FROM DQ TO EQ: UNDERSTANDING DATA QUALITY IN THE CONTEXT OF E-BUSINESS SYSTEMS

A fix for irrelevant information, cognitive overhead, and disorientation—common gremlins endured by every e-business system user.

The bust of the “e” hype in 2000 may have taken the shine from the e-everything paradigm, but not before effecting a fundamental transformation in the way firms do business. The e-bubble has burst, but the e-business paradigm survives and continues to grow on a nearly ubiquitous Internet. The early innovators continue refining their existing e-applications and implementing creative new ones, with an eye on creating competitive advantage for their firms. The laggards have examined the business case and joined the e-bandwagon by launching new e-business systems to maintain competitive parity.
Despite this growth, problems in the use of e-business systems remain. For example, users of B2B systems report a number of serious problems during transactions. The problems include difficulties in locating the required information, completing ongoing transactions, finding timely and accurate information, and finding adequate electronic service functions to complete online transactions.

These problems can be analyzed and understood using the lens of data quality\(^1\) [4, 5, 12]. Data quality (DQ) is a pervasive concept and a key antecedent of information systems success [1]. However, traditional approaches to DQ fall short in the context of e-business systems as they do not adequately encompass and address aspects that are unique to these systems. E-business systems, enabled by the Internet, Web, and hypermedia technologies, are highly dynamic and interactive in nature, utilize rich hypermedia mechanisms in user interfaces for information presentation, and provide a tremendous amount of control over temporal aspects of information delivery to end users.

However, due to their evolution from traditional data processing environments, the focus of current DQ approaches has primarily involved the content of the information (such as its relevance, accuracy, and completeness) and not the interface-related aspects of information presentation and end-user delivery.

Information presentation refers to the form in which a snapshot of information content is presented to users through a Web interface, while information delivery refers to the temporal aspects of providing information content to users through the interface. Current DQ approaches deal only with structured types of conventional data, and not with unstructured types of hypermedia data objects and hyperlinks, which are an integral part of any modern-day Web-based e-business system.

It has been noted that data quality should always be considered in terms of “use-based data quality” [8]. In other words, the vehicle for information presentation and delivery that enables information use should not be separated from the information itself (information content) when thinking about DQ. This is particularly important in e-business systems because ignoring information presentation and delivery aspects will create problems in terms of irrelevant information, disori-
Hence, any discussion about data quality in the context of e-business systems should incorporate not only the quality aspects pertaining to information content, but should also include the aspects of the form in which information is presented, as well as the aspects of time pertaining to the delivery of information content to end users (particularly with respect to unstructured types of hypermedia data objects and hyperlinks).

Recent work in the DQ area has begun to address the aspects of information presentation and delivery by focusing on the process of converting data into information [5], based on the notions of service quality [10]. Such research is a move in the right direction, but still neglects the core issues of information delivery and presentation in terms of the usability challenges of disorientation, irrelevant information, and cognitive overhead in e-business systems. We have been conducting research on these core issues. The DQ framework presented here incorporates the various DQ dimensions in the context of e-business systems, and illustrates the unique mechanisms of information presentation and delivery enabled by Web and hypermedia technologies.

The framework also addresses the usability problems faced by users of e-business systems. This framework, termed the E-Quality (EQ) framework, due to its focus on e-business systems, has been developed by integrating the data quality literature [1, 4, 5, 8, 12] with literature on design and usability of hypermedia and Web-based systems [2, 3, 7, 9, 11]. Here, we discuss the Web usability problems, develop the EQ framework, and reconcile its three dimensions with the major and well-known DQ models—in the process showing the shortcomings of the current DQ approaches in the context of e-business systems.

**Usability Problems in Hypermedia**

The hypermedia literature addresses three main problems users encounter while browsing information through hypermedia systems: irrelevant information, cognitive overhead, and disorientation [7]. The irrelevant information problem pertains to the information content itself. While traversing certain links, users reach Web pages (hypermedia documents) that do not contain the required information. In such cases, users generally tend to return to previous pages to continue their search. Users could make better judgments about what links to follow if information was readily available to them about the

---

3Due to limitations on the number of references that can be included, only key references for the two bodies of literature are included here. A larger list of references is available from the authors upon request.

4We use the terms “Web page” and “hypermedia document” interchangeably, as a Web page is essentially a hypermedia document.
Table 1. The EQ framework: Information quality in the context of hypermedia-based e-business systems.

<table>
<thead>
<tr>
<th>EQ Dimension</th>
<th>Meaning</th>
<th>EQ Constructs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>This EQ dimension deals with the intrinsic information content issues that are geared toward providing users with accurate, relevant, and complete information, thereby addressing primarily the problem of irrelevant information in e-business systems.</td>
<td>Information Accuracy: Freedom from mistakes in the information content and hyperlinks provided within Web pages. Information Relevance: Pertinence to users’ interests of the information content and hyperlinks provided within Web pages. Information Completeness: Availability as needed of the information content and hyperlinks within the Web pages for users to complete specific tasks in an effective manner.</td>
</tr>
<tr>
<td>Form</td>
<td>This EQ dimension deals with information presentation issues that are geared toward enhancing users’ cognition, thereby primarily addressing the problem of cognitive overload.</td>
<td>Interface Structural Quality: Primarily comprises interface consistency and structural awareness. Interface consistency implies consistency in the structural arrangement and style of information content and hyperlinks within an e-business application. Structural awareness implies that the interface makes the user aware of the larger structure of the information content in the Web pages in the e-business application. Information Packaging Quality: Refers to how effectively a variety of information in various media types is packaged within the Web interface for presentation to end users. Includes the amount and cohesiveness of information content and hyperlinks presented within the interface, and semantic relationships among them. Information Accessibility: Refers to the ease and efficiency with which a user can navigate within an e-business application to access and retrieve desired information.</td>
</tr>
<tr>
<td>Time</td>
<td>This EQ dimension deals with information delivery issues that are geared toward providing users’ better control over temporal aspects of their actions thereby providing them with a sense of temporal orientation and, thus, addressing primarily the problem of disorientation in e-business systems. This dimension also deals with providing temporally accurate, or current, information to users of e-business systems.</td>
<td>History Maintenance Quality: Refers to the flexibility and comprehensiveness of features that an e-business application provides to its users for specifying and maintaining history on user actions and data and system states of the application. Information Delivery Quality: Refers to the flexibility and comprehensiveness of features that an e-business application provides to its users for specifying and controlling the temporal relationships among the various hypermedia components for effective delivery of integrated hypermedia information to users. Information Currency: Refers to the temporal accuracy of information content and links on Web pages. This construct also captures the notion of age of information on Web pages which can be measured by the amount of time that has passed since the information on the pages was last updated.</td>
</tr>
</tbody>
</table>

Information presentation only covers the issues involved in presenting a snapshot of information to a user on a Web interface. Information delivery includes the temporal issues involved in delivery of information to a user on a Web interface.

documents/content would reach by following those links. This would help solve the problem of irrelevant information, illustrated through the example in Figure 1. In this example, a user seeking tax forms in order to get an IRS refund initially browses the URL www.irs.com instead of the official IRS Web site at www.irs.gov. The user clicks the hyperlink titled “Individual Tax Forms” (see Figure 1a), and reaches a Web page that provides several links pertaining to taxation (see Figure 1b), but individual tax forms are not directly available on this particular Web page. The outcome in this scenario departs from the user’s original intentions.

The second problem users of e-business systems face is that of cognitive overload. This problem stems from the effort and concentration required to manage and/or maintain several tasks or navigation trails at once. Figure 2a shows an example of this problem. The Web page shown in this figure has too much content and too many links on a single hypermedia document, precluding users from readily finding the information they need due to cognitive overload.

The third problem is disorientation: the tendency to lose one’s sense of location and direction in a set of non-linear documents. As a result of traversing a large number of links, users can end up feeling confused. Users “lost in hyperspace” generally turn to their browser’s back key in order to return to familiar territory. Providing a context of where they are and how they arrived at a particular document might decrease their disorientation. Figure 2b illustrates an example of the disorientation problem in an e-business system. In this particular instance, although a user traversed from page 1 to page 4 (as marked in the figure) using appropriate links in each page, the user is currently “lost in hyperspace” and has lost context in the current traversal session.

EQ for E-Business Systems

E-business systems utilize a variety of Internet, Web, and hypermedia technologies and support both hyper-structured navigation and multimedia presentation of information [3]. These systems provide users with a rich set of alternate traversal routes and with the capability to navigate freely through a variety of information using links provided on Web pages. However, this flexibility with regard to storing and retrieving information requires users to make decisions continually and forces them to constantly use higher-level intellectual processes. Inadequacies in either the content or the presentation/delivery of information can, therefore, lead to one or more of the three commonly encountered hypermedia problems discussed earlier, and in the process stymie the effective usage of information for individual and/or organizational benefit.

Thus, the hallmark of any DQ framework geared toward e-business systems is its ability to explicitly address these hypermedia usability problems by incorporating quality constructs that can help designers and users overcome, or at least minimize, these problems. The EQ framework acknowledges this need and incorporates three quality dimensions...
and nine quality constructs that individually and jointly address the three usability problems.

The three quality dimensions in the EQ framework, shown in Table 1, deal with information content, form, and time.

**Content.** The content dimension in the EQ framework deals with intrinsic information content aspects. It is geared toward providing users with accurate, relevant, and complete information, thereby addressing primarily the problem of irrelevant information in e-business systems. The content dimension consists of three quality constructs including information accuracy, information relevance, and information completeness, as it is commonly agreed that retrieved information should be accurate, relevant, and complete in order to add value to the task for which it is retrieved [4]. Information accuracy essentially means that the information content and hyperlinks provided within Web pages are free from mistakes. Information relevance assures that the information content and hyperlinks provided within Web pages are pertinent to users’ needs and interests. Information completeness means that the information content and hyperlinks provided within hypermedia documents are available as needed for users to complete specific tasks in an effective manner.

For example, when logging into a B2C e-commerce Web site, users may be allowed to use any one of a variety of identification mechanisms, such as a user-ID, an email address, a customer number, an order number, a telephone number, and a credit card number to view or update their relevant information (such as order/shipment information or account profile). This is exemplified in Dell’s e-commerce Web site (support.dell.com/dellcare/orderstatus/orderstatus.aspx?c=us&l=en&cs=gen&-ck=mn), which permits users to use either an email address or an order/invoice number.

**Form.** The form dimension in the EQ framework is concerned with information presentation issues in terms of interface structure, information packaging, and information accessibility—all geared toward enhancing users’ cognition, and thereby addressing the hypermedia problem of cognitive overhead. Because users’ capacity to cope with complexity and high volumes of information is rather limited [6], hypermedia systems should provide functionality that lessens viewers’ cognitive efforts toward comprehension [11]. Comprehension is the ability to construct a mental model that represents objects and semantic relations described in a document.

Information quality of an e-business system pertaining to the form dimension can be viewed through the constructs of interface structural quality, information packaging quality, and information accessibility. Interface structural quality is determined by interface consistency and structural awareness. Interface consistency means the structural arrangement and style of information content and hyperlinks follow a certain standard, and are consistent throughout the application interface. Structural awareness means the interface makes the user aware of the larger structure of the hypermedia content, including the major topics and relationships among them within the overall e-business application. It has been suggested in the literature that high interface structural quality will further reduce users’ efforts toward interface adjustment [11]. Information packaging quality refers to how effectively a variety of information in various media types is packaged within hypermedia interfaces for presentation to end users. This construct, therefore, deals with the notion of the amount and cohesiveness of information content and hyperlinks presented within hypermedia interfaces, and semantic relationships among them. Information accessibility refers to the ease and efficiency with which a user can navigate within a hypermedia application to access and retrieve desired information.

**Time.** The time dimension in the EQ framework deals with information delivery issues geared toward giving users better control over temporal aspects of their actions. This dimension provides them with a sense of temporal orientation and addresses the critical hypermedia problem of disorientation in e-business systems. Time is a very important information quality dimension in the context of e-business systems because users must continually have a sense of temporal history to keep track of their location and direction as they navigate from Web page to Web page within an e-business application. Also, because information is frequently time-sensitive, this dimension also deals with providing temporally accurate and current information to users of hypermedia systems.

Information quality in terms of the time dimension relates to constructs of history maintenance quality, information delivery quality, and information currency. History maintenance quality refers to the flexibility and comprehensiveness of features that an e-business application provides to its users for specifying and maintaining history of their actions, and of the data and system states within the application. One of the key aspects of a history of user actions is the

\[\text{As mentioned earlier, the term presentation of information only refers to a snapshot of information that is provided to a user through a Web interface.}\]
navigation history, as it can assist users in identifying their location in hyperspace. Information delivery quality refers to the flexibility and comprehensiveness of features that an application provides to its users for specifying and controlling the temporal relationships among the various hypermedia components for effective delivery of integrated hypermedia information. Order of presentation of hypermedia components, a key aspect of information delivery quality, includes the notion of synchronization, which provides a collection mechanism for related components to be presented together. Synchronization can also be used for integration of tasks and information. Hence, a well-defined presentation sequence can considerably reduce the disorientation problem in e-business systems. Finally, information currency refers to the temporal accuracy of information content and links on Web pages. This construct also captures the notion of age of information on Web pages, which can be measured by the amount of time that has passed since the information on the pages was last updated.

The Integration of EQ Dimensions
The three dimensions of the EQ framework may at first appear to be self-contained. However, as shown in Figure 3, the form of information presentation and the temporal aspects of information delivery are intimately related to the information content. All three aspects of information quality highlighted in the EQ model must be addressed in an integrated manner or the overall quality of an e-business application will suffer. While each of the three Web usability problems of irrelevant information, cognitive overhead, and disorientation are addressed primarily by one of the three EQ dimensions, each problem can only be fully addressed by increasing quality across the three EQ dimensions simultaneously.

For example, the problem of cognitive overhead can only be addressed by effectively packaging and distributing the given hypermedia content in multiple interconnected Web pages that are delivered in a particular sequence to a user to reduce information overload. In other words, a joint optimization on the EQ dimensions of form (because it addresses the issue of information packaging in the presentation) and time (because it addresses the issue of order of information delivery) is required to overcome the problem of cognitive overhead.

We have mapped Wang et al.'s DQ [4, 5, 12] and DeLone and McLean’s DQ [1] constructs with the EQ constructs in Table 2. While these two DQ models provide a number of important constructs, they do not provide adequate coverage for a number of important DQ dimensions in the context of e-business systems. As can be seen from the table, both models are weak in terms of the content EQ dimension, as the constructs contained in these two models refer to information content only in a traditional sense and not to information content as hypermedia objects and concepts. Both of these models also do not contain constructs for structural awareness in the form dimension, an important construct from the perspective of cognition enhancement. These two
Table 2. The EQ framework and traditional DQ frameworks.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevant Constructs</td>
<td>Assessment</td>
<td>Relevant Constructs</td>
</tr>
<tr>
<td>Content Dimension</td>
<td></td>
<td>Assessment</td>
</tr>
<tr>
<td>Information Accuracy</td>
<td>C X H</td>
<td>C X H</td>
</tr>
<tr>
<td>Information Relevance</td>
<td>C X H</td>
<td>C X H</td>
</tr>
<tr>
<td>Information Completeness</td>
<td>C X H</td>
<td>C X H</td>
</tr>
<tr>
<td>Form Dimension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interface Structural Quality</td>
<td>Representational DQ category: consistent representation</td>
<td>C X H</td>
</tr>
<tr>
<td>Information Packaging Quality</td>
<td>Contextual DQ category: appropriateness of data, conciseness of understanding, conciseness of representation</td>
<td>C X H</td>
</tr>
<tr>
<td>Information Accessibility</td>
<td>Accessibility DQ category: accessibility, access security</td>
<td>C X H</td>
</tr>
<tr>
<td>Time Dimension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History Maintenance Quality</td>
<td>None</td>
<td>X C X H</td>
</tr>
<tr>
<td>Information Delivery Quality</td>
<td>None</td>
<td>X C X H</td>
</tr>
<tr>
<td>Information Currency</td>
<td>Contextual DQ category: currency, timeliness</td>
<td>C X H</td>
</tr>
</tbody>
</table>

Legend:
- C - Strong Support for conventional (structured) data
- C - Moderate Support for conventional (structured) data
- C - Weak Support for conventional (structured) data
- X - No Support for conventional (structured) data
- C - Strong Support for hypermedia (unstructured) objects and hyperlinks
- C - Moderate Support for hypermedia (unstructured) objects and hyperlinks
- C - Weak Support for hypermedia (unstructured) objects and hyperlinks
- X - No Support for hypermedia (unstructured) objects and hyperlinks

models are also quite inadequate in the time EQ dimension, as they do not provide any constructs for history maintenance quality and information delivery quality.

**CONCLUSION**

The EQ framework presented here provides a comprehensive framework for developing and evaluating e-business systems from both the product (information content) and service (information presentation and delivery) aspects of information quality. The three dimensions contained in the EQ framework individually and collectively address the three Web usability problems. The three dimensions also adequately capture the major concepts and constructs used in Web and hypermedia systems modeling, such as information elicitation and structuring, synchronization, human-computer interaction, information presentation, and navigation. The EQ framework can be used as a checklist by e-business system developers to help them build “quality by design” into their systems while users can use this framework to help them comprehensively evaluate e-business systems from an information quality perspective.

**REFERENCES**


Yong Jin Kim (ykim@binghamton.edu) is an assistant professor in the School of Management at the State University of New York at Binghamton.
Rajeev Kishore (kishore@buffalo.edu) is an assistant professor in the School of Management at the State University of New York at Buffalo.
G. Lawrence Sanders (mgtsand@acsu.buffalo.edu) is currently the chair of the Department of Management Science and Systems at the State University of New York at Buffalo.