2. Information Quality Problems and Current Approaches

Chapter Overview

This chapter consists of four parts. In a first step, the information quality perspective will be introduced as a problem lens. Then, a survey of information quality problems from relevant literature and field research will be presented. This will result in a problem typology. In a third step, existing information quality frameworks will be presented and evaluated. In the last step of this chapter, research deficits and consequences of further framework development will be discussed.

2.1 Background and Key Terms

I do not say that definitions may not have a role to play in connection with certain problems, but I do say it is for most problems quite irrelevant whether a term can be defined or cannot be defined, or how it is defined. All that is necessary is that we make ourselves understood.[18]

SIR KARL R. POPPER

Goal of the Section

In this introductory section, key terms and concepts that will resurface in the problem description and in the analysis of information quality frameworks will be described and defined. Specifically, the relevant background and vocabulary of quality management and knowledge management (and their relationship) will be briefly discussed.

---

[18] Popper, 1994, p. 18. While I do not fully agree with this opinion and view definitions as one means of clarification, I do agree with Popper that explaining things by examples can lead to a better understanding than simply providing definitions (Popper, 1994, p.19).

Quality Management & Knowledge

If one examines the quality of information in knowledge-intensive processes, one inevitably has to rely on the concepts and terms of two 'disciplines', namely quality management and knowledge management. Although these two fields share many common goals, such as the documentation of procedural knowledge and the continuous improvement through systematic learning, exchange between the two disciplines seems to be neither frequent nor intensive, nor particularly fruitful.[19] Nevertheless, codification of experiences or insights (an issue that regards both disciplines) can only be productive and useful if the stored content is of high quality and can be turned into actionable knowledge. Hence, quality management may profit from a closer examination of the characteristics of knowledge and knowledge management, in return, may benefit from an analysis of the experiences of quality management and its conceptions of quality. In consequence, we will look at the background, differences, and key terms of both disciplines. This should give us a better understanding of the problems inherent in managing information quality in knowledge-intensive processes.

Different Backgrounds and Origins

Having stated that quality and knowledge management may be two fields with great complementarities and mutual benefits, one has to acknowledge the different backgrounds from which these disciplines originated. This regards both the academic background and the first practical applications of the two concepts. Whereas the quality management movement began with a focus on quantitative analysis and monitoring in a manufacturing context (see Evans & Lindsay, 1999), knowledge management evolved (much later than quality management) out of the organizational learning and organizational memory literature and was first applied systematically in research or consulting contexts (see Lehner, 2000). Although both areas have come to stress the importance of learning and communication, they do so from two distinct perspectives. One could argue that quality management focuses on processes and their reliability and how they lead to customer satisfaction. Knowledge management, in contrast, focuses on innovation (e.g., knowledge creation see von Krogh et al., 2000), and intellectual asset reuse (see Davenport & Prusak, 1998).

Two Mind-sets

This uncommon ground may not always be visible in the respective literature, but it does have consequences for the methodologies and tools that are used by these disciplines.

---

[19] See for example Lim et al. 1999.

Particularly the epistemological paradigms employed by the two disciplines seem far apart. Whereas knowledge management literature often stresses a constructivist or autopoetic view of knowledge (emphasizing the collective, context-dependent sense-making of individuals and groups, see von Krogh et al., 1994), quality management tends to be rooted in a positivistic, Cartesian mind-set, that aims at the objective measurement of clearly given situations or problems.[20] These two different mind-sets may lead to, at times, incommensurable views on issues such as measurement, improvement, or learning. This is also the reason why information quality management for the context of knowledge-intensive processes is not just a replica or slight adaptation of concepts such as Total Quality Management or Quality Assurance. In other words: information quality management cannot be just an adaptation of traditional quality management concepts to the immaterial world. It has to be rooted in a knowledge-based view of the firm that differs in terms of the basic assumptions with many ideas inherent in quality management.

Nevertheless, many insightful ideas from quality management can be used in the knowledge work domain in spite of the great differences between the two disciplines. These ideas relate, for instance, to the mobilization of the workforce for quality issues, the forms of aggregation of data about quality, and the identification of non-quality costs.

Required Definitions

To turn these ideas into concepts and to use them in a new (often immaterial) context, we must first clarify some of the key terms of knowledge and quality management. These terms are *quality*, *(total) quality management, data*, *information*, *knowledge*, *knowledge work(er)*, and *knowledge-intensive process*. While there is a plethora of definitions available for these terms, the next few paragraphs will focus on definitions that represent a terminology which is both representative (of the academic field) and useful for the current context.

Quality as a Concept

In their review of quality management literature, Evans and Lindsay conclude the following about the term ‘quality’ and its definition:

---

[20] This distinction seems true for many typical “representatives” of the two disciplines. It does, of course, not apply to all scholars or practitioners active in the two fields. There are many knowledge management ‘activitists’ with positivistic mind-sets, and there are quality management advocates who stress the socially constructed nature of managerial challenges.

> Quality can be a confusing concept, partly because people view quality in relation to differing criteria based on their individual roles in the production-marketing chain. (Evans & Lindsay, 1999, p. 10).

Standard Definitions of Quality

In spite of this confusion, one can actually find a limited number of recurring definitions that adequately describe the term. Two of the most frequently encountered definitions are the following ones: "Quality is the totality of features and characteristics of a product or service that bears on its ability to satisfy given needs." "Quality is meeting or exceeding customer expectations" (Evans & Lindsay, 1999, pp. 15). Variations of these definitions often include the terms 'high value', 'error-free', 'specifications' (which are met), or 'fitness for use'. Further definitions adapted to the information quality context will be provided in section 2.3.[21]

Twofold Nature of Quality
⇒ Subjective and Objective Indicators

What we can learn from these definitions is that quality has a *subjective* (e.g., meeting expectations) and an *objective* component (e.g., meeting requirements), or in other words an *absolute* ('error free', 'meeting specifications') and a *relative* dimension ('fitness for use', 'satisfy needs'). Any approach to quality, including information quality, has to take this twofold nature of quality into account. The duality of quality can have important consequences for the way that quality is measured. It cannot only be calculated with the help of automatically generated key indicators, but must also be evaluated according to the (subjective) judgments and opinions of the customers.

Quality Management

⇒ Focus on the Present and the Future

This view of quality as both an objective and a subjective phenomenon influences the definition of quality management. Managing quality becomes a task of not only assuring that the processes and their outcomes are under control and within specified limits, but also an activity that strives for customer 'delight' and loyalty. It becomes a systematic activity that is not only focused on the present, but also on future improvements. This understanding becomes clear in the following text book definition of total quality management:

> Total Quality Management is a management concept (and associated tools) that involves the entire workforce in focusing on *customer satisfaction* and continuous *improvement*.[22]

---

[21] For alternative definitions and their implications see also: Reeves, C.A., and Bednar, D.A. (1994) Defining Quality: Alternatives and Implications, in: *Academy of Management Review*, 19, no. 3, pp. 319-445.

[22] See Evans & Lindsay, 1999, glossary.

Quality Management Cycle

Most quality management concepts that strive to reach these two goals, do so by employing a *management cycle*. A quality management cycle that has become well-known is the Deming-cycle[23], which consists of the phases *plan* (the quality activities and goals), *do* (what is necessary to improve quality), *check* (whether this has worked), and *act* (to correct still existing deficiencies). The management cycle is an essential element of any quality management concept, since it offers a sequence of steps and associated tools through which quality can be assured and improved.

Total Data Quality Management

Closely related to this understanding of managing quality is the cycle of total data quality management (TDQM), a field that is of course much younger than TQM, but obviously more closely related to the topic of this book. Huang et al. Define this cycle as follows :

> The definition component of the TDQM cycle identifies Information Quality (IQ) dimensions. The measurement component produces IQ metrics. The analysis component identifies root causes for IQ problems and calculates the impacts of poor quality information. Finally, the improvement component provides techniques for improving IQ. They are applied along IQ dimensions according to requirements specified by the consumer (Huang et al., 1999, p. 16).

Analyzing Causes, Calculating Costs

This cycle (which Huang et al. describe with the steps *define*, *measure*, *analyze*, and *improve*) highlights other important aspects of quality management, namely those of analyzing root causes of problems, and calculating the costs of non-quality information. Although Huang et al. refer to information quality dimensions, their cycle is clearly labeled as a data quality management cycle. Hence, they not dot distinguish information from data. The next section will show (by way of examples) that data and information quality problems are clearly distinct. Here, we will show that they are different by looking at the definitions of data, information, and knowledge. This will also highlight one of the goals of information quality management, namely to turn information into actionable knowledge.

Data ≠ Information Knowledge

Whereas data designates ‘raw,’ unconnected, quantitative or qualitative items, the term information relates to answers to questions, statements about situations or facts. Information in this sense is a difference that makes a difference (Bateson,

---

[23] See for example Deming 1986.

1972).[24] A piece of data is just a distinction or a registered item without context. Data in this sense becomes information when it is related to other data. In other words, when we link various sets of data to form one coherent statement, the resulting entity can be called a piece of *information:* a coherent set of statements that forms a message. In the context of this book we look at information as potential knowledge that has to be internalized by the receiver. We see information as a *production factor* and as the *input* and *output* of certain knowledge-intensive business processes. This information becomes knowledge when it is correctly interpreted and connected with prior knowledge.[25] Knowledge, in the traditional epistemological analysis of the term, only qualifies as such if it is a justified (e.g., it can be argued convincingly), true (i.e., corresponds to facts) belief (i.e., is held by an individual).[26] This 'platonic' notion of knowledge, however, has been extensively criticized.

One of the most prominent critics of this 'purist' approach to knowledge is MICHAEL POLANYI. He associates knowledge closely with an individual's *personal* commitment and passion to hold a certain belief. He stresses the activity of knowing over the reification of knowledge and views it as 'an active comprehension of the things known, an action that requires skill'. (see Polanyi, 1974, p., vii-viii, p. 17, p. 143, p. 252-257). Whereas critical rationalist thinkers like POPPER have argued for the superiority of objective knowledge (Popper's world three, see Popper, 1994), Polanyi stresses the value of subjective knowledge and the personal co-efficient that enters every act of knowing. In doing so, he emphasizes the fact that major discoveries – new knowledge – do not always ensue from deductive reasoning, but are often the result of an initially unjustified personal belief (a process that PEIRCE has

---

[24] Davenport and Prusak (1998) provide a similar definition: Information is contextualized, categorized, calculated, corrected and condensed data that makes a difference. This definition already contains some of the information quality principles that will be presented in chapter 3, namely contextualizing, condensing (integrating), and correcting (validating).

[25] Knowledge management's prime objective in this context is the same as the mission of information quality management: to **make information actionable** or ensure that stored experiences can be used again.

[26] See Dance & Sosa, 1993, p. 234.

[27] In chapter three of this book, the first issue has been incorporated in the validation principle, the second issue in the activation principle.

labeled as abduction, see Dancy and Sosa, 1993, p. 8, p. 332).

The consequences of these two views of knowledge for the current thesis are the following ones: Information can only be of high-quality if is validated, and if it is made actionable and stimulates the prior knowledge of the individual.[27]

Examples

Two brief examples can illustrate this important hierarchy of concepts and the differences between data, information, and knowledge:

Data: When the digits 1 and 9 and 2 are *aggregated* with the help of the dot sign (.) the resulting data is a number: 1.92.

Information: When the data 1.92 is *combined* with other data to produce a coherent message, the resulting information could be the following: 1 US$ = 1.92 Swiss Francs. This piece of information states the fact that one US-Dollar is right now equivalent to one Swiss franc and 92 cents. Hence, information can be the answer to a question, e.g., what is the value of one US-dollar in terms of Swiss Francs.

Knowledge: When I *interpret* this equation and link it to my *prior knowledge*, I may realize that the Dollar is quite expensive right now in relation to the Swiss Franc. If I *apply* this fact to my current situation (my plans to travel to the United States) I may decide not to spend my vacation in the States because of the expensive exchange rate. Having *understood* this relationship, we can speak of knowledge (or, more precisely, in this case factual knowledge, know-what).

High-quality information makes it easier to transform information into knowledge, by helping to interpret and evaluate the information, by assisting the connection to prior knowledge, and by facilitating the application of the information to new contexts. Thus, increasing the quality of information means increasing the probability that information is again turned into knowledge.

Agosta, 2000, cites the following examples to illustrate the distinction between information and data:

An example of data is a knock at the door. The information presented is that my colleague has arrived for a visit. Likewise, that it is 55 degrees Fahrenheit today is data. That it is warm for this time of year is information.

He concludes that as soon as data is evaluated as to its *quality*, the data becomes information. Thus, we see another link between quality and information.

Knowledge Work Definitions

Based on the aforementioned distinctions (and the discussion of knowledge as a balance between objective insights and subjective opinions) we can now define how knowledge is used. The application of knowledge can be described from the point of view of the knower, i.e., who is applying it (e.g., the knowledge worker) or how it is applied (e.g., in a knowledge-

intensive process). Thus, we are first going to define what a knowledge worker is (and what constitutes as knowledge work) and then outline our understanding of knowledge-intensive processes.

Professionals that are typically considered as knowledge workers are consultants, lawyers, professors, engineers, or managers (see Schultze, 2000). But what is it that makes these professionals 'knowledge workers' (a term that was coined by Peter F. Drucker)? Despress and Hiltrop offer the following definition of knowledge workers:

> Knowledge workers manipulate and orchestrate symbols and concepts, identify more strongly with their peers and professions than their organizations, have more rapid skill obsolesce and are more critical to the long-term success of the organization (Despress & Hiltrop, 1995).

A synonym of knowledge workers is therefore symbolic analyst, i.e., somebody who analyses and manipulates (changes) symbols. The simplest definition of a knowledge worker is of course 'somebody who performs knowledge work'. While this is certainly true, it does not help in our understanding of the concept. Hence, we have to look at definitions that describe knowledge work itself. Following are six such definitions that show the scope of meanings that are associated with the term knowledge work:

> Knowledge work is human mental work performed to generate useful information. (Davis & Naumann, 1997.)

> Knowledge work is the production and transmission of knowledge. (Sther 1994)

> Knowledge Work is any creative systematic activity undertaken in order to increase the stock of knowledge of man, culture and society, and the use of this knowledge to devise new applications. It includes fundamental research, applied research, and experimental development work leading to new devices, products and processes (Unesco Definition cited in: Despress & Hiltrop, 1995, p.14).

> Knowledge work is the skillful collection, analysis, synthesis, and application of information. (Eppler 1998)

> Knowledge work involves analyzing and applying specialized expertise to solve problems, to generate ideas, or to create new products and services (Zidel 1998).

> Knowledge work can be classified as white collar work which uses and produces information, but unlike service work is not scripted, but

creative and often non-routine. It relies on idiosyncratic knowledge and requires formal education (abstract, technical, theoretical knowledge). Knowledge workers need to constantly balance objective and subjective knowledge (Schultze, 2000).

What all of these definitions have in common is the fact that knowledge work focuses on information that has to be made useful; hence, the connection between knowledge work and information quality. The constant balance between objective and subjective knowledge mentioned by Schultze was already discussed in the definition of knowledge. The fact that knowledge work is not scripted, however, has not been addressed yet. Although knowledge work is often non-routine and creative, as Schultze writes, it can nevertheless be sequenced and broken down into specific steps. We refer to this standardization of knowledge work as knowledge-intensive processes. They are defined in the next paragraph.

Knowledge-intensive Process

We define a knowledge-intensive process as a productive series of activities that involves information transformation and requires specialized professional knowledge. Knowledge-intensive processes can be characterized by their often non-routine nature (unclear problem space, many decision options), the high requirements in terms of continuous learning and innovation, and the crucial importance of interpersonal communication on the one side and the documentation of (high quality) information on the other. Knowledge work can thus be organized into knowledge-intensive processes. This organization includes the management of knowledge. Knowledge-intensive processes require three types of knowledge: Knowledge about the process (which steps to follow), knowledge within the process (which information has to be gathered or used), and knowledge derived from the process, e.g., experiences and insights from the completed steps (see Eppler et al., 1999).

Table Overview

The following table provides a summary of the discussion so far and compares central parameters of quality management, information quality, and knowledge management. It highlights the fact that information quality in knowledge-intensive processes should combine insights from both existing domains.

Descriptors	Quality Management	Information Quality Management	Knowledge Management
Goal	Reduce errors (before they occur) and meet specifications in the manufacturing process. Increase (internal and external) customer satisfaction.	Assure that information is of high value to knowledge workers who use it in knowledge-intensive processes. The goal is to improve the usefulness and validity of information.	Turn information into actionable knowledge, foster innovation, enable learning from mistakes and best practices and promote effective knowledge sharing. Value and exploit intellectual capital.
Object	Products (and processes) in the manufacturing context, at times also (scripted) services	Knowledge content (information that answers how and why questions)	Implicit and explicit knowledge (know-how, know-what, know-why, know-who) in all forms
Dominating mind-set	Positivistic, rational, Cartesian, quantitative	Systemic, qualitative and quantitative	Constructivist, systemic, qualitative
Implicit assumption	Quality affects the entire organization and every employee: it must be specified, measured, and continually improved through training, team work, and process adaptations.	To describe the quality of information, one can rely on a finite number of criteria which can be grouped into several meaningful dimensions.	Knowledge is often tacit (difficult to articulate) and thus depends on the individual. KM must connect knowledge workers, elicit their knowledge, map their skills, and use their experience.
Time of initial development	Early 1950s	Late 1990s	Early 1990s
Tools and methods	Ishikawa diagram, Pareto chart, House of Quality, Quality Function Deployment (QFD), Six Sigma and statistical analysis tools, Failure Mode and Effects Analysis (FMEA), Quality circles, audits, process manuals	Information quality frameworks, surveys, policies and guidelines, training seminars, portals, abstracts, review work-flows, monitoring software, indicators and standards, rating and ranking schemes	Knowledge maps, expert directories, groupware applications, document manage-ment systems, retrieval and mining software, debriefing workshops, intellectual capital reports, collaboration forms such as communities of practice
Management cycle	Plan – do – check – act (Deming)	Define – measure – analyze – improve (Wang)	Set goals – identify – store – develop – share – use – measure knowledge (Probst)
Advocates/ thought leaders	Deming, Ishikawa, Juran, Crosby, Imai, Feigenbaum	Wang, English, Redman, Lesca, Königer, Reithmeyer	Nonaka, Davenport, von Krogh, Probst, Sveiby

Table 3: A comparative view of Quality Management, Information Quality and Knowledge Management

Conclusion

If one studies the quality of information in knowledge-intensive processes, the context of one managerial discipline may not be enough. While both, knowledge management and quality management have developed powerful concepts in their respective fields, only a combination of both approaches can lead to superior results. A combination must take into account the differing mind-sets and definitions and accommodate them to the new context of knowledge work.

2.2 A Survey of Information Quality Problems in Knowledge-intensive Processes

The principal summary point to make is that the major problems in future information systems will revolve around the processes of reducing the amount of and raising the quality of information brought to the attention of the user.[28]

ROBERT S. TAYLOR

Data vs. Information Quality Problems

In the previous chapter, I have outlined how information quality management can be seen as a combination (or mediation) between quality management and knowledge management. I have argued that traditional quality management can only be directly applied in the information management domain if one focuses on data quality issues. As soon as one moves to the next level – meaningful and context-dependent information – quality management techniques alone no longer suffice. Insights from the domain of knowledge management now need to be incorporated. In order to further clarify this distinction between data quality and information quality, the following table lists typical problems from the two domains. This should illustrate the difference between the two concepts as discussed in the previous chapter (but this time from a problem-perspective).

[28] Taylor, 1986, p. 58.

Data Quality Problems	Information Quality Problems
Duplicates, multiple data sources	Conflicting recommendations in a study or analysis
Missing data relationships	Unclear causal effects in a diagnosis
Garbling (meaningless entries)	Wordy reports that have no logical flow
Spelling Errors	Untidy language that contains grammatical errors
Obsolete or outdated entries	An analysis is not updated according to recent discoveries or changes in the organizational context
Inconsistent data formats or naming conventions	Inconsistent layout or navigation structures
Misplaced data that is saved in the wrong database	Lost or 'buried' documents
Complicated query procedures	Difficult information navigation and retrieval
Wrong data coding or tagging (adding wrong meta-data)	Inadequate or insufficient categorization (in-sufficient meta-information or contextual attributes)
Incorrect data entries because of lack of source validation	Unsubstantiated conclusions with inadequate evidence
Manipulation of stored data (deletion, modification)	Manipulation of decision processes (overloading, confusing, diverting)

Table 4: Data quality versus information quality problems

Key Differences

Data quality is often managed with traditional total quality management approaches and the first experiences show that these may indeed offer feasible solutions for many database or data warehouse applications (see English, 1999, Redman, 1996, Huang et al, 1999).

Having clearly differentiated data from information, we can easily see that information quality problems are distinct from data quality problems. This is particularly evident when one looks at the solutions to these problems.

Data Quality vs. Information Quality Solutions

Whereas data quality problems can be resolved through data cleansing algorithms, data profiling programs, stabilization algorithms (e.g., phonetic manipulation and error correction), statistical process control, or dictionary matching routines (see Strong et al., 1997, or Agosta, 2000), information quality problems can often not be solved through automated processes. They require (as do some data quality problems) fundamental analysis of business issues, a change in work practices and process redesigns, an analysis of the involved information community and its expectations and skills, an evaluation of the relevant knowledge domains and their attributes, as well as a

rating of the content management process and infrastructure. Typical remedies for information quality problems may include design guidelines, publishing policies, authoring training, source validation rules, the purchase of additional information services and infrastructures, a re-design of the review and feedback process, etc.

Section Goals

In the next three sections we will look at typical *information* quality problems, as they are described in the relevant literature, or as we have encountered them in our research. This will not only ground the subsequent work in the problems that need to be resolved, but it will also help to structure and categorize the encountered challenges.

2.2.1 Information Quality Problems in Overview

Section Goal

In this section we look at information quality problems as they have been described and compiled in the relevant academic literature. For that purpose several books and articles that treat the topic of information quality (and knowledge work) problems are reviewed and summarized. The three main criteria for inclusion of a text in this literature survey are the perspective (a general approach rather than a very focused one), the explicit connection to the topic of information quality and knowledge work, and the application within a corporate context. The analysis of these texts, particularly how they categorize information quality problems, will contribute to the problem taxonomy that is developed in section 2.2.2.

Garvin's Three Information Quality Problems

In his analysis of information quality problems, GARVIN distinguishes three types of major information problems (Garvin, 1993, pp. 50-58): First, *biased* information, that is to say information that is inaccurate or distorted due to the interests or motives of the source or information transmitter. Second, *outdated* information that is no longer current due to its tardy delivery or a failure to update it. While the first problem related to content and the second to time, the third relates to (inadequate) format. Garvin refers to this problem as '*massaged'* information. According to Garvin, 'massaging is the putting together of data in a manner that applies to a particular problem at hand.' The problem that lies in this massaging is that the very same information - when massaged (or represented) differently - may lead to different (and sometimes inadequate) interpretations. While Garvin's problem survey is very simple and narrow in scope, it does highlight the fact that information quality problems can be split into content, time and format problems. Garvin does not however, stress another important problem

⇒ Content, Time, and Format Problems

dimension: that of *cost*. The same is true for the next information quality approach by Lesca and Lesca (1995). As we will see below, Lesca and Lesca have analyzed a variety of information quality problems.

Lesca and Lesca's Information Maladies

A more complete (yet still general) analysis of information quality problems was conducted by Lesca and Lesca in 1995. It revealed eight clusters of information 'maladies', as the authors have labeled them. Below, we summarize the symptoms and causes of these eight information quality problems, as they are described by Lesca and Lesca, two French professors of information management (see Lesca & Lesca, 1995, pp. 75-165).

1. *Limited usefulness* of information due to an *overload*[29] of information caused by a lack of cleansing or maintenance activities or by neglecting analysis and synthesis. As counter-measures the authors suggest filtering according to relevance criteria, prioritization and hierarchical structuring of information.
2. *Ambiguity* of the provided information leading to differing or wrong interpretations due to lacking precision or accuracy, the use of abbreviations or jargon, or simply different points of view. Counter-measures suggested by the authors include the use of a glossary, establishing feedback loops, and using richer communication media.
3. *Incompleteness* of information that can lead to inadequate decisions. The main causes for this problem are the fragmentation of work and the resulting specialization that leads to fragmentation of information. Main causes for this fragmentation are infrequent communication and exchange of information between specialists, incompatible IT-systems, and an information management strategy that is not aligned with the business strategy. Counter-measures focus on these three areas.
4. *Inconsistency* of information that leads to confusion. Causes for inconsistencies or contradictory statements are a lacking co-ordination between information authors and distributors, unclear responsibilities, or the use of multiple, inconsistent, information sources. Counter-measures focus especially on clear responsibilities and co-ordination mechanisms.

[29] The connection between information overload and lacking information quality is also pointed out by Simpson and Prusak, 1995.

5. An *inadequate presentation format* that leads to expensive conversion tasks. The main problem lies in the fact that information is not presented in an order, format or style that allows for a direct use, hence conversion is necessary. Causes for this problem are insufficient dialogue between information producers and consumers, constant time pressure, and a lacking adaptation of information to usage needs or styles (which would be the most effective countermeasure).
6. The information is *not reliable* or *trustworthy*, i.e., there is a great risk of errors, and the information's background cannot be checked. Causes for this problem are mistakes in the information production and distribution process, as well as unidentified sources.
7. The information is *not accessible*. It is lost over time because of unclear responsibilities or technological changes. This can demotivate staff and lead them to wrong decisions.
8. Finally, a big problem that Lesca and Lesca see, is the *distortion of information*, e.g., when the original message is no longer the same when it is received. Causes for this problem are too many intermediaries, too much specialization and jargon, or even voluntary distortion (e.g., misinformation) such as modifying, delaying or blocking the information to harm the receiver.

⇒ Information Product and Process Problems

Lesca and Lesca structure these eight problems into two sections. The first one views information as a *product* with problems such as overload, ambiguity, incompleteness, inconsistency, or inadequate format. The second one views information as a *process* and contains the last three problems from the list, namely lacking reliability, lost access, or distortion. Every problem that is described in this way is also related to an information quality criterion, the criterion being the opposite of the problem, e.g., overload versus relevance, ambiguity versus clarity.

⇒ Problems as Opposites of IQ-Criteria

What this approach has shown is that information problems must not only address the final information product, but also the process that leads to this product. This is of course an insight that traditional quality management has considered for quite a long time. In addition, the approach shows that information quality problems can be seen as the opposites of information quality criteria (an approach that other authors have also followed, see below).

Information Quality 'Potholes'

The eight problems described above already provide helpful hints where and how non-quality information may cause problems. The above list is nevertheless neither complete, nor very systematic. A more systematic approach has been developed by Strong et al. (1997) and was later modified in Huang, Lee, and Wang (1999). As Garvin and Lesca and Lesca, the authors of this approach base their findings on incidents at real-life companies to illustrate the identified problems. In these approaches ten information quality problems were identified that relate to information producers, information custodians, and information consumers. The authors thus structure information quality problems along the *life cycle* of information: from origination (information producer/sources), storage and maintenance (information custodian/systems), to information use (information consumer/task environment). Since the authors examine information quality from a database and data warehouse point of view, not all identified problems are relevant to this study. Consequently, purely database-related problems[30] are left out at this point. The other problems, as well as their consequences and possible counter-measures, are summarized in the table below.

⇒ Production, Storage and Usage Problems (Life Cycle)

[30] They relate to non-numeric information and its indexing and to automated content analysis.

Problem Name	Description	Consequences	Solution
1. Multiple sources	Multiple sources of the same information produce different values and lead to confusion, less credibility and acceptance.	Use of information declines. Users distrust information. Maintenance is more difficult and costly.	Develop common definitions and consistent procedures.
2. Subjective roduction	Information is produced using subjective judgements, leading to bias.	The objectivity of the information decreases and information is difficult to evaluate.	More training, better rules, expert systems.
3. Production errors	Systemic errors in information production lead to lost information.	Information searching and correcting increases.	Process improvements, incentives, controls.
4. Too much information	Large volumes of stored information make it difficult to access information in a reasonable time.	Excess time is required to extract and summarize information.	Analyze information needs, develop regular, frequently extracted subsets of relevant information.
5. Distributed systems	Distributed, heterogeneous systems lead to inconsistent definitions, formats, and values.	Information can no longer be easily aggregated or combined, due to the format differences and incompatibilities.	Integrate systems in one platform (e.g., data warehouse) or reduce the amount of systems to one.
6. Changing task needs	As information consumer's tasks and the organizational environment change, the information that is relevant and useful changes.	Mismatches develop between available information and what is needed for tasks.	Anticipate changes in tasks and revise processes and systems before the mismatch becomes a crisis.
7. Security and privacy requirements	Easy access to information may conflict wit requirements for security, privacy, and confidentiality.	Mechanisms for security block or delay access, so the information provides less value.	Develop consistent policies and procedures for secure information.
8. Lack of computing resources	The IT infrastructure is insufficient and limits access to information.	Knowledge workers are demotivated and cannot work productively.	Develop technology upgrade policies so consumers know when to expect more resources.

Table 5: Information quality problems (compiled and adapted from: Strong et al., 1997)

⇒ Relating Problems to Several Criteria

Besides using the information life cycle to structure information quality problems, this approach also relates the problems to specific information quality criteria. Unlike Lesca and Lesca, however, who viewed an IQ-problem as the exact opposite of an individual criterion, this approach relates every IQ-problem to several criteria. Problem number three, for example, is related to the criteria of correctness, completeness, and relevancy. Problem number one is related to both consistency and believability.

⇒ Systemizing IQ Problems into Patterns

In addition to identifying a number of significant information quality problems (and relating them to IQ-criteria), Huang et al. (1999) based on Strong et al. (1997), have grouped these problems systematically in what they call 'IQ patterns.' These patterns relate various IQ-problems to one another and combine them with counter measures. An IQ-pattern thus consists of the symptoms, causes, and solutions of IQ-problems. Such a systematic arrangement can help to better understand the interdependencies between information quality problems. A pattern in this sense is a systematic problem perspective that helps users identify the problem type that they are facing and relate it to possible generic solutions. For our own information quality problem typology, we retain the idea of systematically relating information quality problems to one another and to solution elements.

Knowledge Worker Problems

Besides the information quality problems that are discussed in terms of quality criteria, there is a great body of literature on *knowledge workers* and their problems in dealing with information. The descriptions of knowledge work problems can highlight other issues associated with low information quality, since knowledge work is characterized by the fact that its input and output is information (see Drucker, 1991)[31].

⇒ Input & Output, Process, and Infrastructure Problems

If one scans the vast available academic literature and empirical surveys on knowledge workers (see for example Collins, 1997, Schultze, 2000, or Reuters, 1998), several problem areas become evident. They relate to the adequate *amount* of information (as an input factor), its *structure* and format, the working *habits* or *processes* and *workplace* parameters, the *communication media* or *IT-infrastructure*, and the *output characteristics* of knowledge workers. Especially the two first problem areas (amount and structure) are closely related to the informa-

[31] Drucker writes that 'knowledge work by definition does not yield a product. It yields a contribution of knowledge to somebody else. The output of the knowledge worker is always somebody else's input (Drucker, 1991, p. 173).

tion quality domain. They outline what happens when information is neither provided in the right amount nor in the right format. But the other three domains are also relevant to the IQ-domain, since they show in what ways information has to be integrated into the working habits and workplace of a knowledge worker. They also show how the infrastructure can cause knowledge worker productivity problems. The last problem area, output characteristics, can illustrate the potential of information quality to better specify the products of knowledge work efforts.

The following diagram illustrates the main clusters of knowledge worker problems that are discussed below.

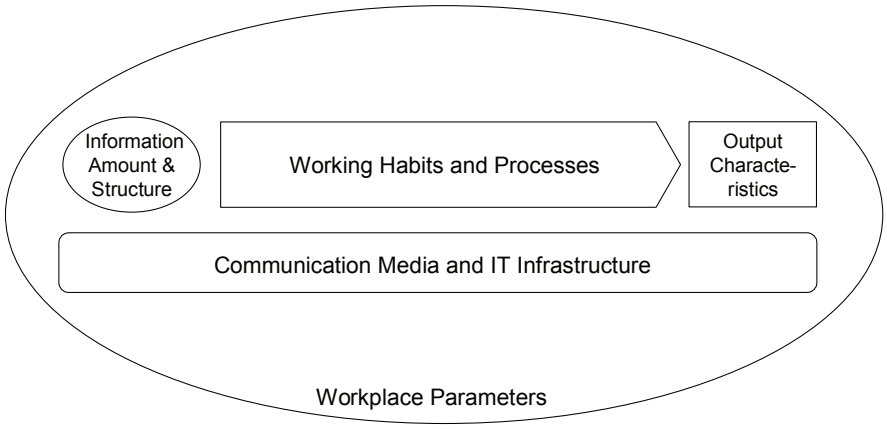

Figure 1: Knowledge work problem categories

Information Amount & Structure (Input)

That the sheer amount of information which has to be screened is a major knowledge worker problem has been documented in numerous studies (for an overview on such studies see Probst et al., 2000). There are literally dozens of investigations that have found information overload to be a key challenge of today's knowledge workers. The term information overload describes situations in which the individual is no longer able to integrate new information for decision-making, due to the great amount of information he or she is exposed to. He or she can no longer productively use the quantity of information in the available time scale. In consequence, decision quality, efficiency, and even well-being may be reduced. The reasons for this phenomenon are that the provided information does not come in the right *amount* or in the right *format* (e.g., lacking structure or focus). It is repetitious, unclear, irrelevant, or simply poorly organized. As a result of this, knowledge workers can no longer distinguish between what is important and what is not, how various pieces of information relate to one another, or which

⇒ Information Overload

information is trustworthy and which isn't. This problem area illustrates that quantity can directly have a negative influence on information quality and that the right amount of information, or its scope, are critical quality criteria. It also highlights the fact that comprehensiveness or completeness are relative terms.

Working Habits and Processes

Whereas the previous problem cluster related to information as an input factor, the problems described in this paragraph focus on information as it is used in the daily routines of knowledge workers. Here, a particular problem of knowledge workers is that they often do not act upon information. That is to say, they know what ought to be done and have all the right information, but don't act on it. Pfeffer and Sutton refer to this problem as the knowing-doing gap (Pfeffer & Sutton, 2000 and Pfeffer & Sutton, 1999). As MINTZBERG (1975) has pointed out, managers need to find a balance between information collection and action taking. If the working habits of knowledge workers consist of merely absorbing information, then they will suffer of what is called 'paralysis by analysis', a tendency to remain in the diagnostic mode and not switching to an action mode.[32] Although this problem relates mostly to the knowledge worker, the role of information quality should not be underestimated. Information that is already organized for action and stimulates a reader or user to remember and apply it, can help in overcoming this knowing-doing gap.

⇒ Knowing-Doing Gap

Workplace Parameters

The cluster labeled as workplace parameters contains such typical problems as constant distraction through co-workers (and their calls, e-mails, questions etc.), administrative overhead that needs to be managed, or generally a high workload without time for concentration and quiet work. Knowledge worker problems that relate to the workplace may also include other factors, such as an ergononomic chair and desk, the general working climate (e.g., no incidences of mobbing etc.), or the larger organizational context (e.g., reorganizations that bind resources, etc.). As far as information quality is concerned, the workplace has to make sure that information is not only provided, but can also be processed adequately (without distractions, interruptions, or other worries). This will have a direct influence on the quality of information as an output of knowledge work. How this output should be characterized is another problem, one which we look at in the next paragraph.

⇒ Concentration Increases Output Quality

---

[32] This danger is also described in one of the case studies, specifically in the description of major problems by Giga's head of research.

Output Characteristics

This problem cluster refers to the difficulties that exist in specifying the output of knowledge workers or measuring their productivity. Schultze (2000) in her in-depth analysis of knowledge work refers to this problem as the subjectivity-objectivity problem. It is based on the insight that many knowledge workers rely on idiosyncratic knowledge that they need to make understandable to their peers and superiors[33] (who may have difficulties evaluating their productivity). However, this is not an easy task. It requires a 'negotiation' of meanings, ideas, and opinions. In addition, it requires that the often intangible results of knowledge work are measured through indicators or criteria that allow for comparisons between employees and analyses over time.

⇒ Specification Enables Measurement and Comparisons

Especially in the context of interdisciplinary knowledge work teams, the lack of specification of knowledge work output can lead to major problems, such as performance gaps, lags, or simple misunderstandings. Specifying the characteristics that make up information quality for a given context can help reduce this problem cluster and it can provide management with metrics to measure knowledge worker productivity, one of the biggest challenges of management today according to PETER F. DRUCKER.[34]

Communication Media and IT Infrastructure

The last problem cluster relates to the communication and information infrastructure that is available to knowledge workers. Here, problems range from complicated interfaces and low user-friendliness of these systems to incomplete deployment, slow performance, or insufficient control (e.g., the systems are not interactive enough). Employees find that knowledge work systems (see Laudon & Laudon, 2000), which should support them in their tasks, are not adjusted to their information needs or inconvenient to use. Standard IT-tools or communication media are purchased and not sufficiently adapted to the context of knowledge workers. The knowledge workers, in return, do

⇒ User-centric Design and Training

---

33 This topic is also discussed at length in Polanyi (1974) as stated in the previous chapter.

34 Drucker (1994) views the two issues of quality and productivity as closely related. He writes: "Wee need systematic work on the *quality of knowledge* and the *productivity of knowledge* – neither even defined so far" (see also Drucker, 1991). In related research, Cornell and Brenner write that "many have tried to measure knowledge worker performance, but with little success." One of the reasons they give for this is that the output is often intangible and abstract. "You can't hold, weigh or put a caliper to the output of a knowledge worker" (Cornell & Brenner, 1993).

not receive enough training to master the offered IT-tools and use their functionalities. In consequence, the communication and information infrastructure contributes to information overload instead of reducing it through filtering mechanisms. The potential of these tools to improve information quality (e.g., making information more accessible, current, and easy-to-use) is hence often not fully exploited. As stated above, this is mainly due to design problems and insufficient training.

If we compare this analysis of knowledge work problems with that of other authors, we see that the presented framework does in fact cover the dominant knowledge work problems. Especially the problem cluster around 'output characteristics' dominates the current discussion of knowledge work.

Other Approaches

Pfiffner and Stadelmann (1995) emphasize the following two knowledge workers problems:

First, Pfiffner and Stadelmann see the difficult *control* of the working progress by the superior as a major problem (because of mostly *abstract* results). We have discussed this issue in the 'output' section above, where we stated that information quality criteria can provide indicators of whether a knowledge worker provides high or low quality work as far as his or her information is concerned (e.g., whether it is accurate, useful, timely etc.). Second, *lags* in the articulation and distribution of new knowledge cause many productivity problems (mainly because the assimilation of knowledge – learning – is always time consuming). The problem of lags was also addressed in the output section. One reason for process lags are time-consuming adjustments because of lacking initial information specifications. Lags of this type can be avoided through information quality specifications. Additionally, the use of high-quality learning content can accelerate the learning process and reduce lags in the application of new knowledge.

Another analysis of knowledge work problems can be found in Davenport et al. (1996). The authors describe the following seven problems associated with knowledge work:

Knowledge workers struggle with the multitude and *insecurity* of information *inputs* and *outputs*. Again, this problem can be directly linked to our discussion of input and output parameters.

Managers struggle with the *unstructured* and highly individual working rules and *routines* of knowledge workers. In our framework in the figure above, we have indicated that individual working habits and processes may cause problems. A way to reduce these individual (and at times conflicting) rules consists of specifying common criteria which everybody should adhere to.

It is difficult to separate work process, result, and original input due to the unstructured nature of knowledge work. For this problem, we have proposed to describe knowledge work in terms of knowledge-intensive processes that offer a greater degree of structure.

There are no clear indicators by which one could measure knowledge worker performance. This problem relates again to output characteristics and to finding reliable indicators.

Every knowledge worker demands a high degree of autonomy. This can lead to lacking coordination. This problem refers to our "habits and process" section.

There are great differences in terms of performance over time and between knowledge workers. As mentioned above, this is a problem of measuring the output of knowledge work.

Knowledge workers often struggle with lacking IT-support. This issue has been explicitly discussed in our last problem cluster. The reasons given were inadequate training and design

From these discussions of knowledge work problems we see that measurement and control problems are consistently affected, two activities that are at the heart of management. For both of these issues, information quality criteria that help to specify the outcome of knowledge work may offer feasible solutions.

⇒ Measurement and Control Problems

Literature on information quality problems categorizes these challenges according to their dimension (content, format, time), their view (information as product or process), or their phase in the information life cycle (production, storage, use). Some approaches directly relate the problems to information quality criteria.

The literature on knowledge work problems revealed that major problems are information overload (resulting from inputs that are too frequent and in an inadequate format), a knowing-doing gap due to the working habits of employees, frequent interruptions, an inadequate IT infrastructure, and missing measurement or control mechanisms (output specifications or indicators).

Conclusion

2.2.2 Information Quality Problem Patterns

You do not understand anything until you understand it in more than one way.

MARVIN MINSKY

Goal of this Section

Having provided an overview on the range of possible information quality problems, we can now summarize the insights from the previous sections in a systematic problem classification. The resulting problem typology – which will reduce the great number of encountered problems to smaller groups – will provide helpful categories or patterns for the framework that will be developed in the next chapter.

Review:

Categorizations of Information Quality Problems

When one analyzes the problems discussed in the previous sections, one can detect various ways of framing or categorizing various information quality problems. There is the abstract problem description that relates information quality problems to either information *content*, *timing*, or *format*. Then, on a more specific level, there are the problems themselves which are often described in terms of (deficient) information *attributes*, e.g., whether information is correct or flawed, to the point or rather wordy, delivered in a timely and convenient manner, consistent or not etc. These specific IQ-problems can be related to either information as a *product* (stressing stable 'entity' aspects of information) or information as a *process* (stressing the dynamic communication aspects).

Another such abstract way of categorizing information quality problems consisted of relating them to the *life cycle* of information from production (or input) problems to consumption (or output) and deletion problems. Yet another way of categorizing the problems consisted of relating them to certain '*master problems'* that represented a whole series of related issues. Problems that were mentioned in this context were information overload (or the 'volume problem'), the knowing-doing gap, and design and measurement problems. Another (crude) distinction that was made in categorizing information quality problems was the one between *real* information quality and *perceived* quality of information. Then, there was the categorization of problems based on the various information *stakeholders*. Problems affecting information producers (such as difficult publication procedures) were distinguished from those that information administrators face (such as maintaining an information repository), or the problems that information consumers face (e.g., judging the

credibility of information). Finally the two categories of *content quality* and *media quality* were introduced to distinguish the different origins of information quality problems.

The section on surveys and focus groups did not yield any additional categorizations, but rather indicated which information attributes represent the greatest problem areas. Based on these crucial areas and the discussed categorizations, we will now group the multitude of problems to sets that allow systematic counter-actions.

Three Bases of Categorization:

Origin, Consequence, Responsibility

In order to structure the problems encountered previously, we can look at them from three insightful perspectives. We can categorize the problems according to their *origin* (i.e., what causes the problems), according to their *consequences* for the information consumer, or according to the *responsibilities* for solving the problems. Below, we structure the encountered problems in those three ways. They seem especially apt because they represent a logical sequence from cause to consequence to remedy.

Problem Origins

As far as the origins of the various information quality problems are concerned, one can distinguish four possible causes. The first cause for information quality problems can be that the provided information is *not targeted* at the community that is supposed to use it. In other words, the problems exist because information is addressed to the wrong audience. This can result in irrelevant, incomplete, or simply not useful information for the information consumers. A second cause of information quality problems may simply be that the information producers create 'bad' information that is often incorrect, inconsistent, or outdated. The origin is not a wrong allocation of the information as in the first cause, but already a *wrong production* to begin with. A third possible cause may be that the information is provided in the *wrong manner*. The information may be correct and targeted to the needs of the information consumer, but it may not be provided in the right way or through the right process. A final, fourth possible origin or cause for a great number of information quality problems may be the *infrastructure* on which information is provided. If the hard- and software on which any information system is built (whether electronic or print-based) is not reliable, the information may not be accessible or secure.

Let me summarize these causes as the following four generic information quality problem categories:

1. **Irrelevant Information**: selection, filter, and profiling problems.
2. **Unsound information**: source or authoring problems.
3. **Sub-optimal content management process**: workflow and management problems.
4. **Unreliable infrastructure**: hardware and software problems.

We can relate many information quality problems to these four causes. We can also articulate strategies against these four types of problems. These counter measures are to align information better to the needs of the user community (problem area one), to improve the training for the information producers and enforce certain minimal quality standards (for problem area two), to improve the co-ordination and the workflows in the content management processes, and to improve the hard- and software in order to make it more reliable and efficient. The following diagram summarizes this view.

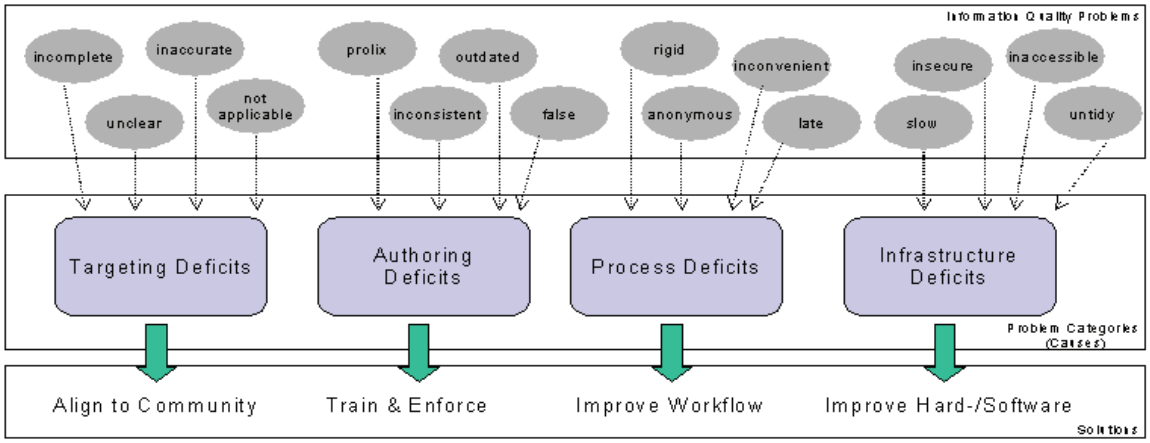

Figure 2: IQ Problems categorized in terms of heir origins

Problem Consequences

Whereas the above figure illustrates the means to overcome problems based on specific root causes, the next categorization helps to resolve information quality problems by analysing their *consequences*. Here, one can also distinguish four main issues. The following four consequences are a direct result of lacking information quality. They are all formulated from an information consumer's perspective:

1. **Information Overload**: I cannot identify the right information.
2. **Misjudgment**: I cannot judge or evaluate the information.

3. **Misinterpretation**: I cannot understand or interpret the information.
4. **Misuse**: I cannot use or apply the information.

Here is the logic behind these four categories: If an information consumer cannot *identify* (find and access) information, then this may be because it is difficult to get an overview on the available information because it is too dispersed, too vast, or because it is incomplete, inconvenient or impossible to access. Once the information has been found, it has to be evaluated by the user. If an information consumer cannot *evaluate*, judge, or trust a piece of information this may be due to inaccurate or inconsistent statements, or due to an insecure platform on which the information was found. If an information consumer can evaluate the credibility of an information source, but he or she cannot *interpret* a piece of information itself, this could be the result of lacking clarity, or simply because the information is not correct or does not refer to its sources. It may also be because the information is not kept current and not maintained and hence the user no longer knows what is still relevant and what has become obsolete. Finally, if the information consumer is not sure how to *apply* the information he or she has been provided with, then the resulting misuse could be the consequence of a wrong information format, a wrong timing or a wrong application context. Once again, a diagram can be used to illustrate this logic.

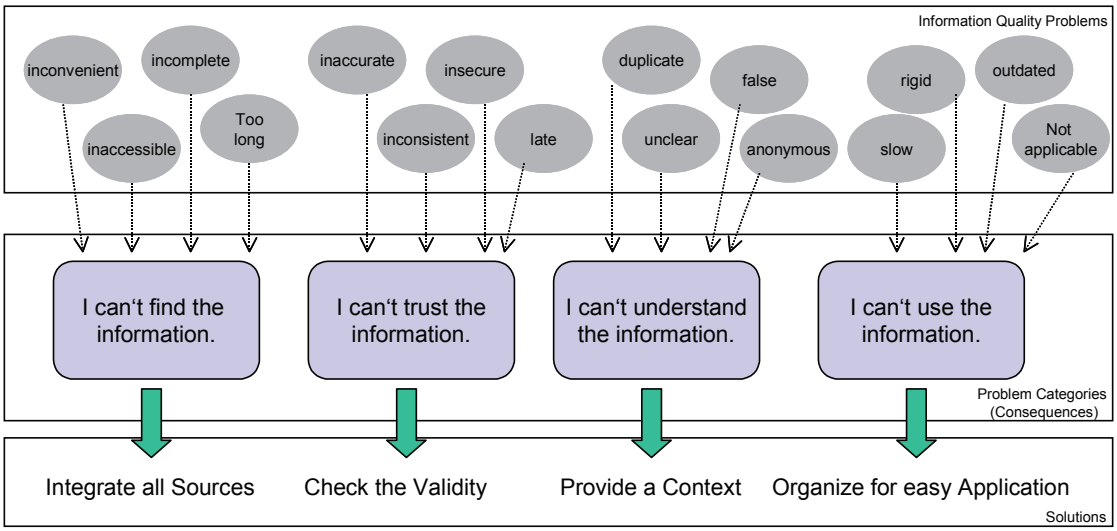

Figure 3: IQ Problems categorized in terms of their consequences for the user

Countermeasures

As the diagram shows, every one of the four problem categories has a specific counter-measure that can help to overcome it. In order to resolve the identification problem, sources need to be integrated or aggregated (e.g., through portals, maps, directories, summaries, etc.). To increase the trustworthiness of information its validity and background must be checked. To foster a better understanding of the information, a context must be provided that allows for an adequate interpretation of a piece of possibly isolated information. Finally, to avoid a misuse of the information or its non-use, it must be re-organized for easy application, that is to say it must be provided in a (interactive) format that can be directly used by the information consumer. An example of this would be a bank statement with account information that can be directly used for income tax forms.

Responsibilities

We have now grouped information quality problems according to their main causes and their main consequences. A third and final possible categorization criterion is the responsibility for the problems – that is to say who should do something about them. Here, we can distinguish between three professional communities: the information producers or *authors*, their superiors or *managers*, and their support staff or *IT-managers*. If the information quality problems result from providing the wrong kind of information (see the problem origins) then the managers must get authors to produce a different kind of information. If the information is relevant, but often false, outdated, or inconsistent, then the authors need to improve their content either on their own or with the help of their management. In contrast, if the way that information is provided is sub-optimal (slow, complicated, untraceable), then the information technology managers need to become active. The same holds true for deficits of the infrastructure, whether it is insecure or simple difficult to maintain. Thus, we can categorize information quality problems as either *content problems* that must be resolved by the information producers and their management, or as *media problems* that need to be resolved with the help of the information technology department that should improve the content management processes and infrastructures. The figure below illustrates this dual responsibility for information quality in a corporate setting.

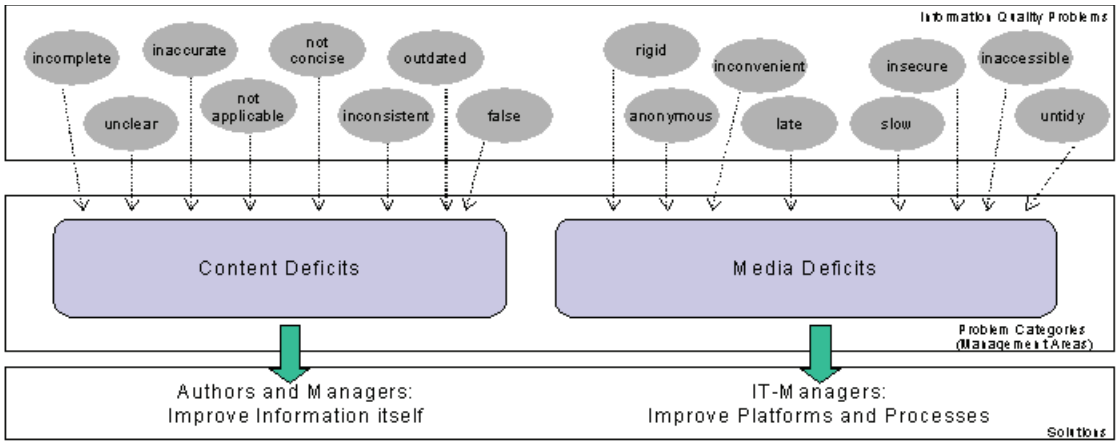

Figure 4: IQ Problems categorized in terms of their responsibility

Resulting Definition

Based on the three ways of categorizing IQ problems, we can now formulate the following definition of information quality problems:

Definition: Information Quality Problem

We define an *information quality problem* as a situation in which the content or medium of information does not meet the requirements of its producers, administrators, or users. It can be expressed in terms of information attributes that are clustered according to their origin, consequence or responsibility.

Information quality problems are the result of either providing the wrong kind of information (i.e., *irrelevant* or incomplete information), providing simply wrong information (i.e., *unsound* information that is inconsistent or incorrect), providing the information in the wrong way (i.e., in a non-convenient and rigid *process*), or providing the information on the wrong kind of *infrastructure* (e.g., not reliable or secure).

Information quality problems ultimately result in higher costs and a lower value of information use; in other words: the greater the information quality problems, the less informative the medium and its content.

Use of the Definition

This definition can help to better frame, delineate and analyze information quality challenges. It can further help to devise effective solution strategies by identifying the origins, consequences, and responsibilities of such problems.

Transition to the Next Section

We have now thoroughly analyzed the problem side of the information quality issue. In the next section, we are looking at the solution side in terms of already existing frameworks. We will examine how well they structure the possible information quality characteristics and their inherent challenges.

Conclusion

By categorizing the observed information quality problems according to their origin, their consequences for the information consumer, and their remedies or responsibilities, we have classified the main problem drivers and possible countermeasures. As far as *origins* are concerned, we distinguished information quality problems due to the lacking value of information for a specific community, problems due to intrinsic information deficiencies, problems caused by an inadequate content management process, and problems due to infrastructure deficits. As far as the *consequences* are concerned we distinguished information quality problems that made the efficient identification of information difficult, problems that lead to a wrong evaluation, problems that lead to a wrong interpretation or allocation of information, and problems that made the information difficult to use. The *responsibilities* for these problems can be split between information producers, line-managers, and information (system) administrators.

2.3 Information Quality Frameworks in Review

2.3.1 Recent Models from Various Domains

In the information quality field, researchers from such distinct domains as media studies, data warehouses, corporate communications, on-line publishing or knowledge management have pondered the question of what can be qualified as "good information". Regardless of the great differences of their research contexts, goals and methods, these scholars have built what seems at times an astonishing consensus in regard to the criteria that can be used to describe the value of information products such as newspapers, databases, websites, Intranets, or technical manuals.

Conceptual frameworks of information quality abound in management, communication, and information technology literature. In our review of information quality literature from the last ten years, we have found twenty information quality frameworks that define and categorize quality criteria for information (i.e., adjectives that describe information characteristics which make information useful for its users) in various application contexts (see Table 1). In this evaluation of information quality frameworks, we have found it useful to