A Generic Framework for Information Quality in Knowledge-intensive Processes

Martin J. Eppler
Institute for Media and Communications Management (=mcm *institute*)
University of St. Gallen
Blumenbergplatz 9
9000 St. Gallen
CH Switzerland
Phone: +41 71 224 24 07
Fax: +41 71 224 27 71
E-mail: Martin.Eppler@unisg.ch

Abstract: Based on an evaluation of existing information quality frameworks, action research with seven partner companies, and empirical surveys among practitioners, this paper proposes an information quality framework that overcomes some of the analyzed deficits of current approaches. The proposed framework is especially apt for the context of knowledge-intensive processes, such as market research, product development or consulting. It consists of four quality levels, namely community or relevance criteria, product or soundness criteria, process criteria, and infrastructure criteria. A main element of the framework are information quality principles which indicate how the criteria contained in the four levels of the framework can be improved. The paper outlines the basic elements of the framework and indicates how it can be applied in various knowledge-intensive processes. Two short case studies are provided, one from a market research company, and one from a book abstraction service. The paper concludes by pointing out future research needs in the domain of information quality frameworks, such as finding measurable and instructive quality indicators on all four levels.

1. Introduction: The Rationale for a New Information Quality Framework

Organizational scientists should be viewed not as engineers offering technical advice to managers but as providers of conceptual and symbolic language for use in organizational discourse.

(Astley, Zammuto, 1992, p.443)

In a prior study related to information quality frameworks, we have analyzed information quality frameworks from the last twelve years and found that they required further development in five areas in order to be useful for practitioners and researchers alike (Eppler, 2001). These improvement areas are:

1. The applicability of the frameworks in more than just one area, e.g. improving their *generic* nature.
2. The development of information quality frameworks that show *interdependencies* between different quality criteria (such as the accuracy timeliness tradeoff).
3. The inclusion of *problem areas* and *indicators* into these frameworks, as well as the inclusion of possible *solution* elements.

4. The development of *tools* which are based on the framework.
5. The development of frameworks that are at the same time *theoretical and practical* (that are at the same time rigorous and relevant, elaborate yet concise).

With the following information quality (IQ) framework, we try to overcome these shortcomings. We strive for a *generic* framework (in the sense of wide applicability) that can be used for any kind of knowledge-intensive process that has information both as an input and an output factor (this means information for direct human – and not direct machine – use; for a definition of knowledge-intensive processes see Eppler et al.)[1]. Generic also means that the framework can be used for various purposes, such as evaluation and assessment, improvement, management or monitoring. The framework will explicitly show tradeoffs between specific criteria (drawn as arrows that connect two criteria). It can be used to position information quality problems, and it consists of four principles which help to find solutions to IQ-problems (the integration, validation, contextualization, and activation principle). It will be shown that the framework is rooted in an existing theory (the media reference model) and that it can be used to provide tools such as checklists, diagnostic questionnaires or information quality guidelines. The framework is both tested against existing information quality theories and in terms of its practical applicability in the information management domain.

The goal of the present framework is neither prognosis nor precise description. It should help (as Porter points out in his discussion on the use of conceptual frameworks) to better *think through a problem* and select among strategic alternatives. Frameworks in this understanding identify the relevant variables and the questions which an analyst must answer in order to develop conclusions tailored to a particular company (see Porter, 1991, p. 955). Frameworks thus provide a *conceptual language* which practitioners can use to facilitate their mutual problem understanding and coordinate their collaborative actions (for this point, see Astley & Zammuto, 1992, p. 443). In the information quality context, a framework should thus provide a systematic and concise set of terms which practitioners and researchers can use to analyze and resolve information quality issues. The existing information quality frameworks do not provide that conceptual language or terminology since they focus on data problems in the data warehousing or information systems context, and not on the use of information by knowledge workers. This crucial distinction is illustrated in the table below. Table 1 compares typical data quality problems with those that are addressed in the context of knowledge-intensive processes.

[1] We define a knowledge-intensive process as a productive series of activities that involve information transformation and require specialized professional knowledge. Knowledge-intensive processes can be characterized by their often non-routine nature (unclear problem space, multiple decision options), the high requirements in terms of continuous learning and innovation, and the crucial importance of interpersonal communication on the one side, and of documentation (or codification) of information on the other.

Data Quality Problems	Information Quality Problems
Duplicates, multiple data sources	Conflicting recommendations or expert opinions 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	Cluttered 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 conventions or navigation structures
Misplaced data	Lost or 'buried' evidence
Complicated query procedures for a database	Difficult information navigation and retrieval in a knowledge base
Wrong data coding or tagging (adding wrong meta-data)	Inadequate or insufficient categorization (insufficient meta-information or contextual attributes) due to a lacking taxonomy
Incorrect data entries because of lacking source validation	Unsubstantiated conclusions with inadequate evidence
Manipulation of stored data (unauthorized deletion or modification of entries)	Manipulation of decision processes (overloading, confusing, diverting attention)

Table 1: Data Quality versus Information Quality Problems

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 or questions, a change in work practices or process designs, an analysis of the involved information community and its expectations and skills, an evaluation of the relevant knowledge domains and its attributes, as well as an evaluation of the content management process and infrastructure. Typical remedies for information quality problems may include information 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.

Having outlined *why* a new information quality framework is necessary for the described context, we can now analyze *how* such a framework can be established.

2. Legitimacy of the Framework: Six Validation Areas

We believe that a *prescriptive* or normative framework such as this one, which does not describe how something does work or will work, but how it *should* work, can be legitimized or validated in the following six ways. They are at the same time the six empirical and theoretical bases of the framework presented in this study.

1. Through a solid existing *theory* which is used to generate the framework and in-/exclude certain elements: in this case the four levels or views of the framework are based on the

knowledge media theory of BEAT SCHMID, chair of communication management at the University of St. Gallen (see Schmid 2000).

2. Through *feedback from practitioners* about its usefulness: the framework presented in this paper has been discussed with practitioners in the fields of consulting, market research, corporate communications, and product development over the course of three years. It has been used to analyze real-life information quality problems and improve information products. The practitioners were given the chance to provide feedback on the framework's design and its components.

3. Through *comparisons with other frameworks* and their deficits: the current framework has been developed based on an analysis of more than twenty information quality frameworks from the last twelve years (see Eppler, 2001).

4. Through the evaluation of the framework through *meta criteria* (see Roehl 2000): The criteria which were used to evaluate the current framework are the following six: 1. Precision (all included terms are clearly defined) 2. Positioning (the context of the framework is made clear) 3. Consistency (the elements of the framework are mutually exclusive and collectively exhaustive) 4. Conciseness or parsimony (the framework uses a relatively small amount of elements) 5. Illustration (the framework can be illustrated through examples) 6. Practicality (the framework can be used as a tool to improve real-life problems) (see Eppler, 2001).

5. Through the use of *empirical surveys*:[2] the relevance of the included criteria has been tested by several surveys among employees of different companies in the consulting and market research context (see Brocks, 2000). In these surveys, the importance of the included criteria has been rated by the practitioners in relation to the process steps that they are involved in. As a result of these surveys, some information quality criteria were no longer included in the framework (such as believability), while others were added to it (such as applicability).

6. Through multiple case studies: As Eisenhardt (1989, p. 545) has pointed out, a conceptual framework can often be the result of case study research. Hence, case studies are a feasible way to see whether a framework does in fact fit with the reality of the corporate world or not. Two such case studies are presented in the last section of this paper. Those and others from the domains of consulting and financial services have contributed to the present framework.

Having stated the six validation possibilities that exist for a conceptual (prescriptive) framework, we can now turn to the framework itself and its elements.

3. Elements of the Framework: Views, Phases, and Principles

The present framework consists of three major elements: The first element is the framework's vertical structure. It consists of *four views* on information quality that categorize crucial information quality criteria according to their relation to the target community, the information product, the information process, and to its infrastructure. The second element of the framework is the horizontal structure, which is divided into *four phases*. The four phases represent the life cycle of information from a user's point of view: it is searched and found, evaluated, adapted to a new context, and applied. The third major element of the framework are the *management*

[2] See for example the empirical research referenced in: Huang, Lee, Wang, 1999.

principles. They help to improve the quality of information in every phase. Below, these three major elements of the framework are described in more detail.

The overall *vertical* structure of the present framework is derived from an implicit convention which most recent information quality frameworks seem to follow, namely that they are divided into two sections: a category section, and a criteria (or dimensional) section. Thus, in most IQ-frameworks, the individual quality criteria are grouped into fewer information quality categories. These categories typically do not include qualifiers, but often have standard names such as intrinsic IQ (Wang & Strong, 1994) or quality of structuring (Königer & Reithmayer, 1998). In the current framework this twofold structure is also used, but with qualifying category names that already include a quality criteria on a higher level. The four IQ-categories or views are:

1. *Relevant information*: This category relates to whether the information is comprehensive enough, accurate enough, clear enough for the intended use, and whether it is easily applicable for the problem at hand. This category is also called *community view* since the relevance of a piece of information depends on the expectations and needs of a certain (writer-, administrator-, or user-) community.

2. *Sound information*: This second category contains criteria which describe the intrinsic or *product* characteristics of information, such as whether it is concise or not, consistent or not, correct or not, and current or not. Whereas the criteria in the first category (relevance) are subjective (indicated through the term "enough") these should be relatively independent of the targeted community (indicated through the term "or not").

3. *Optimized Process*: The third category contains criteria which relate to the content management process through which the information is created and distributed and whether that process is convenient (for writers, administrators, and users), and whether it provides the information in a timely, traceable (or attributable), and interactive manner.

4. *Reliable Infrastructure*: The fourth and final category contains criteria which relate to the infrastructure on which the content management process runs and through which the information is actually provided. Reliability in this context refers to a system's easy and persistent accessibility, its security, its maintainability over time (including aspects of cost-efficiency), and its high (and continuous) speed or performance.

The logic behind these four categories or IQ-views is based on the knowledge media theory of BEAT SCHMID (see Schmid 2000). It states that any knowledge media (in the sense of a platform that enables the transfer of knowledge) design must begin with the analysis of the community of people who need to share knowledge, and analyze their needs, activities, and work practices. Then, the services and information objects which have to be provided to that (and by that) community need to be analyzed and a process has to be designed in order to deliver these information services or information objects. Only then can the infrastructure requirements and parameters be determined. Thus, the following framework is usually used in a top-down approach.

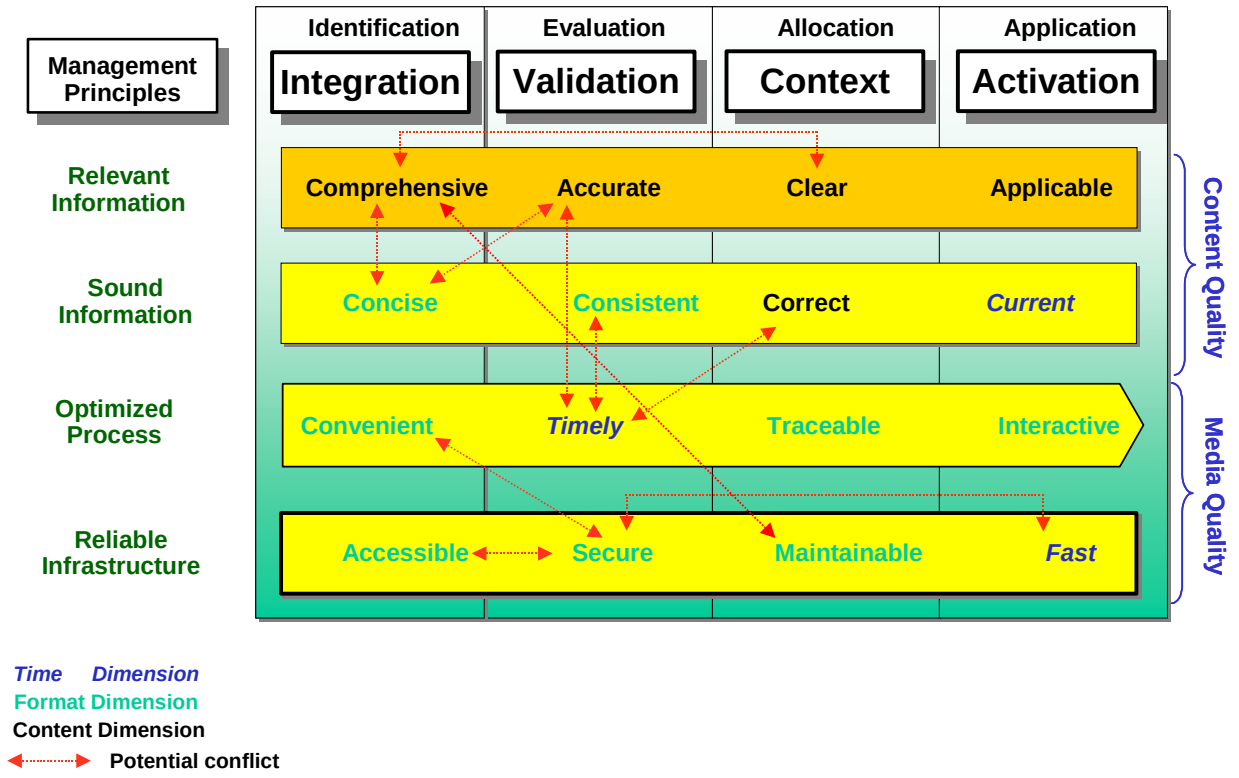

Figure 1: The Information Quality Framework

As figure one shows, the upper two levels of the framework are labeled as content quality, while the lower two are referred to as media quality. The first two categories, relevance and soundness, relate to the actual information itself, hence the term *content quality*. The second two categories, process and infrastructure, relate to the management of that information, and whether the delivery process and –infrastructure are of adequate quality, hence the term *media quality* which stresses the channel in which information is transported. For the end-user, both segments, media and content quality, may be perceived as one final product – the information and its various characteristics. For the information producers and administrators, however, this difference is crucial, since the authors usually cannot influence the media quality, and the administrators only have limited possibilities of influencing the content quality.

The *horizontal* structure of the framework incorporates a chronological sequence (or phases) from the user's point of view. For him (or her) information may be the answer he needs to find, understand and evaluate, adapt to his context and apply in the right way. Thus, a knowledge media should assist him in identifying relevant and sound information. It should help him to evaluate whether the information is adequate for his purpose. It should assist him in re-contextualizing the information, that is to say understand its background and adapt it accordingly to the new situation. Finally, the knowledge media should provide assistance in making the found, evaluated, and allocated information actionable, e.g., use it effectively for decision making. In terms of the key questions of an information consumer which are answered in each phase, they can be described as follows: 1. Where is the information I need? (identification) 2. Can I trust it (evaluation) 3. Can I adapt it (allocation) 4. How should I best use it (application)?

As stated earlier, the current framework cannot only be used as a systematic arrangement of crucial information quality criteria, or as the key questions of users, but also as a systematic problem lens. The four phases of the framework can be used to designate four dominant information quality problems, namely: information overload (information is not integrated), information misjudgment (information is not validated), information misinterpretation (information is not seen in context or contextualized), and information misuse (information is not made actionable).

The third and final element of the framework are the management principles.[3] As mentioned, they provide pragmatic help in implementing the framework and achieving the quality criteria contained in it. The principles are also placed *vertically* along the framework since they follow the same step-by-step logic as the four phases discussed above. Every principle relates to the criteria that are found in the same column as the principle.

The ***integration principle*** states that high-quality information has to be aggregated or compressed (made comprehensive, concise, convenient, and accessible) in order to give the information consumer an overview before details are presented. The application of this principle should make it easier to identify relevant and sound information quickly because information is no longer distributed in various sources and formats. Means of applying this principle are abstracts (content summaries), visualization (e.g. maps or matrices), categorization or taxonomies (e.g., hierarchical content trees), prioritization, or personalization (as in an intranet portal, where information sources are integrated based on a personal user profile). The main IQ-problem that is resolved through this principle is *information overload* or the fact that information is no longer acknowledged but ignored or stored away (and thus loses relevance or impact) when it is fragmented, prolix, inconvenient, or inaccessible.

The ***validation principle*** states that high-quality information has to be validated (in terms of correctness, consistency, timeliness, and security) in order to present only justified information to the information consumer and that the validation mechanisms that lie behind a piece of information be made visible. Means of applying this principle are consistency-checks on the information itself, comparisons with other sources ('second opinions'), an analysis of the primary source of the information (its reputation and competence), and a rating mechanism and rating scale that makes the degree of validation of the information visible (and gives information consumers the chance to provide feedback on the perceived quality of the information). The main IQ-problem that is resolved through this principle is *misjudgment* of (incorrect, inconsistent, late, or manipulated) information.

The ***context principle*** states that high-quality information is always presented with its context of origination and its context of use (where did it come from, why is it important and to whom is it important, how should it be used). Through this, the information should become clearer for the target group, because it can understand the information's background. The target group can also better assess whether the information holds true for the new context and if it is correct even under different circumstances. The context principle should also assure that the information is traceable, that is to say that its various origination steps can be traced back to the original source (this

[3] According to Merrill "a principle is a proven, enduring guideline for human behavior. It is a relationship that is always true under appropriate conditions regardless of program or practice. See: Merrill (1987).

criteria is also known as attributability). Finally, the context principle refers to the infrastructure in which information is stored. This infrastructure should not be neglected, but maintained to serve in future contexts. Means of applying this principle are adding meta-information[4] (such as author, reviewer, origination and expiration dates, target group, etc.), referring to similar pieces of information or to people who have used the information, and referring to prior information of the same kind. The main IQ-problem that is resolved through this principle is the *misinterpretation* (and hence misallocation) of information.

The ***activation principle*** states that high-quality information provides means of activating the information in the mind of the information consumer and thus renders it memorable and consequently easily applicable for later use. The activation principle strives for greater user acceptance by making the information as applicable and current as possible and by providing it in an interactive and fast manner. Specific means of applying this principle are repetitions of crucial information elements, mnemonics (cognitive shortcuts such as abbreviations), stories (vivid plots which make the information more memorable), metaphoric language and metaphoric visualizations, check questions for the information user, simulations or animations that make the information come alive and motivate the information consumer to actively explore and use it, etc. The main IQ-problem that is resolved through this principle is often referred to as *paralysis by analysis* or the fact that information is often not stimulating or motivating actions or decisions, but rather delaying them. A generic term for this information quality problem is information *misuse*.

These four principles are an integral part of the framework. In consequence, the framework can help the analyst to not only think about the crucial information characteristics (the individual information quality criteria) and their inherent conflicts or tradeoffs, but also about how these characteristics can actually be improved. The figure below summarizes how the four principles can be applied in knowledge-intensive processes.

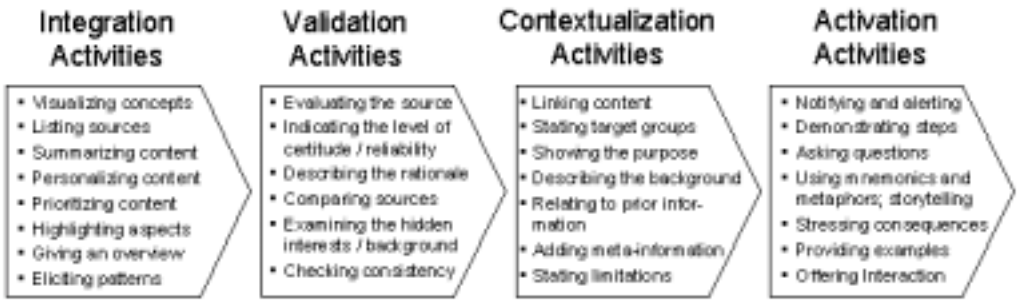

Figure 2: Ways of Implementing the Four Principles

Having described the main elements of the framework in overview, we can now turn to the specific information quality criteria and their relationships and see what they mean in specific knowledge-intensive processes.

[4] This improves the clarity of the information, its traceability, but also its maintainability for information administrators.

4. Discussion of the Included Criteria and their Relationships

In this section, the logic of the individual criteria and their relation to the category names are explained. The potential conflicts between the individual criteria are also discussed in this segment.

The logic for the criteria contained in the first level is the following*: Relevant information* is information that is adequate for the community that requires it. Adequate in this context signifies that the scope (or breadth) of the information is right (comprehensive enough), that the precision and level of detail is sufficient (accurate enough), that the clarity of argumentation is sufficient (comprehensible, interpretable, or clear enough) and that the information is easily applicable for the target community.

The criteria of the second level follow this rationale: *Sound information* is information that has certain intrinsic (product) characteristics which make it of high quality independently of the community that deals with the information. The information can said to be sound if it does not contain superfluous or non-related elements (conciseness), if it is internally consistent (does not contradict itself and uses the same format conventions)[5], if it does not contain errors or false elements (correctness), and if its it not outdated by more recent information (currency).

The criteria of the third level all relate to information as a process: We refer to the information delivery process as an *optimized information process* if the following conditions are met: the information creation, administration, and delivery is as convenient as possible for the three information stakeholders (author, administrator, user), e.g., there are no superfluous or tedious steps; the access to the information is secure in the sense that both the information and the user are protected against unauthorized manipulations; the information is continuously maintained (cleansed, updated); and the way the information is accessed or retrieved can be adapted to one's personal preferences through interactive elements.

The criteria of the fourth level all deal with infrastructure requirements: For an information *infrastructure* to be *reliable*, it is important that it is always accessible (no down-times, otherwise the information itself is not accessible), that it is secure (protected against unauthorized access or information manipulation), that it is maintainable (that is to say that the information can also be accessed easily in the future), and that it enables a fast interaction between the stored information and the users (or the authors or administrators). Infrastructure in this framework does only relate to the hardware and operating system software of information systems. An information infrastructure can be any kind of channel that allows for information to be shared, such as a paper archive, a library, a documentation center, an intranet, a war room like control center, a television studio etc.

As the framework in figure one indicates, all criteria of which the framework consists relate to at least one dimension of either *time, content* or *format*. The first level of information quality, the relevance criteria, contain subjective notions of information quality that mainly relate to the content of the information (although one could argue that applicability is also a question of

[5] For this understanding see also Kahn & Strong, 1998, which view soundness as an IQ-category for criteria such as free-of-error, concise representation, completeness, consistent representation.

format and timing). The second level, the soundness criteria, contain criteria of all three dimensions, since information has to be sound in terms of format, content, and time aspects. Again, one could argue that criteria such as consistency are not only related to a consistent format of the information, but also to a consistent content (that it is free of self-contradictions). While this is certainly true, it is often easier to detect inconsistencies in the format than in the content. For the other criteria in this group, the dimension seem obvious. The last two levels of the framework do not contain any content criteria since they relate only to the media quality and not the content quality of the information. The process and the infrastructure can directly influence the format (the presentation) of the information and its timely delivery. The process can only indirectly affect the content quality of information, for example through rating and validation workflows that double-check the information before it is published or distributed.

As far as possible *tradeoffs* between individual criteria are concerned, one can argue that the most critical criteria are comprehensiveness, timeliness, security, conciseness and accuracy, since they provide the most potential conflicts with other criteria. A tradeoff in this context refers to a possible goal conflict, that is to say when the increase in quality in one criteria leads to a decreasing value in another. One tradeoff that has been discussed in the information quality literature is the accuracy-timeliness tradeoff (see Ballou & Pazer, 1987), which consists of a choice for either accurate or timely information, since the very fact of a timely delivery often impedes the careful consideration of accuracy issues. A similar tradeoff exists between timeliness and correctness: the faster the information has to be delivered, the less time is available to fully check its correctness for a given context. The same may be said for consistency and timeliness: the faster information has to be delivered, the less time can be spent on improving its format and content consistency. Another tradeoff that may exist is between accuracy (in the sense of precision or level of detail) and conciseness: the more accurate information is, the less concise is its presentation. This tradeoff is similar to the one between conciseness and comprehensiveness: the greater the scope of the information, the more difficult is its presentation in a concise format. The quest for comprehensive information may lead to less clarity, since the increased scope decreases the clear distinction between central and peripheral information and thus makes information more difficult to comprehend. A high level of comprehensiveness also makes the infrastructure more difficult to maintain, since more information objects need to be updated (or removed) on the infrastructure. The tradeoffs related to the security criteria are threefold: First, there is a potential conflict between convenience and security, since many security measures lead to inconvenient authorization procedures for information producers, administrators, or consumers. Second, there is a clear conflict between providing quick access to an information infrastructure and keeping the infrastructure secure. A typical example of this tradeoff in the computer context is the number of times one has to enter passwords to access a certain information system. Third, there may be a tradeoff between the speed of an information infrastructure and its security, since security measures require additional resources which in turn may slow down the functioning of an information infrastructure.

Making these (and other) tradeoffs visible can help the designer of an information system or information product in his interaction with information consumers and authors, since it shows them the *constraints* under which one has to operate. In the context of consulting and market research, we have used the tradeoffs in the framework to show clients that is not possible to request a report that is delivered within two weeks (timeliness), contains no errors whatsoever (correctness and consistency), has a high level of accuracy and is very comprehensive and at the same time not more than fifteen pages (conciseness). Finally, the tradeoffs can also show

differences between various user groups of information: while one group may require information in a very comprehensive format, another information consumer group may require the same information in an extremely concise format (due to time constraints).

Table one summarizes the discussed categories or levels, the information quality criteria, as well as their antonyms.

Information Quality Levels	Information Quality Criteria	Opposites
Community Level (Relevance)	1. Comprehensiveness	Incompleteness
	2. Accuracy	Inaccuracy
	3. Clarity	Obscurity
	4. Applicability	Uselessness
Product Level (Soundness)	5. Conciseness	Prolixity
	6. Consistency	Inconsistency
	7. Correctness	Falsity
	8. Currency	Obsolescence
Process Level (Optimization)	9. Convenience	Inconvenience
	10. Timeliness	Lateness
	11. Traceability	Indeterminacy
	12. Interactivity	Rigidity
Infrastructure Level (Reliability)	13. Accessibility	Inaccessibility
	14. Security	Exposure
	15. Maintainability	Neglect
	16. Speed	Slowness

Table 2: Information Quality Criteria and their Opposites

As a result of this juxtaposition, we can define the antipode of quality information as follows:
Low quality information is incomplete, inaccurate, obscure, useless, prolix (or wordy), inconsistent, false, obsolete, delivered in an inconvenient, late, undeterminable and rigid way, on an infrastructure that is inaccessible, exposed to manipulation and other security risks, not maintainable, and slow.

Having described the framework and its logic, we can now turn to its application. For this purpose, we provide short case studies that show how two companies' efforts to improve information quality can be analyzed with the help of the information quality framework.

5. Case Studies on Information Quality Improvements: IHA-GfM Market Research and getAbstract

In section two of this paper, we stated that one way of validating[6] a conceptual framework such as this one consists of applying it in real world cases. This should reveal the usefulness of the framework as a systematic lens for information quality problems and solutions.

[6] The term validation in this context does not relate to the truth value of the framework, but rather to its applicability.

The first company case study deals with the knowledge-intensive process of **market research**. From 1998 to 2001, we have collaborated with a leading Swiss market research company called IHA-GfM (www.ihagfm.ch), a subsidiary of the international market research group GfK. The organization in Switzerland has over 300 employees and thus qualifies as a medium-sized company. The main products of this company are market research reports, market statistics, and market and (food-, non-food, near-food-, pharmaceutical, and media-) product analysis tools, such as media monitoring tools, category management tools or sales analyzers. In working together with the company, we have analyzed its *information process* (from the client briefing and first offer to survey construction, survey use, survey codification and analysis, to the final survey interpretation and client feedback) through workshops and interviews with the specialists of the company, as well as the final information *product* (the market reports) and its *infrastructure* (such as the company's client extranets). Since this company's main products are in fact information products, the quality of information is a crucial competitive component. Until our involvement with the company, however, the quality of information was mainly viewed as accuracy, consistency, correctness, timeliness and currency. But in 1999 (as market data became more and more of a commodity), the company understood that it could only enter a higher margin business if other quality criteria started to become relevant, such as applicability, convenience, conciseness, clarity, or maintainability. This, however, also meant a change in the qualification of its staff , who – up until then – were mostly trained in statistical analysis and not in information design and effective client communication (we will return to this important point when describing the specific improvement activities).

In six workshops with the company's project managers[7] we gathered and analyzed the challenges in the area of information quality that they saw (and realized because of their client satisfaction surveys) and we tried to find ways to improve the identified deficits. The reoccurring themes or challenges in terms of information quality were the following five issues:

1. The timeliness of the information that the company provided to its clients was seen as sometimes inadequate (still too many market research reports were not delivered on-time).
2. The accessibility and convenience of the information for clients was judged to be insufficient (it was argued that the new media were not yet fully used for the benefit of the clients).
3. The applicability of the information for clients was seen as a great improvement area (here it was argued that more added-value needed to be provided with the market data, such as benchmarks, comparisons, trend analyses, recommendations, consulting services etc.)
4. Finding the right scope or level of detail (for a market research report) in order not to overload clients was seen as a constant challenge.
5. Because of the relative autonomy of the various units of the company, the project managers considered it a major challenge to provide information in a consistent structure and layout.

Because of these problems, the company, decided to launch two projects to improve its competitiveness in this area. One project was launched to increase the quality of information and its sharing internally (labeled as *Knowledge Management*), the other project was launched to increase the value of information for clients (labeled as *Value 2000*).[8] The specific measures that

[7] The first such workshop was held in January of 1999 with about fifteen participants. The last was held in March 2001 with twelve participants.

[8] The knowledge management project was sponsored by the head of human resources, while the value 2000 project was directly sponsored by the CEO of the company.

were taken are summarized in the table below, where they are listed with the IQ-criteria that they affect most.

Information Quality Levels	Information Quality Criteria	Activities to improve the IQ-criteria
Community Level (Relevance)	Comprehensiveness	In order to increase the comprehensiveness of the information provided to clients, the company entered a joint-venture with Mediametrix to enlarge its scope to web user data.
	Accuracy	No specific measures were taken since the present label was seen as sufficient. The company had implemented a (certified) quality management system for its processes earlier.
	Clarity	Different layout templates were introduced that should make the information clearer and more easily interpretable. Presentations were introduced to make the information contained in market research reports clearer to the client.
	Applicability	In addition to just reporting the market data, reports now include interpretations of the data, further analysis and cross references, and recommendations for action. The reports are not only presented, but discussed with the client to determine its internal use. The project managers are trained in consulting tools in order to improve the impact of the gathered information.
Product Level (Soundness)	Conciseness	All reports now include executive summaries. Many reports have the statistical information in the appendix and focus on the key results.
	Consistency	All market reports that a client receives have a similar structure, layout and logic.
	Correctness	No specific measures were taken in regard to correctness.
	Currency	The use of adhoc on-line surveys was intensified to provide more up-to-date consumer data to clients.
Process Level (Optimization)	Convenience	The market report is not only delivered as a document, but also as PowerPoint slides, as a CD-ROM, and in the future also in an updated form on the client extranet.
	Timeliness	Pre-tests were intensified in order to eliminate time lags or possible errors early on in the process.
	Traceability	A knowledge map was developed and put on the company's intranet which makes it possible to trace back any tool or method to a tool owner or tool specialist.
	Interactivity	Clients are given more opportunities to provide input (via briefings, e-mails, presentations, telephone conferences etc.) during the information gathering process, before the report is finished.
Infrastructure Level (Reliability)	Accessibility	The client extranet can be accessed from any computer with internet access anytime of the day or night.
	Security	The client extranet is protected through a password and a hidden link. The intranet is protected through a firewall.
	Maintainability	Specialized key accountants are assigned to the client extranets where the market reports are updated or cleansed.
	Speed	No specific measures were taken since the available infrastructure was seen as fast enough.

Table 3 : Implemented Information Quality Improvement Activities at IHA-GfM

One key insight (regarding the information quality framework) from this implementation period at IHA-GfM emerged as more people were involved in the endeavor: As far as the momentum of

implementation was concerned, the great number of crucial information quality criteria was a disadvantage. Even a framework that would consists of only seven or eight core criteria would still make it difficult to focus the workforce on the necessary improvements and to drive the change process at a company. Because of this insight, we started to rely on the aforementioned four IQ-principles that are aimed at improving the IQ-criteria, rather than on the criteria themselves. Applying this insight to the situation at IHA-GfM, the following results were achieved:

- *Integration*: almost every major market report now contains a concise and systematic executive summary and an on-site presentation where the main consequences of a market study are presented in overview.
- *Validation*: almost every market research report contains an appendix that explains (in detail) how the data was gathered, analyzed and condensed.
- *Contextualization*: most market reports now refer to related reports or provide links to other available information or benchmarks that may render the information more meaningful by putting it into perspective and enabling comparisons.
- *Activation*: many market repots are now not just sent to the client or just presented in a management meeting, but actually discussed and analyzed in a workshop-like setting where clients can ask questions or probe deeper together with the consultants of the market research company. Most resources were spent on this aspect, especially in the area of training and tools.

The last issue from the above list points at a second key insight that emerged from the experience at IHA-GfM, namely that information quality improvements do not always require major information system changes. One of the key activities to improve the relevance dimension of information at IHA-GfM was **training** project managers on how to tailor and activate information for their clients through better presentations and better information visualization.

The second company case study deals with the knowledge-intensive process of **keeping up with relevant management know-how**. It was not gathered through participatory research like the case study on market research, but through a four hour interview with the CEO of the company, through being a client of the company for four months, and through a document and website analysis. The analyzed company is getAbstract.com Inc., the analyzed information products are book abstracts.

GetAbstract is a knowledge compression and rating company based in Lucerne, Switzerland with additional offices in Fort Lauderdale, Paris, Hamburg, Beijing and Hong Kong. It has about twenty full-time employees, and a network of 120 part-time collaborators. It is, according to the company, a “leading provider of compressed knowledge” mainly in the area of business books. GetAbstract states its mission as follows:

> To get the latest business trends and knowledge into the hands and heads of executives, managers and business students worldwide through concise Abstracts (summaries) of the newest and most important books on the market.

The company was founded in 1998 and incorporated in 1999. It has received major funding from two Swiss banks and various institutional and private investors. The company provides its (about

3000) private and more than sixty corporate clients and subscribers with concise, five-pages abstracts, available in four languages and across multiple platforms (as Adobe PDF files or PalmPilot files sent by e-mail, as audio abstracts, or in the form of a repository as a part of a company's intranet). The abstracts of current business books are written by a network of 120 professional writers and edited by three professional editors in Switzerland and the United States. These editors also rate each book according to its overall appeal, applicability, innovation, and style. The revenues of the company are generated by annual subscriptions of 299 US$ per client (this fee provides access to all available abstracts and to one new abstract every week automatically sent by e-mail) or through its corporate clients who integrate the service (with its library of over 1200 abstracted books) into their intranets.

Every book abstract has the same consistent structure: A thumbnail of the book next to the title and publishing information, a half-page of key take-aways, a rating on a scale of five to ten (books which receive a rating below five are not summarized), a one paragraph long review and recommendation, the actual abstract itself (about three pages long), and information about the author(s). The last two lines of every abstract provide a list of buzz-words used in the book.

On its web-page, GetAbstract provides full access to all abstracts which are categorized in so called knowledge channels, such as leadership, strategy, or technology. One can also search the database of summaries through a keyword search or browse a list of top downloads or new abstracts. Table three lists the main functionalities of the getAbstract service and shows which information quality criteria are influenced by it.

Information Quality Levels	Information Quality Criteria	Functions of GetAbstract.com
Community Level (Relevance)	Comprehensiveness	At the book level: the most important elements of the book are represented in the abstract. At the portal level: most general management bestsellers are summarized and about 8000 business books are screened per years.
	Accuracy	The accuracy of the provided information is determined by the editorial guidelines that are provided to the writers and by the quality of the writers who are mostly professional journalists.
	Clarity	Only professional writers are hired to write abstracts. Professional editors review the abstracts for clarity and style.
	Applicability	Regular feedback from abstract users is acknowledged and incorporated. The abstracts focus on take-outs and main new terms.
Product Level (Soundness)	Conciseness	Every book abstract is limited to five pages. Reviews are limited to one paragraph per book.
	Consistency	Every book abstract has the same structure: take-away, rating, author, buzz words. Authentic quotes from the book are included on the side.
	Correctness	The book abstracts are corrected by a team of editors.
	Currency	New book summaries are added every week, about five hundred new books are summarized every year.
Process Level (Optimization)	Convenience	The book abstract can be simply clicked at and is directly mailed to the inbox of the client where it can be read as a PDF-file or on the palm-pilot.
	Timeliness	New abstracts that fit the profile of a client are automatically sent out by e-mail.
	Traceability	Author and publisher information is always given. However, the reviewer's name is not disclosed.

	Interactivity	The getAbstract client can interactively edit his account and his interest profile. He can browse various book categories or do a key word search.
Infrastructure Level (Reliability)	Accessibility	All abstracts are accessible all of the time from any computer with Internet-access.
	Security	The getAbstract site is protected by a firewall. The client account is protected through a password.
	Maintainability	The site is continually updated and improved by a team of technical experts.
	Speed	The response time of the server seems immediate.

Table 4: Characteristics of GetAbstract Services that Increase Information Quality

Major competitors of getAbstract are Summaries.com (which offers the most inexpensive book summaries and delivers them over the Internet) Soundview's summary.com – whose thirty summaries per year are longer than getAbstracts but also available on tape and in print - and meansbusiness.com which has a significantly lower number of book abstracts in total, but also provides concise summaries of content across various books in its so called concept suites. A concept suite is a summary of book chapters from various books that deal with the same topic. The biggest entry barriers for this type of service are the legal obstacles (getting publishers to agree to the book abstraction) and finding qualified writers who can provide consistently instructive abstracts over a long period of time.

Again, we can summarize GetAbstract's major benefits or innovations with the four information quality principles:

- *Integration*: GetAbstract integrates on two levels: every book is reduced to five pages, and a great number of books is integrated in one web site.
- *Validation*: GetAbstract pre-selects and filters new business books and rates them according to a defined set of criteria.
- *Contextualization*: GetAbstract provides information on the author and his or her background, it states possible target groups of a book, and it will add references to similar books or to books others have found to be useful.
- *Activation*: GetAbstract stresses the key take-outs of every book and provides them through a push-mechanism to the reader (based on a user-defined profile).

The getAbstract case study has shown that high-quality information (in this case especially in the area of conciseness, comprehensiveness, and convenience) may be considered as a growing industry in its own right. The case study has also shown that a framework like the one presented in this paper can indicate possible future market niches, such as the one discovered by getAbstract to increase the conciseness of business knowledge in the forms of books.

6. Conclusion and Outlook

In this last section of the paper, we summarize the central findings in the first paragraph and provide an outlook to future research steps in the second one.

To increase the quality of information, it has to be targeted at a specific community in order to be relevant, it has to be managed as a product (with intrinsic qualities that we call soundness) and as a (continually optimized) content management process, and the platform on which information is provided has to be managed in order to be reliable. We refer to this reasoning as the four views or levels of information quality: the *community* level, the *product* level, the *process* level, and the *infrastructure* level. To manage the quality of information one has to pay attention to the author's, administrator's, and user's point of view and to their specific needs. One has to be aware of the potential conflicts between various information quality criteria and make these constraints visible for information consumers, authors, and administrators. The information user needs to be able to find and access the information (*identification* phase), he needs to be able to assess the information (*evaluation* phase), he has to be able to see the information in context and adapt it to his specific situation (*allocation* phase), and he has to be able to use the information for decision making or other applications (*application* phase). In order to assure that this is possible, certain management activities must take place at every one of the four described levels and in every one of the described phases. One can summarize these management or value-adding activities with the help of four principles: the identification-, validation-, contextualization-, and activation-principle. These principles make it easier to communicate and implement an information quality improvement program (versus having to explain a great number of criteria).

This paper has to be seen as an element of a larger research project. This project consists of finding an adequate application context for information quality, evaluating existing information quality frameworks, finding their improvement areas for the examined context, devising a modified framework for the application context, and illustrating and applying the framework with the help of documented case studies. As of now, the first four steps of this research project have been completed: 1. An application context has been defined, namely knowledge-intensive processes. 2. Existing information quality frameworks have been screened and evaluated according to specific meta-criteria. 3. Five deficits have been identified. 4. A new framework has been proposed. What remains to be done at this point, is to fully document the researched case studies and show how the framework's application can actually improve information quality in real-life situations. A specific challenge in this endeavor will be the compilation of a realistic set of indicators that make information quality improvements in knowledge-intensive processes measurable.

References

Agosta, L. (2000) Data Quality Methodologies and Technologies, Giga Information Group, Planning Assumption, August 30, 2000. URL: www.gigaweb.com [14.08.2001]

Astley, W.G.; Zammuto, R.F. (1992) Organization Science, Managers, and Language Games, in: *Organization Science*, 3: 4, pp. 443-460.

Ballou, D. P.; Pazer, H. L. (1987) Designing Information Systems to Optimize the Accuracy-Timeliness Tradeoff, in: *Information Systems Research*, 6(1), pp.509-521.

Brocks, M.: *Wissensarbeit im Consultingcontext: Die Rolle der Informationsqualität*, Diplomarbeit MBE-HSG, St. Gallen: Universität St. Gallen, 2000.

Eppler, M., Röpnack A., Seifried P.: Improving Knowledge Intensive Processes through an Enterprise Knowledge Medium, in Prasad, J. (Ed.): PROCEEDINGS OF THE 1999 ACM SIGCPR CONFERENCE Managing

Organizational Knowledge for Strategic Advantage: The Key Role of Information Technology and Personnel, 1999, pp. 222-230.

Eppler, M.: The Concept of Information Quality: An Interdisciplinary Evaluation of Recent Information Quality Frameworks, in: *Studies in Communication Sciences*, No. 1, 2001, pp. 167 -182.

Eisenhardt, K.M. (1989) Building Theories from Case Study Research, in: Academy of Management Review, Vol. 14, No. 4 (October), pp. 532-550.

Huang, K.-T.; Lee, Y.W.; Wang, R.Y. (1999) *Quality Information and Knowledge*. New Jersey: Prentice Hall.

Kahn, B. K.; Strong, D. M. (1998) Product and Service Performance Model for Information Quality: An Update, in: Chengalur-Smith, I.; Pipino, L. L. (1998) *Proceedings of the 1998 Conference on Information Quality*, Cambridge, MA: Massachusetts Institute of Technology.

Königer, P.; Reithmayer, W. (1998): *Management unstrukturierter Informationen*, Frankfurt.

Merrill, M. D. (1987) The New Component Design Theory: Instructional design for Courseware Authoring in: *Instructional Science*, 16, pp. 19-34.

Porter, M.E. (1991) Towards A Dynamic Theory of Strategy, in: *Strategic Management Journal*, Vol. 12, pp. 954-117.

Roehl, H. (2000) *Instrumente der Wissensorganisation*, Wiesbaden: Gabler.

Schmid, B. (2000) *Knowledge Media*, St. Gallen: NetAcademy Press.

Strong, D. M., Lee, Y. W., Wang, R. Y. (1997) 10 Potholes in the Road to Information Quality, in: *Computer IEEE*, 30(8), pp. 38-46.