

---

# Knowing-Why About Data Processes and Data Quality

YANG W. LEE AND DIANE M. STRONG

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 Publishers, 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 Cochair for International Conference on Information Quality (ICIQ) in Cambridge. She received her Ph.D. from MIT.

DIANE M. STRONG is an Associate Professor in the Management Department at Worcester Polytechnic Institute and Director of the MIS program. She received her Ph.D. in Information Systems from Carnegie Mellon University. Dr. Strong's research centers on data and information quality and on the organizational impacts of MIS application systems, especially ERP systems. Her publications have appeared in leading journals such as *Communications of the ACM*, *Journal of Management Information Systems*, *ACM Transactions on Information Systems*, *Journal of Systems and Software*, and *Information & Management*. She is a member of AIS, ACM, and INFORMS, is serving on the AIS Council, and was Program Chair for AMCIS in Boston.

**ABSTRACT:** Knowledge about work processes is a prerequisite for performing work. We investigate whether a certain mode of knowledge, knowing-why, affects work performance and whether the knowledge held by different work roles matters for work performance. We operationalize these questions in the specific domain of data production processes and data quality. We analyze responses from three roles within data production processes, data collectors, data custodians, and data consumers, to investigate the effects of different knowledge modes held by different work roles on data quality. We find that work roles and the mode of knowledge do matter. Specifically, data collectors with why-knowledge about the data production process contribute to producing better quality data. Overall, knowledge of data collectors is more critical than that of data custodians.

**KEY WORDS AND PHRASES:** data processes, data quality, knowing-why, knowledge, knowledge modes, work roles.

---

KNOWLEDGE IN GENERAL IS A PREREQUISITE FOR WORK. When we discuss knowledge for work, we mean knowledge about work processes and the resulting work performance. For a data production process, the work domain for this study, knowledge about work processes encompasses knowledge about the three key processes within a data production process: collection of raw data, storage and maintenance of data in computer systems, and user retrieval and manipulation of data [43]. Knowledge about work performance is knowledge about producing high-quality data from data production processes.

Whereas conventional research has recognized the importance of knowledge and the corresponding organizational learning [1, 2, 10, 13, 17, 49], actual relationships and patterns of various modes of knowledge for specific performance have not been explored. Little attention has been paid to understanding how one's scope and depth of knowledge about organizational work processes contributes to better process performance. For example, would an information systems (IS) group with more knowledge about a data production process produce better quality data for data consumers?

We identify three modes of knowledge pertinent for shaping organizational capabilities, knowing-what, knowing-how, and knowing-why. *Knowing-what* is rich in information on activities and relationships, and thus, is a fact-based mode of knowledge. For example, equipped with and using knowing-what, one can identify activities and facts associated with a data production process. *Knowing-how* is characterized as step-by-step procedures executable in a specific sequence [9, 19], and thus, is a procedure-based mode of knowledge. Using knowing-how, one can solve routine data quality problems, as these problems involve well-defined procedures and rules. These two modes of knowledge, knowing-what and knowing-how, draw upon existing distinctions, such as the one between declarative and procedural knowledge [26, 40], tacit and explicit knowledge [41, 30, 31, 32, 35], and component and architectural knowledge [11].

Initial proposals for a hybrid third mode of knowledge stretch back at least to John Locke. As a third mode of knowledge, Nosofsky [33] suggests representation process, Wisniewski and Medin [50] suggest selective attention, and Sakmann [38] suggests axiomatic knowledge, why-knowledge. *Knowing-why* is characterized as contextual knowledge that enables inquirers to direct questions based on understanding relevant purposes and underlying principles. We explicitly include the aspects of knowledge that question contextual reasons and axiomatic principles underlying the work practices in organizations, knowing-why [28].

Knowing-why is gained from experience and understanding of the objectives and cause-effect relationships underlying the activities (knowing-what) and procedures (knowing-how) involved in work processes in organizations. Knowing-why is the contextual and axiomatic mode of knowledge that raises deeper inquiries than knowing-what and knowing-how. With the knowing-why mode of knowledge, organizational members are more likely to raise a more meaningful inquiry beyond mechanical and routine work. For example, they understand the deeper axiomatic principles of why they collect, store, and use data in their organizations.

Our inquiry into knowing-why is guided by the classical findings that creating and using organizational knowledge is difficult in the absence of organizational inquiry [4, 15]. Mismatches between expectations and realities, and difficulties experienced in problem solving situations lead members to initiate organizational inquiries [39]. As knowledge is socially constructed [8, 22] and, thus, context-dependent [7, 24, 25, 27], we posit that the knowledge about the context of work processes—that is, knowing-why—enables organizational members to raise meaningful organizational inquiries about questionable or problematic data quality.

Knowing-why may be the underlying mode of connection between the fact-based mode, knowing-what, and the procedure-based mode, knowing-how. Particularly, in the context of data quality-related knowledge, knowing-why is a critical mode of knowledge because understanding principles and purposes behind the routine data production activities and procedures provides one with the capability for questioning poor quality data and suggesting new solutions.

Controversy remains about how the modes of knowledge are involved in work practice. Some [33, 38, 50] argue for no distinction between the modes of knowledge, but accept the interaction among the modes of knowledge. For example, preexisting declarative knowledge plays a role in absorbing and creating new knowledge in innovation activities [14]. Our view is that the modes of knowledge are conceptually separate, but, during work, they are used interactively. We offer a clear explanation that we need to view knowledge as constituting multiple modes of operation that are used in combination. Clearly, there is a need for further empirical investigation into how knowledge works in specific work domains.

The work domain for our inquiry into knowing-why is data production processes. A data production process involves different work roles and subprocesses to accomplish the work of producing high-quality data for data consumers. The differential knowledge of these work roles may explain the problems that many organizations have with producing high-quality data. Not all knowledge is shared and used by all organizational members. The process of communication and integration of cognitions leads to the formation of collective organizational knowledge [16]. Organizations with seemingly knowledgeable IS groups and well-established organizational rules, procedures, and routines, which are the evidence of organizational knowledge, are not exempt from producing poor-quality data and being affected by them [5, 29]. Whereas an IS group is typically very knowledgeable about storage and maintenance of data in its systems, it may know little about how and why data consumers use data. Knowledge about user processes may help IS groups to understand the reasons why they store and maintain organizational data and thus contribute to the production of higher-quality data for data consumers.

In general, the IS groups we investigated in this study are knowledgeable. Our question is not about existence of organizational knowledge, but about the lack of a critical mode of knowledge, knowing-why. Different modes of knowledge direct the inquirer to ask different questions in search of solutions to the questionable quality of data. The solutions found, therefore, are bound to the questions raised. We examine how utilizing different modes of knowledge is related to producing data with different quality.

An inquiry into poor quality data rests on the three modes of knowledge. In the area of data quality, *knowing-what* is defined as the understanding of the activities involved in data production processes. *Knowing-how* is defined as understanding procedures to handle known data quality problems. *Knowing-why* is defined as the ability to analyze underlying principles and discover previously unknown data quality problems or solutions. The three modes of knowledge become further distinguishable as an inquirer assumes a specific work role, data collector, data custodian, or data consumer. *Data collectors* provide initial input of organizational data; *data custodians* are responsible for storage and maintenance of the data; and *data consumers* utilize the data for further integration, aggregation, presentation, and interpretation of data.

This paper contributes by advancing theoretical development both conceptually and empirically. It also advances theories at both general and domain-specific levels. Finally, this research offers practical insights.

First, it advances the concept and effects of knowledge in organizations and it differentiates explicitly that knowledge includes contextual and axiomatic knowledge, that is, knowing-why, in addition to knowing-what and knowing-how. It also operationalizes these ideas via a questionnaire based on the constructs we conceptualized. The data collected from the field are used to investigate the differential effects of modes and domains of knowledge. As a result, this study makes both conceptual and empirical contributions to our understanding of knowledge. Second, these contributions apply at both the general level and the domain-specific level by anchoring the study in the domain of data process knowledge and data quality performance. By starting from general knowledge theory, we assure that our results are not just domain-specific. Since measuring knowledge only makes sense by grounding the questionnaire at the domain-specific level so that questionnaire respondents can respond clearly, the results offer insights into the domain-specific area, that is, data process roles, knowledge modes, and data quality performance. Third, by anchoring the questionnaire in a specific domain, this study provides practical insights of use to organizational managers in addition to the theoretical insights.

## Background and Hypotheses

---

TO FOCUS OUR INQUIRY INTO KNOWING-WHY, we develop four hypotheses ranging from general to more specific. We start with the nature of knowledge about data production processes. Then we develop hypotheses about the relationship between knowledge and our performance measure data quality, and, subsequently, whether that relationship differs by work role. Finally, we focus our attention specifically on knowing-why and its relationship to data quality and whether that relationship differs by work roles.

### Knowledge About Data Processes

Processes to produce data have many similarities to processes that produce physical products, and should be viewed as producing data products for data consumers [6,

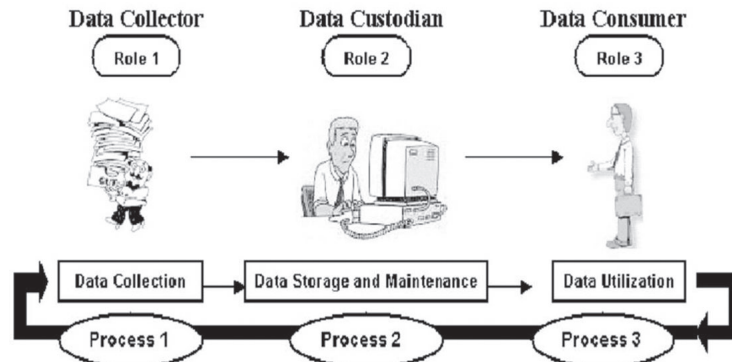


Figure 1. Data Production Process

18, 20, 23, 34, 42, 44, 45, 46, 48]. As evidenced in the production of physical products, a data production process, too, is divided into distinctive work processes, collection, storage, and use work processes [21, 43]. We identify three roles within a data production process: data collectors (people, groups, or other sources who generate information); data custodians (people who manage computing resources for storing and processing data); and data consumers (people or groups who use data). Each role is associated with a process: (1) data collectors are in charge of data-collection processes, (2) data custodians of data storage and maintenance, and (3) data consumers of data-utilization processes, which may involve retrieval of data, additional data aggregation, and integration (Figure 1).

Knowledge about work processes represents scope of knowledge, that is, how much of the entire data production process is known. The three modes of knowledge capture aspects of knowledge depth, from knowing the activities to procedural know-how, and to contextual and underlying purposes. In sum, knowledge about data processes consists of knowledge held about work processes, for example, data collection, data storage, and data utilization, and of the modes of knowledge held, for example, knowing-what, knowing-how, and knowing-why. Hypothesis 1 states this conceptualization of the nature of knowledge about data production processes.

*Hypothesis 1. Knowledge about data processes is distinguishable into (1) knowledge about distinctive work processes within the data production process, data collection, data storage, and data utilization, and (2) the three modes of knowledge, knowing-what, knowing-how, and knowing-why.*

## Relationships Between Knowledge and Data Quality

The purpose of data production processes is to produce data for data consumers. Data of high quality means data that are fit for use by these data consumers [43, 47]. Fitness for use involves multiple dimensions.

We focus on five dimensions of data quality in this research: accessibility, relevancy, timeliness, completeness, and accuracy. *Accessibility* denotes the extent to which data

are available, or easily and quickly retrievable. *Relevancy* denotes the extent to which data are applicable and helpful for the task at hand. *Timeliness* denotes the extent to which the data are sufficiently up-to-date for the task at hand. *Completeness* denotes the extent to which data are not missing and are of sufficient breadth and depth for the task at hand. *Accuracy* denotes the extent to which data are correct and free-of-error [47]. Each of these quality dimensions is a performance goal of the data production process. That is, the overall goal is to produce accurate, complete, and timely data that are accessible to data consumers and relevant to their tasks.

As knowledge is an indicator for organizational learning and performance in general, knowledge about a data production process is an indicator for data quality performance. As with producing any other product, producing high-quality data requires organizational knowledge about data production processes. Knowledge about a data production process indicates the competency for raising and acting upon inquiries for data quality performance, that is, producing high-quality data. We establish a general rubric of relationship between knowledge and data quality performance.

*Hypothesis 2. Knowing-what, knowing-how, and knowing-why about the three data production processes are associated with higher data quality along all five dimensions (accuracy, completeness, accessibility, timeliness, and relevancy).*

### Relationship Between Knowledge and Data Quality by Work Role

Individuals who assume a work role are more familiar with and capable of knowing about their immediate work process. Data collectors are knowledgeable about data collection processes, whereas data custodians, such as IS groups, are knowledgeable about data storage and maintenance. Data consumers are knowledgeable about data utilization. Knowing more about certain work processes may lead to performing better in one area than in others. For example, data collectors may know more about collecting accurate and complete data. Data custodians may know more about making data accessible. Data consumers may know more about making data relevant to their task.

*Hypothesis 3. The relationship between knowledge and data quality differs by work roles within the data production process.*

- *Data collectors' knowledge is highly associated with accuracy and completeness of data.*
- *Data custodians' knowledge is highly associated with completeness, accessibility, and timeliness of data.*
- *Data consumers' knowledge is highly associated with relevancy of data.*

### Relationship Between Knowledge and Data Quality by Work Role and Knowledge Mode: Examining Knowing-Why

Modes of knowledge indicate aptitudes of inquiries in problem solving. Roles that organizational members assume and perform indicate their domain areas of compe-

tency. Combining the two indicators, modes of knowledge and roles assumed in a data production process form a cue for associations with different data quality dimensions.

Different modes of knowledge lead to raising inquiries on different aspects of data quality. Knowing-what implies that one can identify activities and facts associated with the data production process. Knowing-how implies that one can solve routine problems easily. Routine problem solving involves executing well-defined procedures and rules for known problems. Holding and using know-why implies that one can solve new problems and can raise inquiries that deal with contextual questions of data quality.

*Hypothesis 4a. The relationship between knowledge and data quality differs by both modes of knowledge and work roles within the data production process.*

In particular, knowing-why indicates the competency for solving nonroutine problems and understanding contextual knowledge about data production processes. In a sense, knowing-why can link knowing-what with knowing-how. This view leads to the assumption that knowing-why also indicates the competency for solving data quality problems that are important for data consumers. Because competency requires at least minimal knowledge of work activities (knowing-what) and procedures for doing them (knowing-how), knowing-why is likely to be a critical differentiator for performance.

*Hypothesis 4b. The relationship between knowledge and data quality differs more between work roles for the knowing-why knowledge mode than for knowing-what and knowing-how knowledge modes.*

## Research Method

---

### Research Sites

SIX COMPANIES SERVED AS DATA COLLECTION SITES for this research, two financial institutions, three health-care organizations, and a consumer product manufacturing company. In all six companies, the research focus was the quality of their customer activity data. Specifically, in the financial institutions, the focus was the quality and production of data about investors and their investment activity. In the health-care organizations, it was the quality and production of data about patients and the health-care services they used. In the manufacturing firm, it was customers and their purchasing activity.

### Sample

The sample consisted of 155 respondents from these six companies. Respondents were selected to ensure coverage of all the roles in the data production process (data collectors, data custodians, and data consumers) for the customer activity data being studied. The sample consisted of 48 data collectors, 45 data custodians, and 62 data consumers.

## Procedures

Two sessions were held at each company. All people selected for potential participation were invited to the first session. At this session, an overview of the research and a questionnaire was presented. This overview motivates subjects to complete the questionnaire carefully and thoroughly. Pilot tests of the questionnaire indicated that such a session was necessary to ensure that the questionnaire was completed in full with quality responses. The respondents were instructed to fill out the questionnaire for the customer activity database and its associated data production process. After the overview, respondents completed the questionnaire before leaving the session. This ensured nearly 100 percent response rate. The second session provided summary feedback on the questionnaire results and invited participants to provide comments and explanations.

## Independent Variables

Knowledge is the independent variable. Knowledge is measured at the level of the three modes, knowing-what, knowing-how, and knowing-why. Each of these three modes was measured for the three data production processes, data collection, data storage, and data utilization. Thus, there are nine sub-measures of knowledge.

## Dependent Variables

Data quality is the dependent variable. Data quality is measured at the level of data quality dimensions [47] using five key dimensions of importance to data consumers—accuracy, completeness, timeliness, relevance, and accessibility.

## Measurement Instrument

All independent and dependent variables were collected using a questionnaire completed by each respondent. Four to seven items were included for each measure. (See the Appendix for measures and items.) For the nine knowledge measures, the scale was 1 to 10, where 1 is “very small extent,” 5 is “average,” and 10 is “very large extent.” For the five data quality measures, the scale is 0 to 10, where 0 is “not at all,” 5 is “average,” and 10 is “completely.”

The data quality measures have been well tested in previous research [29, 46]. Development of the knowledge measures proceeded as follows. Researchers generated items for each construct to be measured. These items were reviewed for construct coverage by other researchers, and also reviewed for meaningfulness by data collectors, data custodians, and data consumers in organizations. A pilot study was conducted to test item understandability by respondents, construct reliability, and questionnaire procedures. Further details about questionnaire development can be found in Huang et al. [21].



## Analyses

Statistical analysis is performed using SPSS for Windows. Cronbach's alpha provided evidence of the reliability for the measures. These range from 0.87 to 0.96, which was acceptable reliability with no problem constructs. (See the Appendix for the reliability indicators for each measure.) Factor analysis is employed to test whether the knowledge measures separated into the nine knowledge sub-measures. For the knowledge about processes and the knowledge modes in Hypothesis 1 to hold, the factor analysis must produce nine factors.

The analysis is performed at the level of individual respondents. The findings presented below for the hypotheses are based on simple correlations among independent and dependent measures for each role. Correlations are calculated among the nine knowledge sub-measures and the five data quality sub-measures. At this stage in the research and measurement of knowledge in organizations, simple correlations are sufficient for exploring the efficacy of measuring knowledge by questionnaire and using these measures to understand knowledge patterns and their relationship to one indicator of organizational performance, the quality of organizational data.

## Findings

---

OUR CENTRAL INTERESTS IN THIS STUDY ARE (1) distinctions among different knowledge modes and among different data production processes, (2) the relationship between knowledge and data quality performance, (3) how this relationship differs by work role, and (4) the effects of knowledge modes, especially why-knowledge. The findings for these four hypotheses are discussed below.

### Hypothesis 1

Hypothesis 1 posits distinctions among three knowledge modes—knowing-what, knowing-how, knowing-why—and knowledge about the three data production processes—data collection, data storage, and data utilization. The factor analysis shows distinctions only among knowledge about the data production processes (Table 1); that is, data-related knowledge falls into three distinct constructs, namely, collection, storage, and utilization processes. Knowing-what, knowing-how, and knowing-why about collection, storage, and use are intermixed within the respective collection, storage, and use factors.

### Hypothesis 2

Hypothesis 2 posits a positive relationship between knowledge and data quality. Hypotheses 3 and 4 posit detailed relations under the rubric of Hypothesis 2. The association of knowledge with higher levels of data quality is clearly confirmed with the analysis. All modes of knowledge are associated with data quality (Table 2). This

Table 1. Knowledge Factor Analysis Results

Rotated Component Matrix			
Item label	Data storage component	Data collection component	Data utilization component
KWYS06	0.827		
KHWS03	0.795		
KWYS11	0.783		
KHWS02	0.767		
KHWS01	0.759		
KHWS04	0.753		
KWTS03	0.753		
KWYS03	0.740		
KWYS04	0.719		
KHWS05	0.716		
KWYS10	0.707		
KWTS04	0.706	0.473	
KWTS06	0.692		
KWTS05	0.689		
KWYS08	0.676		
KWYS02	0.603		0.504
KWTS01	0.472		0.464
KWTS02	0.440	0.431	0.414
KWYC02		0.780	
KWYC04		0.780	
KHWC02		0.774	
KHWC03		0.771	
KWTC03		0.763	
KWTC04		0.748	
KHWC04		0.728	
KWTC01		0.727	
KHWC01		0.719	0.408
KWTC06		0.704	
KHWC05		0.694	
KWYC05		0.680	
KWYC06		0.656	
KWTC02		0.636	
KWTU05			0.744
KWTU04			0.734
KWTU03			0.713
KHWU02		0.447	0.693
KWYU06			0.691
KHWU01			0.691
KWTU01			0.690
KHWU03			0.689
KHWU05	0.421		0.675
KWTU02		0.430	0.670
KWYU03			0.657

*(continues)*

Rotated Component Matrix			
Item label	Data storage component	Data collection component	Data utilization component
KWYU07			0.594
KWYU15			0.593
KWYU17			0.564
KWYU16			0.555
Eigenvalue	27.478	3.118	2.111
Percent variance	58.5	6.6	4.5
Cumulative percent variance	58.5	65.1	69.6

*Notes:* KWT—knowing-what; KHW—knowing-how; KKY—knowing-why, about data; S—storage; C—collection; U—utilization processes; values < 0.400 are shown as blanks; extraction method: principal component analysis; rotation method: varimax with Kaiser normalization.

supports the hypothesis that knowledge is an important prerequisite for producing high-quality data.

### Hypotheses 3

Hypothesis 3 posits that the relationship between knowledge and data quality differs by work roles within the data production process. The correlational analysis clearly shows that the knowledge held by different roles differentially impacts dimensions of data quality (Table 3).

Specifically, the knowledge held by *data collectors* plays a key role in data quality. All knowledge modes for all three data production processes are highly correlated with the dimensions of data quality except for the timeliness dimension. From the statement of Hypothesis 3, the highly significant correlations with accuracy and completeness were expected. The additional, highly significant correlations with accessibility and relevancy indicate the importance of the knowledge of data collectors. While we cannot assume causality, data collector's knowledge of what, how, and why data are collected, stored, and used appears to be critical for achieving accurate, complete, accessible, and relevant data.

*Data custodians'* knowledge is most highly correlated with accuracy, completeness, and timeliness. The efforts of data custodians to ensure that all fields are completely filled, and that processing is completed on time support these findings. The highly significant correlations with accuracy, rather than accessibility, is surprising. The hypothesized findings for accessibility are there, but weak, because the knowing-how correlations are significant, but not highly significant. We did not expect the accuracy finding because we assumed that data accuracy is most influenced by the accuracy of the data as collected, rather than by work activities of data custodians.

Table 2. Knowledge Modes and Data Quality

Knowledge mode	Domain process	Accuracy	Completeness	Accessibility	Timeliness	Relevancy
Knowing-what	Collection	0.365**	0.409**	0.386**	0.325**	0.444**
	Storage	0.436**	0.488**	0.461**	0.397**	0.496**
	Utilization	0.356**	0.416**	0.397**	0.279**	0.501**
Knowing-how	Collection	0.361**	0.418**	0.397**	0.371**	0.445**
	Storage	0.417**	0.467**	0.415**	0.388**	0.420**
	Utilization	0.379**	0.416**	0.405**	0.344**	0.428**
Knowing-why	Collection	0.249**	0.309**	0.284**	0.260**	0.362**
	Storage	0.306**	0.367**	0.353**	0.314**	0.365**
	Utilization	0.349**	0.366**	0.342**	0.295**	0.407**

Notes:  $N = 155$ ; \*\*  $p < 0.01$ ; \*  $p < 0.05$ .

Table 3. Work Role and Data Quality

Data role	Knowledge mode	Domain process	Data quality performance dimensions					
			Accuracy	Completeness	Accessibility	Timeliness	Relevancy	
Data collector (N = 48)	Knowing-what	Collection	0.463**	0.500**	0.518**	0.201	0.649**	
		Storage	0.476**	0.472**	0.480**	0.171	0.578**	
		Utilization	0.428**	0.469**	0.480**	0.137	0.572**	
	Knowing-how	Collection	0.399**	0.411**	0.498**	0.210	0.555**	
		Storage	0.477**	0.468**	0.556**	0.275	0.588**	
		Utilization	0.360**	0.388**	0.447**	0.121	0.484**	
	Knowing-why	Collection	0.345**	0.328**	0.373**	0.194	0.470**	
		Storage	0.443**	0.465**	0.560**	0.293*	0.506**	
		Utilization	0.489**	0.467**	0.482**	0.329*	0.547**	
	Number of highly significant correlations			9	9	9	0	9
	Data custodian (N = 45)	Knowing-what	Collection	0.316**	0.393**	0.387**	0.397**	0.125
			Storage	0.480**	0.564**	0.527**	0.516**	0.461**
Utilization			0.403**	0.523**	0.470**	0.319**	0.532**	
Knowing-how		Collection	0.291	0.378*	0.329*	0.426**	0.252	
		Storage	0.403**	0.486**	0.355*	0.508**	0.328*	
		Utilization	0.347**	0.432**	0.345*	0.382*	0.346*	
Knowing-why		Collection	0.101	0.220	0.187	0.236	0.088	
		Storage	0.208	0.265	0.184	0.218	0.116	
		Utilization	0.217	0.206	0.154	0.143	0.205	
Number of highly significant correlations			5	5	3	5	2	

(continues)

Table 3. Continued

Data role	Knowledge mode	Domain process	Data quality performance dimensions				
			Accuracy	Completeness	Accessibility	Timeliness	Relevancy
Data consumer ( <i>N</i> = 62)	Knowing-what	Collection	0.253*	0.262*	0.245	0.258*	0.442**
		Storage	0.286*	0.325*	0.358**	0.329**	0.409**
		Utilization	0.250	0.285*	0.289*	0.285*	0.430**
	Knowing-how	Collection	0.302*	0.355**	0.315*	0.338**	0.440**
		Storage	0.275*	0.311*	0.286*	0.223	0.273*
		Utilization	0.344**	0.340**	0.361**	0.358**	0.405**
	Knowing-why	Collection	0.194	0.259*	0.220	0.220	0.403**
		Storage	0.112	0.153	0.215	0.193	0.311*
		Utilization	0.269*	0.303*	0.298**	0.276*	0.366**
	Number of highly significant correlations			1	2	3	3

Notes: *N* = 155; \*\* *p* < 0.01; \* *p* < 0.05.

For *data consumers*, their knowledge across modes and production processes is most highly correlated with the relevancy dimension, as hypothesized. Although the other dimensions of data quality can often be objectively defined, it is only the data consumers who know whether the data are relevant.

Thus, Hypothesis 3 is supported. There are clear differences in the relationship between knowledge and data quality depending on the work role of the knowledge holder. By data quality dimensions, data collectors' knowledge is highly correlated with accuracy, completeness, accessibility, and relevancy, but not timeliness. Data custodians' knowledge is most highly correlated with accuracy, completeness, and timeliness, although the number of highly significant correlations is lower than those of data collectors. Data consumers' knowledge is most highly correlated with relevancy.

#### Hypothesis 4

Hypothesis 4a further investigates differences in the relationship between knowledge and data quality by work role. We examine the findings in Table 3 from the view of knowledge modes for each work role. For data collectors, all knowledge modes are highly correlated with data quality dimensions. The key observation, however, is that *data collectors* are the only role for which *knowing-why* is clearly important for most dimensions. For *data custodians*, *knowing-what* has the most highly correlated results. For *data consumers*, *knowing-how* has the strongest results, although *knowing-what* is a close second.

The differing importance of the three modes of knowledge is most pronounced for data custodians, the IS professionals responsible for storing and maintaining data. In general, data custodians' *knowing-what* is most significantly correlated with achieving high-quality data. Their *knowing-how* is important for some data production processes and for some data quality dimensions, but their *knowing-why* is *not* correlated with high-quality data.

Hypothesis 4b investigates the role of why-knowledge. Since *knowing-why* is the critical knowledge mode for meaningful problem solving, we further examine *knowing-why* by work roles and its effect on the data quality dimensions (Table 4). Overall, *data collector's knowing-why* is highly associated with all dimensions in all production processes except for timely collection. *Data custodian's knowing-why* is *not* associated with any data quality dimensions. *Data consumer's knowing-why* has a mixed degree of association, with relevancy as one dimension that is highly associated with data consumer's *knowing-why* about collection and utilization. These differences between work roles for the *knowing-why* knowledge mode are more pronounced than the differences between work roles for *knowing-what* and *knowing-how*.

For four of the five data quality dimensions, it is most important that data collectors know why the data are stored and used, that is, why the data are needed. This finding about data collectors' *knowing-why* about data storage and data utilization processes is worth noting. These processes are not a data collector's immediate process domain responsibility. *Knowing-why* about not directly assigned processes holds the key to effective use of *knowing-what* and *knowing-how* about the entire data production

Table 4. Knowing-Why By Work Roles

Data role	Knowledge mode	Domain process	Data quality performance dimensions				
			Accuracy	Completeness	Accessibility	Timeliness	Relevancy
Data collector ( <i>N</i> = 48)	Knowing-why	Collection	0.345**	0.328**	0.373**	0.194	0.470**
		Storage	0.443**	0.465**	0.560**	0.293*	0.506**
		Utilization	0.489**	0.467**	0.482**	0.329*	0.547**
		Number of highly significant correlations	3	3	3	0	3
Data custodian ( <i>N</i> = 45)	Knowing-why	Collection	0.101	0.220	0.187	0.236	0.088
		Storage	0.208	0.265	0.184	0.218	0.116
		Utilization	0.217	0.206	0.154	0.143	0.205
		Number of highly significant correlations	0	0	0	0	0
Data consumer ( <i>N</i> = 62)	Knowing-why	Collection	0.194	0.259*	0.220	0.220	0.403**
		Storage	0.112	0.153	0.215	0.193	0.311*
		Utilization	0.269*	0.303*	0.298**	0.276*	0.366**
		Number of highly significant correlations	0	0	1	0	2

Notes: *N* = 155; \*\* *p* < 0.01; \* *p* < 0.05.



process. This finding is direct evidence for the generally assumed benefit of cross-functional knowledge.

## Discussion

---

THE RUBRIC HYPOTHESIS THAT KNOWLEDGE MATTERS for work performance is confirmed as generally reflected in common sense. In our findings, data process knowledge and data quality work performance are highly correlated.

To strategically create and use knowledge for performance, we need to understand the inner workings and nonperformance of knowledge. That is, we need to delve below the aggregate-level high correlations and understand aspects of knowledge at work for work performance. We need to treat knowledge not as a sacred black box but as a system that may have subsystems interacting among each other that produce different performance results.

In our research, interesting results are discovered by teasing out knowledge into three modes and situating all modes of knowledge held by different work roles in a specific work context of a data production process designed to produce data of high quality. We further examined the lack of certain modes of knowledge, and compared and examined the patterns of weak and highly significant correlations.

## Modes of Knowledge

The most important finding about modes of knowledge lies in two related points. First, all modes of knowledge, combined about all processes, held by all roles, contribute to the overall data quality. Second, each mode of knowledge about a process held by a role contributes to a specific performance measure, a data quality dimension.

Based on the analysis of the questionnaire data and after-survey discussion sessions, we conclude that pure forms of any modes of knowledge are seldom exhibited. The combined knowledge of all three modes gives one's knowledge the power of application in data production processes. Knowing-what provides directions to knowing-how. Knowing-how provides implementable steps and action sequences that raise the level of applicability of knowing-what. Knowing-why provides the underlying reasons for connecting knowing-what and knowing-how in a specific context of solving data quality problems. As Raelin [36] explains about the difficulties of separating modes of knowledge, knowledge is "fluid" and it shifts between construction and transformation. The modes of knowledge, however, become differentiable as they are used for a specific performance goal, in a specific domain, by a specific role: achieving high data quality in the data production process by three data roles.

## Knowing-Why

We began this study with a question of whether the existence or lack of a certain mode of knowledge is associated with different data quality results. Our assumption

underlying the question was recognizing the critical role that knowing-why might play in solving data quality problems.

The discovery that data custodians' knowing-why is not associated with producing high-quality data disconfirms the conventional notion that data custodians working as IS professionals must understand the contextual mode of knowledge to perform their task. Specifically, data custodian's knowing-why is *not* significantly associated with achieving high data quality. Data consumer's knowing-why is highly associated with selected dimensions of data quality.

Instead, we find that data collector's knowing-why is the most critical prerequisite for high data quality across data production processes. Data collector's knowing-why is more highly associated with high data quality than that of data consumers. This critical role of data collectors should be recognized and exploited in organizations.

Based on our findings, organizations can make a strategic decision as to which work roles and modes of knowledge need to be enhanced to improve a specific dimension of data quality. For example, to achieve high accuracy, data collector's knowing-why about the utilization process is most critical. Data custodian's knowing-why is not associated with dimensions such as completeness. On the other hand, relevancy calls for data collector's knowing-why about the data utilization process. Making a completely filled database does not require data custodian's knowing-why, but collecting relevant data requires contextual knowledge, knowing-why about data utilization by both data consumers and data collectors.

## Roles, Knowledge, and Performance

The production process in which members work and the role that members play set the specific context for focusing on immediate data quality concerns, thereby asking specific questions that utilize a specific mode of knowledge. There are three distinctive but related questions that the three data roles ask. Data collectors ask *why* do people need these data; data custodians ask *what* data should they be storing; and data consumers ask *how* to use these data.

For example, data custodians pay more attention to completeness of data. At the time of data collection in a hospital case, data collectors did not fill the data field designated for diagnosis. Not filling in a record has serious implications on the tasks by both data custodians and data consumers. To data custodians, missing fields may mean that the database violates the data integrity rules [12, 37]. If the fields are required in the database, a value in the right range has to be filled. Data custodians specify these completeness rules so that records are rejected and returned to data collectors to fill any missing fields. For fields critical to the database or to data consumers, data custodians may specify rules that provide a default value to avoid a missing field when data collectors do not specify a value, thus generating possibly inaccurate data.

Data consumer's knowing-how is highly associated with high data quality. One example dimension is accessibility. Data consumers with higher knowing-how assess

data accessibility higher. One financial analyst experiences difficulties in accessibility and states:

I need to know more about how to access these data. Sometimes, the data are there, but I don't know how to access them. It takes a long time to figure out the right "joins" for the new tables I need to create from the existing tables. [Comment from a financial analyst in a hospital]

The findings from Hypothesis 3 offer insights into different performance impacts associated with the three data roles. Data collectors collect relevant, accurate, and complete data. Data custodians are involved in storing complete data and in maintaining them to be timely and accurate. Data consumers are involved in identifying relevant data.

These findings present an interesting observation: there is no overlap between the performance dimensions correlated with custodians' knowledge and those correlated with consumers' knowledge. One conjecture from these findings is that a key function of data collectors is to understand the needs of data consumers, the relevancy dimension, and to collect accurate and complete data for storage by custodians. Thus data collectors may serve the role of data quality brokers or intermediaries between custodians and consumers.

Another interesting observation is the overall weak contribution of the knowledge held by data custodians. This contradicts the conventional responsibility of data custodians for designing and implementing organizational information systems. For four out of five data quality performance dimensions, the knowledge held by data collectors, rather than data custodians, has the most highly significant associations. This contradicts the generally weak responsibility of clerical data collectors in today's organizations. In a typical IS requirements analysis process, data custodians interview data consumers about data requirements. Our results suggest that data consumer's knowledge about data needs other than relevancy of data to their tasks will have limited effects on data quality. Data collectors seem to hold key data quality knowledge, but, in today's IS requirements process, they rarely play a significant role.

### Data Roles: Added Division of Labor

The results from the factor analysis confirm the three data production subprocesses as distinct factors, that is, distinct domains of knowledge. The further correlation analysis also confirms the differential impacts of data roles on data quality problem solving.

Regardless of organizational members' conventional functional task roles, such as marketing, finance, or manufacturing, the three data roles are distinctively differentiated along their knowledge about their primary production process: collection, storage, and utilization. For example, a data consumer's work on data for any conventional functional domain, that is, marketing, finance, or strategic planning, involves accessing data and using them for business purposes by reformatting, aggregating, inter-

preting, and representing data. To fully understand how best to perform organizational work and to design organizational processes effectively, our new finding of data roles based on the knowledge about data production process needs to be incorporated into the conventional division of labor, which is based on hierarchical and functional differentiation.

We are not stating that data roles ought to replace the current divisions based on functions and hierarchies. We argue that the distinct data roles based on the knowledge about corresponding data production process need to be included along with the conventional understanding of division of labor, as data increasingly play a larger role in organizational work. Particularly for emerging flat organizations where conventional hierarchy and functional divisions are collapsed into larger blocks of processes, our finding of data roles can be used to understand, reassign, and facilitate data-based collaboration within and across the integrated work process.

## Conclusion

---

A MAJOR FINDING OF THIS STUDY is the complexity of knowledge at work in a specific problem solving situation, namely, a data production process. The evidence of the three modes of knowledge becomes clearer when that knowledge is being applied to the data production process by a particular data role. We find that the critical mode of knowledge that impacts organizational performance differs depending on the roles of organizational members and on the domain of the problem solving process.

Some might argue that our findings of relationships among knowledge at work for data quality can be explained by either lack of communication or exhibition of defensive routines [3]. For example, the typical actions taken by data custodians can arguably be interpreted as “defensive routines.” The notion of defensive routine, however, does not fully explain the knowledge held and used by the data custodians. It assumes that data custodians have full knowledge of how and why data consumers use the data. It also does not take into account the complex constraints built in the database systems for technical and operational purposes. With the understanding of modes of knowledge and their impacts on organizational performance, communication and coordination issues can be handled in an informed way.

## Implications for Research

The findings have important implications for both researchers and practitioners. For researchers, the findings can be used to extend the theories in several areas. *First*, the findings can be used to extend theory on how knowledge impacts organizational performance in different work settings. For example, some work settings may require organizational members to hold and use why knowledge to improve organizational performance. On the other hand, some organizational processes need to be designed to embed knowing-why into activities and procedures. *Second*, the findings can be used to explore the transformation process of knowledge modes in work. For ex-

ample, knowing-how embedded in a problem solving process can be structured into knowing-what. This transformation explains how IS-related work changes problem solving patterns. *Third*, the findings can be used to extend and revise the theories of division of labor in work by incorporating how data roles in the organizational data production processes interact with conventional or emerging work processes. The results will reflect today's work situation more realistically. *Finally*, the research process presented here can be applied to advance theories at both general and domain-specific levels as the process used here exploits both the clarity and the contextual articulations from including theories at general and domain-specific levels.

### Implications for Practice

For practitioners, the implications are directly applicable for managing organizational data and for managing work processes for general organizational performance. To process organizational data, a firm's data production process is conceptually divided into three distinct areas: data collection, data storage, and data utilization. Members in each process, regardless of one's functional specialty, focus on collecting, storing, or utilizing data. To achieve high data quality, all three processes must work properly.

Most organizations handle data quality problems by establishing routine control procedures in organizational databases. To solve data quality problems effectively, the members in all three processes must hold and use sufficient knowledge about solving data quality problems appropriate for their process domains. At minimum, data collectors must know what, how, and why to collect the data; data custodians must know what, how, and why to store the data; and data consumers must know what, how, and why to use the data.

In addition, knowledge about work processes other than their immediate work processes contributes to producing high-quality data. This is in line with the conventional notion that cross-functional knowledge contributes to organizational performance. Our results show that this applies to some roles, but not to others. For example, data custodians' knowledge about other work processes does not contribute to producing high-quality data, but data collector's does.

In sum, the three modes of knowledge held by different data roles in a data production process work together to identify and solve problems in the process, and thus to improve organizational data quality. Understanding the differentiated relationships among work processes, knowledge modes, and organizational performance opens up new avenues for future research and practice that can study and advise knowledge-intensive, global, and loosely structured virtual work environments.

### 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. Argote, L.; Beckman, S.L.; and Epple, D. The persistence and transfer of learning in industrial settings. *Management Science*, 36, 2 (1990), 140–154.

3. Argyris, C. *Knowledge for Action: A Guide to Overcoming Barriers to Organizational Change*. San Francisco: Jossey-Bass, 1993.
4. Argyris, C., and Schön, D.A. *Organizational Learning II*. Boston: Addison-Wesley, 1996.
5. Ballou, D.P., and Tayi, G.K. Enhancing data quality in data warehouse environment. *Communications of the ACM*, 42, 1 (1999), 73–78.
6. Ballou, D.P.; Wang, R.Y.; Pazer, H.; and Tayi, G.K. Modeling information manufacturing systems to determine information product quality. *Management Science*, 44, 4 (1998), 462–484.
7. Becerra-Fernandez, I., and Sabherwal, R. Organizational knowledge management: A contingency perspective. *Journal of Management Information Systems*, 18, 1 (Summer 2001), 23–56.
8. Berger, P., and Luckman, T. *The Social Construction of Reality*. New York: Anchor Books, 1990.
9. Brown, J.S., and Burton, R.R. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2, 2 (1978), 155–192.
10. Brown, J.S., and Duguid, P. Organizational learning and communities of practice: Toward a unified view of working, learning and innovation. *Organization Science*, 2, 2 (1991), 40–57.
11. Clark, K.B., and Fujimoto, T. *Product Development Performance*. Boston: Harvard Business School Press, 1995.
12. Codd, E.F. A relational model of data for large shared data banks. *Communications of the ACM*, 13, 6 (1970), 377–387.
13. Cohen, M.D., and Sproull, L.S. (eds.). Special issue on organizational learning: Papers in honor of (and by) James G. March. *Organization Science*, 2, 1, 1991.
14. Cohen, W.M., and Levinthal, D.A. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 1 (1990), 128–152.
15. Dewey, J. *How We Think*. Boston: D.C. Heath, 1910.
16. Duncan, R., and Weiss, A. Organizational learning: Implications for organizational design. In B.M. Staw (ed.), *Research in Organizational Behavior*, vol. 1. Greenwich, CT: JAI Press, 1979, pp. 75–123.
17. Epple, D.; Argote, L.; and Devadas, R. Organizational learning curves. *Organization Science*, 2, 1 (1991), 58–70.
18. Funk, J.; Lee, Y.; and Wang, R. Institutionalizing information quality practice. In I. Chengalur-Smith and L. Pipino (eds.), *International Conference on Information Quality*. Cambridge, MA: ICIQ, 1998, pp. 1–17.
19. Hiebert, J. Children's mathematics learning: The struggle to form link and understanding. *Elementary School Journal*, 84, 5 (1984), 497–513.
20. Hu, J.; Huang, K.T.; Kuse, K.; Su, G.; and Wang, K. Customer information quality and knowledge management: A case study using knowledge cockpit. *Journal of Knowledge Management*, 1, 3 (1998), 225–236.
21. Huang, K.; Lee, Y.; and Wang, R. *Quality Information and Knowledge*. Upper Saddle River, NJ: Prentice Hall, 1999.
22. Jarvenpaa, S.L., and Staples, D.S. Exploring perceptions of organizational ownership of information and expertise. *Journal of Management Information Systems*, 18, 1 (Summer 2001), 151–183.
23. Kahn, B.K., and Strong, D.M., Product and service performance model for information quality: An update. In I. Chengalur-Smith and L. Pipino (eds.), *International Conference on Information Quality*. Cambridge, MA: ICIQ, 1998, pp. 102–115.
24. Knorr-Cetina, K. *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science*. Oxford: Pergamon, 1981.
25. Knorr-Cetina, K. The micro-social order: Towards a reconception. In N.G. Fielding (ed.), *Actions and Structure: Research Methods and Social Theory*. Newbury Park, CA: Sage, 1988, pp. 21–53.
26. Kogut, B., and Zandler, U. What firms do? Coordination, identity, and learning. *Organizational Science*, 7, 5 (1996), 502–518.
27. Lave, J., and Wenger, E. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press, 1991.

28. Lee, Y. Why “know-why” knowledge is useful for solving information quality problems. In *Americas Conference on Information Systems*. Phoenix, AZ: ACIS, 1996, pp. 200–202.
29. Lee, Y.; Strong, D.; Kahn, B.; and Wang, R. AIMQ: A methodology for information quality assessment. *Information & Management*, 40, 2 (2002), 133–146.
30. Lyles, M.A., and Schwenk, C.R. Top management, strategy and organizational knowledge structures. *Journal of Management Studies*, 29, 2 (1992), 155–174.
31. Nelson, R.R., and Winter, S.G. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Harvard University Press, 1982.
32. Nonaka, I., and Takeuchi, H. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press, 1995.
33. Nosofsky, R.M. Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 1 (1984), 104–114.
34. Pipino, L.; Lee, Y.; and Wang, R. Data quality assessment. *Communications of ACM*, 45, 4 (April 2002), 211–218.
35. Polanyi, M. *The Tacit Dimension*. Garden City, NY: Doubleday, 1966.
36. Raelin, J.A. Action learning and action science: Are they different? *Organizational Dynamics*, 26, 1 (Summer 1997), 21–34.
37. Rob, P., and Coronel, C. *Database Systems: Design, Implementation, and Management*. Boston: Course Technology, 1997.
38. Sackmann, S. *Cultural Knowledge in Organizations: Exploring the Collective Mind*. Newbury Park, CA: Sage, 1997.
39. Schön, D.A. Generative metaphor: A perspective on problem-setting in social policy. In A. Ortony (ed.), *Metaphor and Thought*, 2d ed. New York: Cambridge University Press, 1993, pp. 1–16.
40. Shelling, T. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press, 1960.
41. Stenmark, D. Leveraging tacit organizational knowledge. *Journal of Management Information Systems*, 17, 3 (Winter 2001–2), 9–24.
42. Storey, V.C., and Wang, R.Y. Modeling quality requirements in conceptual database design. In I. Chengalur-Smith and L. Pipino (eds.), *Conference on Information Quality*. Cambridge, MA: ICIQ, 1998, pp. 64–87.
43. Strong, D.; Lee, Y.; and Wang, R. Data quality in context. *Communications of the ACM*, 40, 5 (1997), 103–110.
44. Strong, D.; Lee, Y.; and Wang, R. Ten potholes in the road to information quality. *IEEE Computer*, 30, 8 (August 1997), 38–46.
45. Wang, R.Y. *Models of Information Quality*. Norwell, MA: Kluwer Academic, 1998.
46. Wang, R.Y. A product perspective on total data quality management. *Communications of the ACM*, 41, 2 (1998), 58–65.
47. Wang, R.Y., and Strong, D.M. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12, 4 (Spring 1996), 5–34.
48. Wang, R.Y.; Lee, Y.; Pipino, L.; and Strong, D. Manage your information as a product. *Sloan Management Review*, 39, 4 (1998), 95–105.
49. Weiss, C. Knowledge creep and decision accretion. *Knowledge*, 1, 3 (1980), 381–404.
50. Wisniewski, E.J., and Medin, D.L. On the interaction of theory and data in concept learning. *Cognitive Science*, 18, 2 (1994), 221–281.

## Appendix. The Measures

---

### Knowledge (Independent) Measures

ALL ITEMS ARE MEASURED ON A 1 TO 10 SCALE, where 1 is “very small extent,” 5 is “average,” and 10 is “very large extent.” Items labels with (R) are reverse coded.

Knowing what about data collection (five items, Cronbach’s alpha = 0.92)

- (KWTC01) I know who creates this information.
- (KWTC02) I know which group collects this information.
- (KWTC03) I know the procedures by which this information is collected.
- (KWTC04) I know the steps taken to gather this information.
- (KWTC06) I know the sources of this information.

Knowing what about data storage (six items, Cronbach’s alpha = 0.94)

- (KWTS01) I know who maintains this information in our computers.
- (KWTS02) I know which group maintains this information in our computers.
- (KWTS03) I know the procedures used to store this information in our computers.
- (KWTS04) I know the steps taken to store and maintain this information in our computers.
- (KWTS05) I know which of our computers stores this information.
- (KWTS06) I know which software is used for storing this information in our computers.

Knowing what about data utilization (five items, Cronbach’s alpha = 0.93)

- (KWTU01) I know who (individual or group) uses this information.
- (KWTU02) I know which group uses this information.
- (KWTU03) I know the procedures in which this information is used.
- (KWTU04) I know the steps taken when using this information.
- (KWTU05) I know the tasks which require the use of this information.

Knowing how-to about data collection (5 items, Cronbach’s alpha = 0.94)

- (KHWC01) When typical problems arise with collecting this information, I know how we handle them.
- (KHWC02) I know the usual solutions for problems with collecting this information.
- (KHWC03) I know how to fix routine problems with collecting this information.
- (KHWC04) I know how to fix recurring problems with collecting this information.
- (KHWC05) I know the standard procedures for correcting deficiencies in information when collecting it.



Knowing how-to about data storage (five items, Cronbach's alpha = 0.96)

(KHWS01) When typical problems arise with storing this information in our computers, I know how we handle them.

(KHWS02) I know the usual solutions for problems with storing this information in our computers.

(KHWS03) I know how to fix routine problems with storing this information in our computers.

(KHWS04) I know how to fix recurring problems with storing this information in our computers.

(KHWS05) I know our standard procedures for correcting deficiencies in information when storing it in our computers.

Knowing how-to about data utilization (four items, Cronbach's alpha = 0.94)

(KHWU01) When typical problems, such as interpretation or access, arise with using this information, I know how we handle them.

(KHWU02) I know the usual solutions for problems with using this information.

(KHWU03) I know how to fix routine problems with using this information.

(KHWU05) I know our standard procedures for correcting deficiencies in information when using it.

Knowing why about data collection (four items, Cronbach's alpha = 0.91)

(KWYC02) I know the problems encountered in collecting this information.

(KWYC04) I understand the information collection procedures well enough to recognize why this information is collected incorrectly.

(KWYC05) I can detect sources of new problems in collecting this information.

(KWYC06) I can recognize new problems as they arise in collecting this information.

Knowing why about data storage (seven items, Cronbach's alpha = 0.93)

(KWYS02) I know why this information is displayed in this form in our computers.

(KWYS03) I know some of the problems in storing this information appropriately in our computers.

(KWYS04) I know why it is difficult to store this information in our computers in an easy-to-interpret manner.

(KWYS06) I understand our computing environment well enough to analyze why this information is stored inadequately.

(KWYS08) I can recognize new problems as they arise in storing and maintaining this information in our computers.

(KWYS10) I know why people have difficulty with computer access procedures for this information.

(KWYS11) I know why it is difficult to store all this information in our computers.

Knowing why about data utilization (six items, Cronbach's alpha = 0.88)

(KWYU03) I know some of the problems in ensuring that this information is used appropriately.

(KWYU06) I can detect sources of new problems in using this information.

(KWYU07) I can recognize new problems as they arise in using this information in a new task.

(KWYU15) I cannot diagnose problems in using this information. (R)

(KWYU16) I cannot find the causes of new problems in the use of this information. (R)

(KWYU17) I cannot recognize when new problems arise in using this information in a new task. (R)

### Data Quality (Dependent) Measures

All items are measured on a 0 to 10 scale, where 0 is "not at all," 5 is "average," and 10 is "completely." Items labels with (R) are reverse coded.

Accuracy (four items, Cronbach's alpha = 0.91)

This information is correct.

This information is incorrect. (R)

This information is accurate.

This information is reliable.

Completeness (six items, Cronbach's alpha = 0.87)

This information includes all necessary values.

This information is incomplete. (R)

This information is complete.

This information is sufficiently complete for our needs.

This information covers the needs of our tasks.

This information has sufficient breadth and depth for our tasks.

Timeliness (five items, Cronbach's alpha = 0.88)

This information is sufficiently current for our work.

This information is not sufficiently timely. (R)

This information is not sufficiently current for our work. (R)

This information is sufficiently timely.

This information is sufficiently up-to-date for our work.

Relevancy (four items, Cronbach's alpha = 0.94)

This information is useful to our work.

This information is relevant to our work.

This information is appropriate for our work.

This information is applicable to our work.

Accessibility (four items, Cronbach's alpha = 0.92)

This information is easily retrievable.

This information is easily accessible.

This information is easily obtainable.

This information is quickly accessible when needed.

Copyright of Journal of Management Information Systems is the property of M.E. Sharpe Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.