IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com
ITB12692

This paper appears in the book, *Emerging Trends and Challenges in Information Technology Management, Volume 1 and Volume 2*
edited by Mehdi Khosrow-Pour © 2006, Idea Group Inc.

A Comparison of Quality Issues for Data, Information, and Knowledge

Elizabeth Pierce, Dept of MIS & Decision Sciences, The Eberly College of Business and Information Technology, Indiana University of Pennsylvania, 664 Pratt Dr, Rm 203, Indiana, PA 15705-1087, P (724) 357-5773, F (724) 357-4831, empierce@iup.edu

Beverly Kahn, Information Systems & Operations Management Department, Sawyer School of Management, Suffolk University, 8 Ashburton Place, Boston, MA 02108, P (617) 573-8642, F (617) 994-4228, bkahn@suffolk.edu

Helinä Melkas, Helsinki University of Technology, Lahti Centre, Saimaankatu 11, FI-15140 Lahti, Finland
P +358 3 525 0285, F +358 3 525 0204, Helina.melkas@tkk.fi}

ABSTRACT
Most agree that although related, data, information and knowledge differ from each other. Given this distinction, it is logical to ask if there is also variation in how quality is defined, measured, and improved for these three concepts. In this paper, the definitions for data, information, and knowledge are compared and their quality characteristics are explored.

INTRODUCTION
Ask a manager whether he would like an additional data set, another report, or more knowledge about his business; it is safe to say that most managers would choose more knowledge. Even if the data set or report were of excellent quality and the knowledge less so, one might imagine that most managers would still choose knowledge. There is an implied hierarchy between data, information, and knowledge with knowledge being perceived by many as the most desirable. If indeed knowledge is the ultimate product produced by an organization's systems then it is important to understand the relationship between these three concepts and how the quality of one affects the others. To begin, consider the differences in how people describe data, information, and knowledge.

BACKGROUND
Most scholars ([1], [2], [3], [4], [7], [8], [12], [13]) refer to a datum as the most basic descriptive element. Whether it is symbolized as a number, text, or figure, a datum essentially represents a perception or measurement about some object of interest. By itself, a datum's value typically lacks content, meaning, or intent. Data is the plural of datum and its usage is more common because for the most part, organizations work with collections of datum. For example, consider the kinds of datum that are used to describe a customer sales order. Individual datum like the customer's name, the item's description, the quantity sold, and the price are grouped together to form data. Data are often organized as a record, i.e., a set of attributes whose values describe some entity or event. Each attribute's value can be considered a datum that describes some observation to be retained about the sale to that customer.

Although some use the term data interchangeably with information, others consider information to be more than just data. They view information as the output of some process that interprets and manipulates data into some prescribed format ([1], [2], [3], [4], [7], [8], [12], [13]). Some authors prefer to use the phrase, information product, which is identified in terms of the raw, source data and semi-processed component data items required to manufacture it [17]. The expression, information product, emphasizes the idea that this item is determined by more than just its input data, but also by the procedures used to construct it. Examples of information products include sales orders, packing lists, shipping labels, and customer invoices.

Unfortunately not all information products are characterized by as much stability or simplicity in form or content as a shipping label. Some information products are more ad-hoc in nature. For instance, the results of queries are typically based on an assortment of data, presented in a variety of formats, generated on demand, and are typically used by only a few consumers. This type of information product is similar in nature to a one-of-a-kind manufactured product. Other information products are characterized by their complexity. Consider information products like data warehouses, hypertext documents, catalogs, and reference materials which may contain text, images, and audio objects. Such complex information products are often custom-made by a few people and then disseminated to a large audience. These complicated information products are particularly vulnerable to quality issues regarding their reliability, organization, content, accessibility, and presentation.

Finally, while some view knowledge as information that has been further enriched so its value, context, and meaning are enhanced; others consider knowledge as being intrinsically different from either data or information products ([1], [2], [3], [4], [7], [8], [12], [13]). The idea that knowledge is more than information stems from the notion that knowledge is more process than product. The knowledge process occurs when an individual mentally synthesizes together an assortment of inputs: information, experiences, beliefs, relationships, and techniques to determine what a specific situation means and how to handle it [1]. For example, if the Marketing Vice President wants to devise next year's sales strategy, he cannot solely rely on viewing information products like the end of year sales report or consumer market survey to acquire this understanding. Using his own internal reasoning, he must combine his assessment of these information products with his other accumulated experiences to come up with a plan for how to act.

To make better use of the knowledge that will benefit their employees, processes, products, and performance, many companies are seeking to improve their Knowledge Management Systems (KMS). A KMS is not a computer system; rather it is a way of doing business that promotes knowledge management. Firms interested in improving their ability to discover, capture, share and apply knowledge must implement a variety of organizational means and technologies. Organizational means include such practices as collaborative creation of documents, face-to-face meetings, on-the-job training, rotation of employees across departments, and corporate retreats. Technologies that enhance knowledge management include database management systems, video-conferencing, e-mail, groupware, and web portals. To be successful in knowledge discovery, capturing, sharing, and application, both the organizational means and the technologies employed must be compatible with the underlying organization's culture, structure, information technology (IT) infrastructure, common knowledge base, and physical environment.

In particular, it is through the IT infrastructure which includes data processing, storage, communication technologies, and computer systems that data and information products are linked to the people and their actions for creating, storing, distributing, and exploiting knowledge of all types within an organization. Failure to understand this connection between the IT architecture's information systems and a company's ability to manage knowledge can lead to practical difficulties in organizations such as poor information sharing between different functional areas [14]. To avoid these problems, companies should identify as part of their system's requirements, the kinds of knowledge needed to conduct day-to-day operations and to make decisions. Because some aspect of this knowledge must be conveyed and stored in a physical format, this necessitates the design and development of information products. Only when the design criteria for information products are well understood can an organization proceed to make sound decision about how to model, represent, and process the raw data upon which the information products will be based. Once the IT infrastructure's systems are constructed and operational, data from the transactional processing systems are funneled into the management information systems which in turn help to support the organization's KMS. Thus, while the design of systems and processes seem to be knowledge-driven, the operations of the systems appear to be data-driven. Given this interrelationship, the main thrust of the remainder of this paper will be to compare how quality dimensions, measurements, and methods for improvement compare between data, information products, and knowledge.

COMPARING DATA, INFORMATION, AND KNOWLEDGE QUALITY

Quality Dimensions

Most scholars agree that data quality is multidimensional in nature (Table 1: [10], [15], [16]). Although not always explicitly stated in the literature, it seems reasonable that one can apply these dimensions to data, information products or knowledge and their general meanings continue to apply.

Quality Measurement

Although the meanings of the quality dimensions are similar for data, information products, and knowledge, they differ in their measurability. Consider the quality dimension of completeness. Completeness refers to the extent to which something is not missing and is of sufficient breadth and depth for the task at hand. Completeness can be measured using subjective perceptions supplied by the consumer or by using quantifiable, objective measures which may either be task dependent or independent. Table 2 illustrates that it is relatively easy to derive objective, task independent completeness measures for data. Even if domain experts are needed to help ascertain whether the data's schema or population is sufficiently complete to satisfy the requirements of a particular application, it is still fairly straight forward to derive objective, task dependent completeness measures. This occurs because data records are structured and well defined. The same data record is also probably used by many different functional areas (sales, shipping, billing, etc.) so that certain quality characteristics of the data record such as its completeness, consistency, accuracy, and currency can be measured consistently across applications that require it.

Table 1. Summary of quality dimensions for data, information products, and knowledge

Quality Category	Dimensions Associated with this Category
Quality of Values Collected (Also known as Intrinsic Quality)	Accuracy, Objectivity, Believability, Source Reputation, Completeness, Unambiguous, Meaningfulness, Currency
Quality of Application (Also known as Contextual Quality)	Value-added, Relevancy, Timeliness, Comprehensiveness, Appropriate Amount, Appropriate Use, Proficiency in Use
Quality of Presentation and Storage (Also known as Representational Quality – Deals with Format and Definition)	Ease of Interpretation, Ease of Understanding, Representational Consistency, Concise Representation, Appropriate Precision/Granularity, Good Organization/Structure/Definition
Quality of Accessibility via System	Availability, Diffusion, Ease and Speed of Retrieval, Ease of Manipulation, Security, Privacy
Quality of System Support Services	Feedback, Measurement, Improvement Track Record, Help Services, Ability to Handle Special Requests, Architecture, Portability, Commitment to Quality Policy

Table 2. Completeness measures for data, information product, and knowledge

Examples of Quality Metrics for Completeness Dimension			
Metric Types	**Data Record**	**Information Product**	**Knowledge**
Objective, Task Independent Measure	*- Single Record -* Within a record, how many attributes contain values. *- Group of Records -* Proportion of records that contain all their values. Proportion of records that contains a value for a given attribute. Of the total cells contained in a table, the proportion that contains values.	*-Single IP –* Within an IP, the proportion of values which are present and accounted for. *- Group of IP's -* Proportion of IP's that contain all their values.	*-knowledge area –* Codification: How complete is the definition and structure of the knowledge be it explicit, tacit, or some other type such as self-transcending.
Objective, Task Dependent Measure	*- Single Record -* Number of attributes missing from the record. *- Group of Records -* Number of records that are missing from the data set. Number of attributes whose domain of values are incomplete.	*-Single IP –* Within an IP, are there any values that are partially incomplete? (e.g. missing items from a list or an incomplete sum) An IP's design lacks certain pieces of information required by the task. *- Group of IP's -* Number of IP's missing from a group.	*- knowledge area -* Proficiency: How complete is the depth and comprehension of knowledge. Diffusion: How complete is the distribution and networking of knowledge capabilities across relevant stakeholders Value: How complete is the impact associated with the knowledge's contribution to employees, processes, products, and performance.
Subjective Measure	Is this data record sufficiently complete for the task at hand?	Is this information product sufficiently complete for your needs?	Do you have the knowledge you need to analyze this situation?

Objective quality measurements can also be developed for information products. Their quality metrics tend to be more context specific than those of data. This is reasonable since many information products are developed specifically for use by a particular business functional area. Thus one might examine a group of information products like a stack of sales orders for the purpose of determining if any sales orders were omitted or if the sales order design itself addresses all the information needs of the marketing group. Information products also tend to have more processing associated with their creation than data. This occurs because an information product builds upon the prior processing used to collect and store the raw data by adding additional steps that convert the raw data into a specified form. Hence, one might also wish to develop quality metrics that examine how complete were the data inputs and activities employed during each stage of an information product's construction.

Because knowledge is essentially the result of a process by which a variety of inputs are combined together by someone who wishes to determine what a specific situation means and how to handle it, it can be a challenge to find objective, application independent quality metrics for knowledge. One possible measure in this area is to identify the degree to which the knowledge can be made explicit or codified. This could be recorded on a scale ranging from gut feelings and mind models on the low end, then onto discussions, presentations, reports, followed by best practices and standards on the high end [5].

For certain types of knowledge, it is possible to obtain objective, task dependent quality measures, especially in the case of explicit know-how or know-why knowledge. For example, completeness for explicit know-how knowledge can be defined as whether all the steps in the process are described. In addition for an individual step, one can ask if all the necessary details were included. Note that the completeness of the "step knowledge" is dependent not only on the nature of the task but on the level of the user's expertise. A novice may require more explicit knowledge than a journeyman who needs only a subset of this knowledge. Explicit knowledge plus the individual's tacit knowledge form the complete set of knowledge proficiencies used to accomplish the step. Completeness of knowledge can also be evaluated from the perspective of its diffusion across the enterprise or in its impact on the work accomplished by the firm.

These last points demonstrate that of the three concepts: data quality, information product quality, and knowledge quality that it is knowledge quality that is most defined by both its context and by the individuals who must apply the knowledge within that context. This sentiment is echoed by the Knowledge Management Professional Society who state that the criteria used to evaluate knowledge should be based upon the standards and conditions unique to the individual or group seeking to validate the knowledge [6]. Thus knowledge quality needs to be measured and investigated from a more personal and task oriented point of view than either data quality or information product quality. As a consequence, subjective measures that inquire as to which parts of the knowledge process were sufficiently present may be the principle means by which to judge the different quality aspects of one's knowledge for a given purpose.

Quality Improvement

Most researchers agree that improvement of data, information products, and knowledge depends on applying quality principles and practices to the processes that create, store, manage, and present them ([4], [9], [17], [18]). The manufacturing literature provides many examples of using a total quality management philosophy to raise the satisfaction levels for many different types of products and services. Within the data quality field, it has been demonstrated that these quality improvement tools and techniques can be successfully adapted to the special characteristics of data, information products, and knowledge which possess atypical quality dimensions like "believability" and exhibit a simultaneous, non-depleting, multi-use nature [18].

In terms of how the quality improvement effort differs between data, information products, and knowledge, several observations can be noted. Although the creation and management of data, information products, and knowledge all involve processes, the complexity of those processes increases as one moves from the production of data through the production of information products to the production of knowledge. As the complexity of the processes increases, so does the difficulty associated with the definition, measurement, analysis, and improvement of quality. In part, the growing complexity of the processes stems from the need for a greater number of "manufacturing" steps over a longer length of time. Another complicating factor is that data, information products, and knowledge follow a life cycle which can be characterized by four major stages: Creation, Growth, Maturity, and Decline [17]. In practice, the distinctions between these four stages may not be clear cut. For example in healthcare, an information product such as a patient's medical file tends to change continuously during its different phases of utilization and it is not necessarily complete even at the very end of the process. This raises the concern that the quality measured at only one stage of the process may give an incomplete picture that does not guide improvement efforts sufficiently.

SUMMARY AND RECOMMENDATIONS FOR FUTURE RESEARCH

Whether it is a routine task such as recording a sales order or a complex activity such as devising next year's marketing campaign, organizations must ensure their employees have the knowledge they need to complete tasks and make appropriate decisions. Making sure this knowledge is adequately developed, captured, shared, and used is the goal of the KMS which among its many facets includes the IT infrastructure where one finds the systems used to collect, manipulate, and store data records and information products. Improving the quality of knowledge requires a holistic approach to the entire knowledge management process which includes an understanding of the role that quality improvements in data records and information products can play. To better this understanding, more research is needed to address the following questions.

- Should organizations concentrate on measuring quality separately within the various systems that manage data, information products, and knowledge or should organizations concentrate on obtaining quality measurements at the boundary points where these systems interact?
- How best to define and capture quality measures for data, information products, and knowledge?
- How do quality assurance costs for the data, information products, and knowledge compare?
- How do legal issues between data quality, information product quality, and knowledge quality compare?
- How do quality policy and personnel issues compare between data, information product, and knowledge quality?
- Sarbanes-Oxley and Health Insurance Portability and Accountability Act have legal requirements for the privacy and security of data. This has also been applied to some information products. Can these be extended to knowledge as well?

While it is apparent that this paper asks more questions that it gives answers, it nonetheless serves the purpose of highlighting further research that must be done if organizations are to fully integrate their data, information products, and knowledge systems. What is needed now is a more comprehensive literature review, comparison, and framework of quality issues related to data, information products, and knowledge.

REFERENCES

[1] Becerra-Fernandez, I.; Gonzalez A.; and Sabherwal, R. *Knowledge Management: Challenges, Solutions, and Technologies*. Prentice Hall, Upper Saddle River, NJ, 2004.

[2] Davenport, T. *Information Ecology: Mastering the Information and Knowledge Environment*. Oxford University Press, New York, NY, 1997.

[3] Earl, J.J. "Knowledge as Strategy: Reflections on Skandia International and Shorko Films." In C. Ciborra and T. Jelassi (Eds.), *Strategic Information Systems: A European Perspective*,. John Wiley & Sons, Chichester, UK, 1994, pp. 53-69.

[4] English, L. P. *Improving Data Warehouse and Business Information Quality*. John Wiley & Sons, New York, NY, 1999.

[5] Hofer-Alfeis, J. http://www.i-know.at/previous/i-know04/presentations/Hofer-Alfeis_authorized. pdf. Published in 2004, Accessed on June 28, 2005.

[6] Knowledge Management Profession al Society. http://kmpro.org,, Accessed on June 28, 2005.

[7] Lillrank, P. "The Quality of Information." *International Journal of Quality & Reliability Management*, 20 (6), 2003, pp.691-703.

[8] Miller, B.; Malloy, M. A.; Masek, E.; and Wild, C. "Towards a framework for managing the information environment." *Information and Knowledge Systems Management*, 2, 2001, 359–384.

[9] Redman, T. *Data Quality for the Information Age*. Artech House, Boston, MA, 1996.

[10] Redman, T. *Data Quality: The Field Guide*. Butterworth-Heinemann, Boston, MA, 2001. p. 74.

[11] Scharmer, C. O. "Self-transcending knowledge: Organizing around emerging realities." in Nonaka, I., Teece, D. (eds.), *Managing industrial knowledge: Creation, transfer and utilization*, Sage Publications, London, 2001, pp. 68–90.

[12] Spek, R.v.d., and Spijkervet, A. *Knowledge Management: Dealing Intelligently with Knowledge*. Ultrecht: Kenniscentrum CIBIT, 1997.

[13] Tuomi, I., "Data is More than Knowledge." *Journal of Management Information Systems*, 16 (3) , Winter 1999-2000, pp. 103-117.

[14] Uotila, T.; Melkas, H. "Book review – Ikujiro Nonaka and David Teece (eds.): Managing industrial knowledge. Creation, transfer and utilization." *Knowledge and Process Management*, 10(4), 2003, pp. 277-281.

[15] Wand, Y., and Wang, R. Y. "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM*, 39 (11), 1996, pp. 86-95.

[16] Wang, R. Y. and Strong, D. M. "Beyond Accuracy: What Data Quality Means to Data Consumers." *Journal of Management Information Systems*, 12 (4), 1996, pp. 5-34.

[17] Wang, R. Y., Lee, Y., Pipino, L. and Strong, D. "Manage Your Information as a Product." *Sloan Management Review*, 39(4), Summer 1998, pp. 95-105.

[18] Wang, R. Y. *"A Product Perspective on Total Data Quality Management,"* Communications of the ACM, 41 (2), 1998, pp. 58-65.