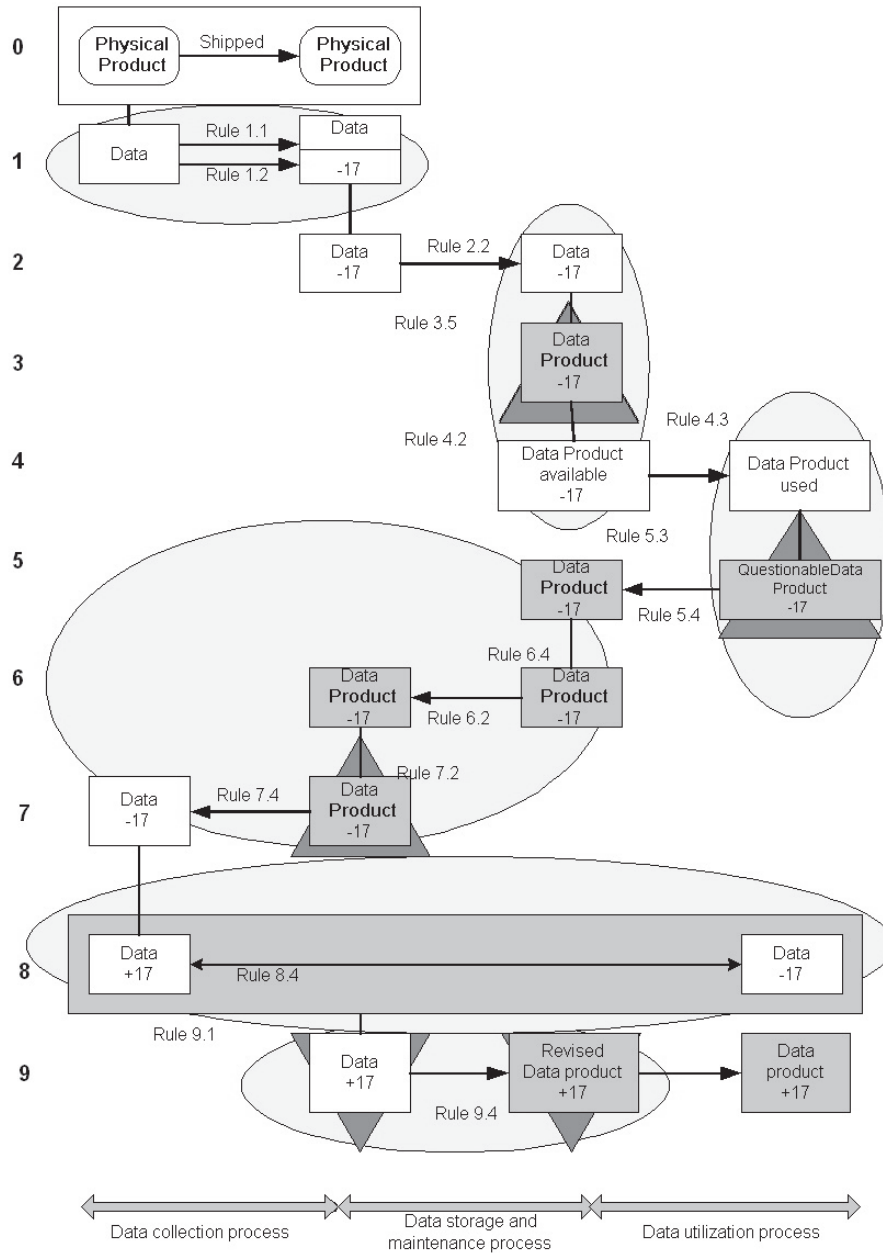
Figure 2. Context Reflected in Example Project Episode at Matcom

mation discourse. Critical reflection depicted in the dark rectangles in Figures 1 and 2 show the activities by Matcom's data quality manager with the team in the organization. A similar case was made for other contexts, such as goal, information role, place, and time contexts.

Table 5. Contexts Involved in Problem Solving

Moves	Contexts involved
1	Local contexts of *paradigm* (accounting and database) were applied in parallel.
2–4	*Role* context (data custodian) played a key role.
4–5	*Role* context (data consumer) played a key role.
5–7	Two *role* contexts (data collector and data custodian) are involved. *Paradigm* for data custodian played a key role for reflection.
8	All *role* contexts are reflected. *Paradigm* context reflected. *Time, place* context reflected.
9	Institutionalized data with revised rules, involving *all contexts.*

Crafting Rules Revises Practice

Rules are embedded in work routines. The logic behind the rules is often masked by coded computer software, integrated work processes, and further hidden by specialization and division of labor. Thus, when challenged by "wicked problems" or intractable problems, conventionally assumed and practiced rules need to be examined. Data quality administrators and data quality analysts work to resolve problems by reflecting further on the information discourse: "Who from where used which book rules to the data?" Explicating contexts underlying the information discourse is a salient and critical activity that practitioners undertake to understand and resolve different rules used in data quality problems.

Rules used at each process are grounded on the dominant logic that is assumed and sanctioned by the organization. When a problem solver, often a data quality manager, is challenged by "indeterminate zones" [75], that is, situations where contested contexts reveal in choosing and adequacy rules in the conventional combination of principles and techniques, one sees an opportunity for revising the rules. This activity of revising the rules in context revises the dominant logic, and thus revises the practice of work as one solves the problem in a new way.

His own thinking schema or framework also guides the reflective practitioner, what Bateson calls an appreciative system [10]. On the basis of one's initial appreciation of the situation, the reflective practitioner forms intentions of changing the situation into a more significant one, the situation that is a better fit with the appreciative system. A reflective practitioner tries to shorten the distance between the materials of the situation and his appreciative system. This process also influences explicating contexts and revising rules in problem solving [76].

Discussion, Implications, and Lessons Learned

THIS RESEARCH OFFERS IMPLICATIONS for both research and practice. For research, next steps might include related topics such as data source tagging, meta data, and knowledge about data. Applying existing research on collective knowledge to data

quality can be a useful first step for this stream of inquiry. They include the research reported in [6, 9, 11, 17, 18, 27, 35, 36, 37, 42, 43, 44, 48, 56, 61, 68, 72, 83, 84].

Another stream of research might examine how the problem solving process and mechanisms work together to institutionalize data quality improvement processes [1, 20, 30, 31, 34, 38, 51, 52, 53, 69]. These research streams provide further insights into context-reflective data quality problem solving.

This research has practical implications as well. *First,* rules involved in problem solving activities need to be recorded and shared. *Second,* contexts of business processes and data processes need to be further explicated. *Third,* organizations need to pay more attention to data needs in aligning strategic, managerial, and operational levels of business processes and integrate these processes horizontally across functional areas. Particularly without connecting to strategic plans, data quality problem solving will impede future and global use of data. Data quality problem solving should be outward-bound, context-bound, and future-bound.

The organizations involved in the study also learned lessons. When practitioners reviewed data quality problems only within their domain, functional areas, and information roles, each move and all activities are considered normal activities with no problems. Most data passed all typical evaluations. Only when they reviewed the connections of multiple moves did they find problems. It is the relationship, dependency, and the connectivity between the areas, where typical data quality problems reside. The Matcom's experience, however, shows that data that exhibit violations of each data production process tend to get the attention and focus of further reviews. In short, explicating contexts is a pivotal activity in the information discourse that moves the search for a solution farther.

The contexts can be further refined in future study to differentiate and identify distinct contexts in data quality problem solving. This will take large pools of data for an empirical testing to confirm or revise the context. For our current study, we used caution as they were related constructs and treated as related. The focus here was not formalizing the specific contexts, but fleshing out their roles: how these contexts are contested, reflected, and used in rule-crafting.

Problem solving may involve several more contexts, which may complicate the PBG diagram. For example, the organization's defensive routines, politics, and cultures also can be added as major contexts. In this study, we treat these important issues as fundamental influencing and contributing factors that eventually impact five key contexts analyzed in this study. Future study is needed to further differentiate these contexts. Nonetheless, a clear example depicted here shows how an analysis can be conducted to understand contexts involved in other problem solving areas as well.

Conclusion

THIS STUDY FINDS THAT PRACTITIONERS solve data quality problems by understanding contexts in the entire information discourse on why an organization collects or creates data and how the data are stored and used. Thus, the discourse on contexts of data connects otherwise separately managed processes, that is, collection, storage, and use.

First, this study suggests that having rules and routines are common and not problematic by themselves. Not having mechanisms to record, review, and assess the rules is, however, problematic. Organizational processes change as business logic changes. Changes in rules for data to support these changes are often temporary and not recorded and shared. *Second,* paradigm-based solutions such as data integrity rules and associated technology must be recorded, understood, and used in the entire context for data quality problem solving. *Third,* the research suggests that data quality knowledge, problem contexts, and solution mechanisms must work together in data quality problem solving. *Fourth,* the research also suggests the use of data quality practitioners for incorporating both technical and organizational insights into solving data quality problems.

The research results are timely. Emerging business processes increase the visibility of data quality problems. At the same time, slack time between major processes for resolving problems is rapidly shrinking. Much more is at stake as data are used as strategic weapons beyond operational use. As is true with any rapidly expanding area, the data quality area is being flooded with experimental solutions ranging from software tools for cleansing data to organizational mechanisms such as enterprise level data quality czars. Companies have adopted such solutions with mixed results in actually resolving data quality problems in the long run. Much research is needed now for solving the backlog of data quality problems.

New information systems and changing business strategies are common scenes in today's business environments. In the thick of these unsynchronized changes, quality of data is often overlooked. Data quality is not a stated priority in most organizations. There is a need to manage all these changes in an integrated and coordinated way. Quality of organizational data could be a critical reference point for how well these various changes are coordinated for strategic and effective management of business.

NOTE

1. Espoused procedures are typically specified in the organization's statement of policy and procedures. Here, I include procedures and rules that are actually used or assumed beyond the documented policy and procedures.

REFERENCES

1. Adler, P., and Cole, R. Designed for learning: A tale of two auto plants. *Sloan Management Review, 34,* 3 (Spring 1993), 85–94.

2. Ancona, D.; Kochan, T.; Scully, M.; Van Mannen, J.; and Westney, E. *Managing for the Future: Organizational Behavior and Processes.* Saddle Stitch, OH: South-Western College Publishing, 1999.

3. Argyris, C., and Schön, D. *Organizational Learning: A Theory of Action Perspective.* Reading, MA: Addison-Wesley, 1978.

4. Argyris, C., and Schön, D. *Organizational Learning II: Theory, Method, and Practice.* Reading, MA: Addison-Wesley, 1996.

5. Bacharach, S.; Bamberger, P.; and Sonnenstuhl, W. The organizational transformation process: The micropolitics of dissonance reduction and the alignment of logics of action. *Administrative Science Quarterly, 41,* 3 (1996), 477–506.

6. Badaracco, J. *The Knowledge Link: How Firms Compete Through Strategic Alliances.* Boston: Harvard Business School Press, 1991.

7. Ballou, D.; Wang, R.; Pazer, H.; and Tayi, G. Modeling information manufacturing systems to determine information product quality. *Management Science, 44,* 4 (1998), 462–484.

8. Baskerville, R. Conducting action research: High risk and high reward in theory and practice. In E. Trauth (ed.), *Qualitative Research in IS: Issues and Trends.* Hershey, PA: Idea Group, 2001, pp. 192–217.

9. Bassellier, G.; Reich, B.; and Benbasat, I. Information technology competence of business managers: A definition and research model. *Journal of Management Information Systems, 17,* 4 (Spring 2001), 159–182.

10. Bateson, G. *Steps to an Ecology of Mind.* St. Albans, UK: Paladin, 1973.

11. Bougon, M.; Donnellon, A.; and Gray, B. Communication, meaning, and organized action. *Administrative Science Quarterly, 31,* 12 (1986), 43–55.

12. Bourdieu, P. *Outline of a Theory of Practice.* Cambridge: Cambridge University Press, 1977.

13. Bourdieu, P. *The Logic of Practice.* Stanford: Stanford University Press, 1990.

14. Chen, M. A model-driven approach to accessing managerial information: The development of a repository-based executive information system. *Journal of Management Information Systems, 11,* 4 (Spring 1995), 33–64.

15. Chiklin, S., and Lave, J. (eds.). *Understanding Practice: Perspectives on Activity and Context.* Cambridge: Cambridge University Press, 1993.

16. Codd, E. *The Relational Model for Database Management: Version 2.* Reading, MA: Addison-Wesley, 1990.

17. Cohen, M., and Bacdayan, P. Organizational routines are stored as procedural memory: Evidence from a laboratory study. *Organization Science, 5,* 4 (1995), 554–568.

18. Cohen, W., and Levinthal, D. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35,* 1 (1990), 128–152.

19. Collins, R. *Theoretical Sociology.* New York: Harcourt Brace Jovanovich, 1988.

20. Daft, R., and Huber, G. How organizations learn: A communication framework. In B.M. Staw (ed.), *Research in the Sociology of Organizations,* vol. 5. Greenwich, CT: JAI Press, 1987, pp. 1–36.

21. Daft, R., and Weick, K. Toward a model of organizations as interpretive systems. *Academy of Management Review, 9,* 2 (1984), 284–295.

22. Date, C. Referential integrity. In G. Gardaun, M. Edelberg, C. Zaniolo, H. Biller, and C. Delobel (eds.), *Proceedings of the Seventh International Conference on Very Large Data Bases.* Cannes, France: VLDB, 1981, pp. 2–12.

23. Dewey, J. *How We Think.* Mineola, NY: Dover, 1997.

24. DiMaggio, P. Interest and agency in institutional theory. In L. Zucker (ed.), *Institutional Patterns and Organizations: Culture and Environment.* Cambridge, MA: Ballinger, 1988, pp. 3–21.

25. DiMaggio, P. The iron cage revisited: Institutional isomorphism and collective rationality. In W. Powell and P. DiMaggio (eds.), *The New Institutionalism in Organizational Analysis.* Chicago: University of Chicago Press, 1991, pp. 63–82.

26. DiMaggio, P. Culture and economy. In N. Smelser and R. Swedenberg (eds.), *The Handbook of Economic Sociology.* Princeton: Princeton University Press, 1994, pp. 27–57.

27. Dougherty, D., and Hardy, C. Sustained product innovation in large, mature organizations: Overcoming innovation-to-organization problems. *Academy of Management Journal, 39,* 5 (1996), 1120–1153.

28. Ein-Dor, P., and Spiegler, I. Natural language access to multiple databases: A model and a prototype. *Journal of Management Information Systems, 12,* 1 (Summer 1995), 171–197.

29. Eisenhardt, K. Building theories from case study research. *Academy of Management Review, 14,* 4 (1989), 532–550.

30. Funk, J.; Lee, Y.; and Wang, R. Institutionalizing information quality practice. In I. Chengalur-Smith and L. Pipino (eds.), *Proceedings of the Conference on Information Quality.* Cambridge, MA: ICIQ, 1998, pp. 1–17.

31. Galbraith, J. Organization design: An information processing view. *Interfaces, 4,* 5 (1974), 28–36.

32. Giddens, A. *New Rules of Sociological Method.* Stanford: Stanford University Press, 1993.

33. Goffman, I. *Frame Analysis.* New York: Harper, 1974.

34. Hanson, J. Informal networks: The company behind the chart. *Harvard Business Review, 71,* 4 (July–August 1993), 104–111.

35. Hayek, F. Rules, perceptions, and intelligibility. In M. Hope (ed.), *Proceedings of the British Academy.* London: Oxbow Books, 1962, pp. 321–344.

36. Helfat, C. Know-how, asset complementarity and dynamic capability accumulation: The case of R & D. *Strategic Management Journal, 5,* 18 (1997), 339–360.

37. Henderson, R., and Clark, K. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms technological change and the management of architectural knowledge. *Administrative Science Quarterly, 35,* 1 (1990), 9–30.

38. Herriott, S.; Levinthal, D.; and March, J. Learning from experience in organizations. *The Theory of Economic Organizations, 75,* 2 (1985), 298–303.

39. Huang, K.; Lee, Y.; and Wang, R. *Quality Information and Knowledge.* Upper Saddle River, NJ: Prentice Hall, 1999.

40. Jacobson, A. The game of the laws. *Political Theory, 27,* 6 (1999), 769–788.

41. Katz-Haas, R., and Lee, Y. Understanding hidden interdependencies between information and organizational processes in practice. In C. Fisher and B.N. Davidson (eds.), *Proceedings of the Seventh International Conference on Information Quality.* Cambridge, MA: ICIQ, 2002, pp. 18–30.

42. Kogut, B., and Zander, U. Knowledge of the firm combinative capabilities, and the replication of technology. *Organization Science, 3,* 3 (1992), 383–397.

43. Kogut, B., and Zandler, U. What firms do? Coordination, identity, and learning. *Organizational Science, 7,* 5 (1996), 502–518.

44. Kozlowski, S., and Ostroff, C. Organizational socialization as a learning process: The role of information acquisition. *Personnel Psychology, 45,* 3 (1992), 849–874.

45. Kuhn, T. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1970.

46. Lave, J., and Wenger, E. *Situated Learning: Legitimate Peripheral Participation.* Cambridge: Cambridge University Press, 1991.

47. Lee, A. A scientific methodology for MIS case studies. *MIS Quarterly, 13,* 1 (1989), 33–50.

48. Lee, Y. Why "know why" knowledge is useful for solving information quality problems. In *Proceedings of the Americas Conference on Information Systems.* Phoenix, AZ: AIS, 1996, pp. 200–202.

49. Lee, Y., and Strong, D. Knowing-why about data processes and data quality. *Journal of Management Information Systems, 20,* 3 (Winter 2003–4), 13–39.

50. Lee, Y.; Strong, D.; Kahn, B.; and Wang, R. AIMQ: A methodology for information quality assessment. *Information & Management, 40,* 2 (2002), 133–146.

51. Levinthal, D., and March, J. A model of adaptive organizational search. *Journal of Economic Behavior and Organization, 2,* 4 (1981), 307–333.

52. Levitt, B., and March, J. Organizational learning. In W. Scott and J. Blake (eds.), *Annual Review of Sociology.* Palo Alto, CA: Annual Review, 1988, pp. 319–340.

53. MacDonald, S. Learning to change: An information processing perspective on learning in the organization. *Organization Science, 6,* 5 (1995), 557–568.

54. Madnick, S. Integrating information from global systems: Dealing with the "on- and off-ramps" of the information superhighway. *Journal of Organizational Computing, 5,* 2 (1995), 69–82.

55. Madnick, S. Database in the Internet age. *Database Programming and Design, 10,* 1 (1997), 28–33.

56. Markus, M. Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems, 18,* 1 (Summer 2001), 57–93.

57. Miles, M., and Huberman, A. *Qualitative Data Analysis: An Expanded Sourcebook.* Thousand Oaks, CA: Sage, 1994.

58. Miles, R., and Snow, C. Network organizations: New concepts for new forms. *California Management Review, 28,* 3 (1986), 62–73.

59. Newell, A. *Unified Theories of Cognition.* Cambridge: Harvard University Press, 1990.

60. Newell, A., and Simon, H. *Human Problem Solving.* Upper Saddle River, NJ: Prentice Hall, 1972.

61. Nonaka, I., and Takeuchi, H. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation.* New York: Oxford University Press, 1995.

62. O'Shaughnessy, J. *Inquiry and Decision: A Methodology for Management and the Social Sciences.* New York: Allen and Unwin, 1972.

63. Ozawa, C.P. Science in environmental conflicts. *Sociological Perspectives, 39,* 2 (1996), 219–230.

64. Pipino, L.; Lee, Y.; and Wang, R. Data quality assessment. *Communications of the ACM, 45,* 4 (April 2002), 211–218.

65. Popper, K. *The Logic of Scientific Discovery.* London: Hutchinson, 1968.

66. Pounds, W. The process of problem finding. *Psychological Review, 60,* 1 (1953), 55–63.

67. Prahalad, C., and Bettis, R. The dominant logic: A new linkage between diversity and performance. *Strategic Management Journal, 7,* 6 (1986), 485–501.

68. Quinn, J.; Baruch, J.; and Zien, K. *Innovation Explosion: Using Intellect and Software to Revolutionize Growth Strategies.* New York: Free Press, 1997.

69. Rein, M., and Schön, D. *Problem-Setting in Policy Research.* Lexington, MA: Lexington Books, 1977.

70. Rob, P., and Coronel, C. *Database Systems: Design, Implementation and Management.* Cambridge, MA: Course Technology, 2000.

71. Robey, D., and Boudreau, M. Accounting for the contradictory organizational consequences of information technology: Theoretical directions and methodological implications. *Information Systems Research, 10,* 2 (1999), 167–185.

72. Robey, D.; Ross, J.; and Boudreau, M. Learning to implement enterprise systems: An exploratory study of the dialectics of change. *Journal of Management Information Systems, 19,* 1 (Summer 2002), 17–46.

73. Schank, R. *Dynamic Memory Revisited.* Cambridge: Cambridge University Press, 1999.

74. Schön, D.A. Organizational learning. In G. Morgan (ed.), *Beyond Method.* Beverly Hills, CA: Sage, 1983, pp. 114–128.

75. Schön, D.A. *The Reflective Practitioner: How Professionals Think in Action.* New York: Basic Books, 1983.

76. Schön, D.A. *Educating the Reflective Practitioner.* San Francisco: Jossey-Bass, 1990.

77. Stokes, R., and Hewitt, J. Aligning actions. *American Sociological Review, 41,* 5 (1976), 838–849.

78. Strong, D.; Lee, Y.; and Wang, R. Data quality in context. *Communications of the ACM, 40,* 5 (1997), 103–110.

79. Suchman, L. *Plans and Situated Actions: The Problem of Human Machine Communication.* Cambridge: University of Cambridge Press, 1987.

80. Susman, G. Action research: A sociotechnical systems perspective, In G. Morgan (ed.), *Beyond Method: Strategies for Social Research.* Newbury Park, CA: Sage, 1983, pp. 95–113.

81. Taylor, C. To follow a rule. In C. Calhoun, E. Lipuma, and M. Psotone (eds.), *Bourdieu: Critical Perspectives.* Chicago: University of Chicago Press, 1993, pp. 45–60.

82. Van Maanen, J. *Tales of the Field: On Writing Ethnography.* Chicago: University of Chicago Press, 1988.

83. Vessey, I., and Conger, S. Learning to specify information requirements: The relationship between application and methodology. *Journal of Management Information Systems, 10,* 2 (Fall 1993), 177–202.

84. Von Hippel, E., and Tyre, M. How learning by doing is done: Problem identification in novel process equipment. *Research Policy, 24,* 1 (1995), 1–12.

85. Wand, Y., and Wang, R. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39,* 11 (1996), 86–95.

86. Wang, R.; Lee, Y.; Pipino, L.; and Strong, D. Manage your information as a product. *Sloan Management Review, 39,* 4 (1998), 95–105.

87. Weick, K. Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly, 3,* 38 (1993), 357–381.

88. Weick, K. *The Social Psychology of Organizing.* Reading, MA: Addison-Wesley, 1979.

89. Yin, R. *Case Study Research: Design and Methods.* Newbury Park, CA: Sage, 1989.

Appendix A. Data Quality Projects in Selected Organizations

Agcom: Project Episode

AT AGCOM, A MAJOR AGRICULTURAL COMMODITY COMPANY, multiple databases in different functional areas served different business purposes. Historically, business processes, expectations, and understanding of the operation of the company's business differed by department. The production department was interested in how beef was cut. The cutting guidelines were based mainly on the weight of the beef. Different breeds of the same weight were cut in the same way. For example, an 800-pound Angus (of a breed that has fat layers in between muscles, is tender, and returns a higher premium) and 800 pounds of any other type of beef were cut in the same way. There was one process for all 800-pound beef, just as there was one process for all 1,400-pound beef. What mattered to production was how heavy the animals were. The sales department, however, was indifferent to weight. What was important to them was the 20–30 percent premium for Angus cows versus 5 percent for some other breed. Thus, the sales department needed to track and record different breeds and their prices. The production department tracked weight and cutting processes, whereas the sales department tracked how prices for different breeds compared. The information from the production system was accurate for the purposes of production. The information from the orders management systems (sales) was accurate for sales purposes. Both systems contained information that was accurate in its particular context, but whose information was not useful, usable, or accessible for making business decisions at a company-wide (global) level. The result was a collision. The challenge for Agcom management was how to interpret and use high volumes of accurate, but unusable, information for its critical business decision-making. It was nearly impossible to evaluate Agcom's profit and loss meaningfully, let alone comprehend and plan future business strategies. An Agcom senior manager summed up their problems: "The things we produce are not the things we sell." Obviously, all departments at Agcom needed to recognize the overarching view at a company-wide level to resolve the mismatching and disjointed information problems.

Healthcom: Project Episode

Healthcom was revisiting detailed data flows from the point of data generation through data collection and storage to data usage. Healthcom collected data about patients, for example, whether a patient was a smoker or a nonsmoker. One attribute of the patient entity or (i.e., one column of the "Patient" table) was designated for this purpose. Valid values for the field included "yes" and "no" to indicate that the patient was a smoker or a nonsmoker. Several years later, the hospital decided to include information from living wills in the database. Not realizing the original purpose of the column, and also recognizing the binary values as yes/no, one department used the same column to indicate whether or not a patient had a living will. Instead of adding a new column to the table and incorporating the new column into their data

model, the department used this attribute (smoker/nonsmoker) as a temporary column to record the existence or nonexistence of living wills. Later, hospital information quality initiatives discovered that the field contained values from two different domains, violating data model rules.

Appendix B. Sample Results from Data Integrity Project

User-Defined Integrity - Records Meeting Condition

Change Fields

CASE_BK	DOL_BK	AST_CUST_ID	PROD_ID	WK_ID
-17	0	69000	90095	199841
-7	0	69000	90096	199841
-6	0	155450	1943	199738
-1	0	348000	90097	199643
-12	0	465600	90095	199633
-49	0	541800	90095	199710
-2	0	541800	90097	199645
-1	0	562800	90917	199815
-4	0	562800	90917	199821
-5	0	607500	90097	199648
-42	0	763800	90095	199745
-12	0	763800	12725	199741
-12	0	763800	12725	199744
-10	0	763800	32726	199744
-7	0	763800	52726	199744
-2	0	786028	2095	199709
-28	0	805500	90096	199650
-28	0	805500	90097	199650
-1	0	960432	90095	199639
-1	0	960432	90096	199639
-2	0	960667	90917	199812
-12	0	960706	90917	199820
-57	0	999999	90095	199737
-1	0	999999	90095	199833
-1	0	999999	90097	199651
-1	0	999999	90917	199819