INFORMATION QUALITY AND THE RAISING DEMANDS OF REGULATORS: REENGINEERING THE CUSTOMER INVESTIGATION PROCESS AT CREDIT SUISSE

Completed Paper

Mario Klesse, Clemens Herrmann
University of St. Gallen, Switzerland
{Mario.Klesse|Clemens.Herrmann}@unisg.ch

Pius Brändli, Thomas Mügeli, Daniel Maier
Credit Suisse, Switzerland
{Pius.Braendli|Thomas.Muegeli|Daniel.Maier}@credit-suisse.com

Abstract: Customer investigations in the banking industry are carried out in connection with prosecutions, administration of estates or other legal actions. The Investigation & Inquiries department of Credit Suisse has to handle approximately 5000 customer investigations per year. So far, the investigation process was very complex time consuming and costly: to be on the safe-side, several redundant query-processes needed to be performed to achieve bullet-proof results. In addition, new regulatory requirements led to a massive increase of investigations to be performed in the last few years. This case study describes how these requirements could be met by redesigning the process and building an investigation application that automates most parts of the process. These two measures significantly improved the customer investigation process, resulting in considerable cost and time savings for Credit Suisse. The case is structured using a theory based information quality framework proposed by Liu and Chi (2002).

Keywords: Data Quality, Customer Investigation, Address Management, Data Warehouse, Fuzzy Search

INTRODUCTION

Customer investigations are common in the banking industry and are carried out in connection with prosecutions, administration of estates or other legal actions. This paper focuses on information quality in the customer investigation process (CIP) at Credit Suisse, a leading provider of comprehensive financial services Worldwide. In 2003 Credit Suisse (not including Credit Suisse First Boston, Bank Leu and Neue Aargauer Bank) had over 2.6 million clients and 20,000 employees worldwide and had CHF 740 billion assets under management. Credit Suisse has to handle about 5000 individual customer investigation requests per year plus a varying but steadily increasing number of special embargo requests, such as the various terror related search lists. The primary objective of the CIP is to recover all business relationships the bank has or has had with a certain customer or with individuals related to a certain customer. In the last years the importance of customer investigations in the financial industry increased due to the following external developments:

- *Risk management*: Banks are enforced by regulatory authorities and market developments to implement improved procedures for managing reputational, operational and legal risks. Reputational risk plays a major role in the banking industry, since the nature of its business requires maintaining the confidence of all stakeholders. Operational risk can be defined as the danger of direct or indirect losses caused by the potential failure or inadequacy of internal processes. Legal risk is the risk that lawsuits, adverse judgments or unenforceable contracts negatively affect the operations or condition of a bank [4]. By implementing an effective and efficient CIP, banks are engaging in due diligence in

identifying customers and understanding their business. This can reduce a bank's reputational and legal risks. The quality of the CIP is also affecting the operational risk of a bank, which in turn is also important in the context of the New Basel Capital Accord (also known as Basel II) [5]. By decreasing the operational risk banks are able to lower the required capital buffers for risk compensation.

- *Combat terrorism*: Since September 11th, 2001 many countries have issued anti terror bills (e.g. USA Patriot Act 2001 [1]) which affect the banking industry. Having a terrorist as customer increases legal and reputational risk besides being unethical. All members of the Wolfsberg Group – an association of twelve global banks aiming to develop financial services industry standards – have committed to cooperate with governance in combating terrorism, to seize measures for identifying suspected terrorists quickly, and to support the Financial Action Task Force (FATF) Special Recommendations on Terrorist Financing [2, 3]. The ongoing fight against terrorism leads to a continuously growing amount of customer investigation inquiries and a demand for monitoring transactions in order to detect those appearing suspicious**.** In addition, the increasing number of blacklists, such as terrorist lists, the Office of Foreign Assets Control list [7], the FBI Control lists etc., have to be checked continuously against all customer information of a bank to both comply with regulatory requirements and to avoid reputational or legal risks. Furthermore, clients need to be checked against other sanction lists, e.g. politically exposed persons [6], in order to be able to apply additional due diligence procedures. All these efforts need to be performed in addition to existing standard banking operations and must be executed in a timely manner.
- *Anti money laundering*: Anti money laundering laws and policies, which were adapted after September 11th, 2001, also lead to an ever increasing amount of customer investigation inquiries and a demand for high quality investigations analyzing relations between individuals and organizations.

These developments force banks to ensure a high quality CIP which enables preventive risk management and fast response to the increasing number of legal inquiries. Therefore, the objective of this paper is to present how Credit Suisse improved and streamlined the process by means of organizational and technical measures. In order to give a structured and sound case description, an appropriate information quality framework based on a literature review is selected in the next section. Upon this theoretical foundation the original CIP, its problems and the major challenges for process improvement are presented. In the subsequent section the workflow of the revised and partly automated CIP is illustrated. The architecture and functionality of the supporting information system and the achieved information quality improvements are described in detail. The paper concludes with a summary of the case.

INFORMATION QUALITY FRAMEWORK

In literature several definitions of the terms 'data quality' and 'information quality' can be found and often they are used synonymously. Since a clear distinction between data and information seems to be impossible for the purpose of this paper, both terms are used interchangeably. A standard information quality definition does not exist yet but information quality is generally regarded as a multidimensional and a hierarchical concept [8-10]. Three different approaches for deriving and specifying quality dimensions can be distinguished [8, 11]. The *intuitive approach* proposes information quality attributes based on personal experience or on subjective insights about what dimensions or attributes are most relevant. The *empirical approach* quantitatively captures the data consumers' point of view about what quality dimensions are important according to their tasks. The *theoretical approach* builds upon an established theory and proposes quality dimensions corresponding to this theory. Table 1 gives an overview of selected publications for each approach.

The major drawback of the intuitive and empirical approaches is the strong influence of the researcher's personal experience on the selection and deduction of information quality attributes and dimensions. The lack of a theoretical foundation results in missing justifications and understandings on why and how certain information quality classifications and definitions are proposed. Therefore, this paper adopts a

theoretical approach for defining information quality. The information quality framework presented by Liu and Chi [11] seems to be well suited for the purpose of a structured case description because their generic approach can easily be adapted to the characteristics of the case and the proposed quality stages correspond to the different steps of the customer investigation process.

AUTHOR(S)	BRIEF DESCRIPTION
Intuitive Approaches	
Wang, Reddy and Kon 1995 [9]	A methodology for tagging data at the cell level of database tables with quality indicators is developed so that end-users can assess the data quality for their intended application. The authors present a hierarchical data quality framework. They distinguish the four dimensions accessibility, interpretability, usefulness, and believability. Each dimension consists of further characteristics.
Miller 1996 [12]	The author proposes ten attributes of information quality focusing on how they can be used as benchmarks to improve the effectiveness of information systems and to develop information quality strategies.
Redman 1996 [13]	The quality dimensions are associated with a conceptual view, data values, and data representation and recording. A total of 27 dimensions are described and specified in detail.
English 1999 [14]	Data quality is characterized by four dimensions: data definition quality, information architecture quality, data content quality, and data presentation quality. For each category a list of detailed quality attributes is set up.
Empirical Approaches	
Wang and Strong 1996 [15]	Based on a multi-staged empirical survey (355 filled-in questionnaires in the main analysis step) a hierarchical framework with 16 data quality dimensions is presented. Accuracy and correctness were identified as the quality attributes most important to end-users. The final results of this study comprise four categories (Intrinsic, Contextual, Representational, Accessibility), each containing several detailed quality attributes.
Helfert, Zellner and Sousa 2002 [16]	The authors conducted a survey among 25 German and Swiss organizations covering data quality issues of data warehouse systems. The results of the survey comprise a list of data quality attributes according to their importance for practice. The most highly ranked attributes are consistency, correctness, and completeness.
Theoretical Approaches	
Ballou and Pazer 1985 [17]	On the basis of a data flow diagram for multi-system environments, the impact of data and process quality can be assessed. The model can address the four quality dimensions accuracy, timeliness, completeness and consistency.
Te'eni 1993 [18]	Based on the data production process and the person-environment fit theory the different data quality requirements of data producers and data users are discussed. The resulting data production model identifies reasons for data quality problems on human and organizational levels.
Wand and Wang 1996 [19]	Ontological mappings are used to determine the conformance between real-world states and representations in information systems. Data are considered incomplete if a real-world state has no corresponding information system state. Conversely, data are regarded as meaningless if an information system state has no real-world representation. Data are deemed ambiguous if a single information system state can be mapped to several real-world states.
Liu and Chi 2002 [11]	Based on the four data evolution activities collection, organization, presentation, and application corresponding data quality stages are defined. Measurement attributes for each data quality stage are derived from typical data quality root causes during data evolution.

Table 1: Literature overview of selected information quality approaches

Figure 1 depicts the so called data evolution lifecycle which is used by Liu and Chi as a theoretical basis to derive four data quality stages. The lifecycle characterizes the typical sequence of data evolution stages consisting of data collection, organization, presentation, and application. First of all, data are captured, e.g. by observing or measuring real world processes or objects. Then data are organized according to certain structures, e.g. in file-based data stores or more sophisticated databases. After that, data are processed and presented. Finally, data are utilized for a certain application purpose which in turn can trigger further data capturing [11]. At every stage of the lifecycle specific techniques, methods, models or other approaches are applied which influence the evolution of data. For example data are organized in different ways depending on the modeling paradigm (e.g. relational or object-oriented) being used. Depending on the applied techniques, methods or models during the lifecycle, different errors may occur and therefore different quality dimensions and attributes have to be measured. Accordingly, Liu and Chi introduce the concept of evolutional data quality consisting of the four quality stages collection quality, organization quality, presentation quality, and application quality. The quality of data at earlier stages of the lifecycle contributes to that at later stages, i.e. the quality measure is accumulative [11]. Liu and Chi exemplify their evolutionary data quality approach by presenting typical root causes of poor data quality

and deriving specific measurement attributes and models at each data quality stage.

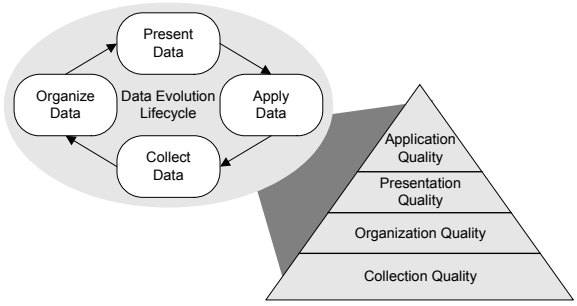

Figure 1: Data evolution lifecycle and corresponding quality stages (cf. [11])

For the purpose of this paper the concept of evolutional data quality is used to point out the information quality issues as well as the improvements in the customer investigation process of Credit Suisse. The process consists of two major phases. In the request phase a new customer inquiry is placed whereas in the reply phase the results to a specific inquiry are produced (see Figure 2). Thus, the process consists of two connected data evolution lifecycles. First, data specifying the customer inquiry are processed. The application of the customer inquiry data triggers the second lifecycle which handles the inquiry results. This two-phase-structuring of the process is used throughout the paper.

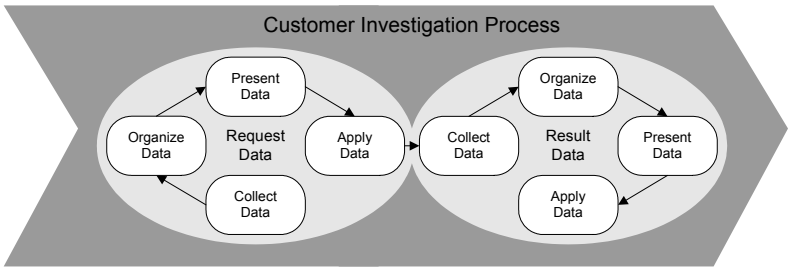

Figure 2: Data-oriented perspective on the customer investigation process

THE ORIGINAL CUSTOMER INVESTIGATION PROCESS

The following section describes the original CIP as it was carried out within Credit Suisse in the past and analyzes the information quality issues according to the framework depicted above.

Activities and Workflow of the Original Customer Investigation Process

The CIP was initiated by an inquirer (e.g. external/government or internal person who requested information) who first had to identify one or more appropriate receiver or consignees for that specific inquiry. Having identified those, the inquiry was sent to all relevant consignees. This starts the investigation process within Credit Suisse. The original CIP within Credit Suisse consisted of the following activities (cf. Figure 3): Each receiver of the inquiry accepted the request (1) and started to identify departments who might have relevant information concerning the inquiry (2). These departments are owners of applications or information archives that store information about customers. For all application owners identified, a fax request was prepared and distributed to them (3). The application owning department then had to prepare queries based on the received request (4) and enter the query into the corresponding system (database) manually in order to extract the relevant information (5). On average, about 15 applications had to be queried by their system owners and about 40 faxes had to be sent for coordination. After running the query, results were collected by each application owner and sent back to the inquiry receiver (6). The inquiry receiver then tried to consolidate the results received from the

application owners (7). If further information was needed or the data quality turned out to be inadequate the process started again with step 2, e.g. further departments got involved and requests were repeated. If the collected information was considered sufficient, results were summarized in a dossier by the inquiry receivers (8). Finally, the dossier was sent to the inquirer (9).
The original process was not only quite complex and time-consuming for Credit Suisse (steps 1-9) but also required considerable coordination efforts for the inquirer. The inquirer had to find an appropriate consignee for his request (B). Therefore he could trigger multiple redundant investigation processes simultaneously (C), if more than one receiver was identified. If multiple inquiries were sent by the inquirer, the dossiers received needed to be consolidated (D).

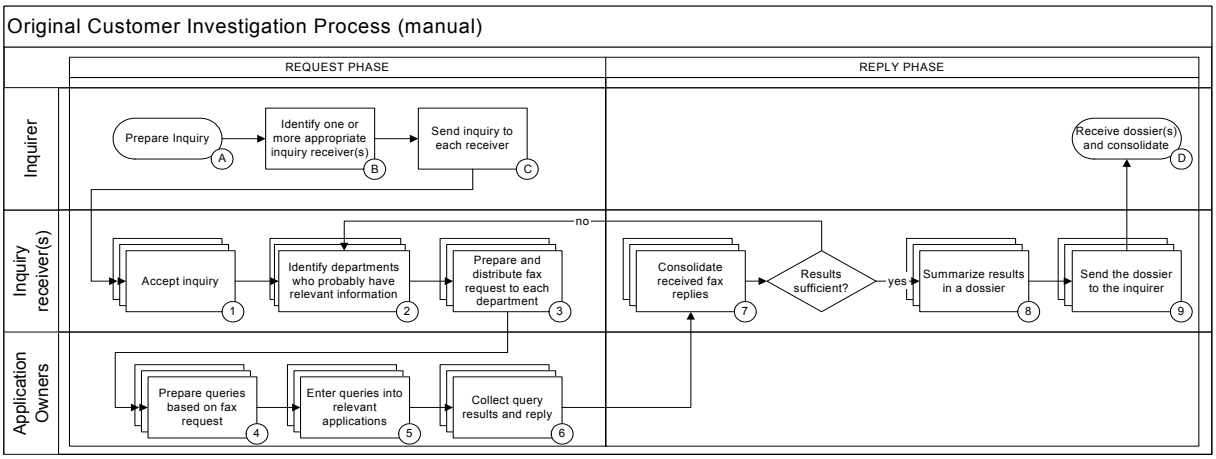

Figure 3: Original customer investigation process at Credit Suisse

Analysis of Information Quality Issues in the Customer Investigation Process

To analyze the critical path in the original CIP we decompose the process into two main parts: First, the distributing and processing of the inquiry itself ("request phase", step 1-6) and second, the processing of the inquiry results by the inquiry receiver ("reply phase", step 7-9). We distinguish the two phases because in each phase a unidirectional data flow occurs: In the request phase information flows top down, from inquirer to application owners and in the reply phase it is collected and consolidated bottom up, from application owners to the inquirer. This allows for analyzing the information quality issues in the process along the data evolution lifecycle as described in the section above. We concentrate on the process steps 1-9, because these actions are carried out within Credit Suisse. The part of the process carried out by the inquirer (A-D) is only analyzed concerning utilization quality, since the inquirer may be an external entity whose information usage process can be unknown.
Furthermore we assign the actions of the process to the elements of the information evolution lifecycle according to Figure 1. Within the request phase, step 1 can be considered as information *collection* by the inquiry receiver, steps 2-5 may be assigned to information *organization* by the inquiry receivers and the application owners and step 6 to information *presentation* from the perspective of the application owners. Within the reply phase, information is *collected* from the application owners by the inquiry receivers (step 7), (re-)*organized* (steps 7 and 8), and *presented* in a dossier (step 8 and 9). The information about the customer investigated is *utilized* in the inquirers process (step D).
Figure 4 depicts the problems and effects on quality categories concerning the information evolution lifecycle in the request phase of the CIP in a fishbone diagram. The assignment of the process activities to the information quality types of the lifecycle is indicated by the numbers of the process steps.

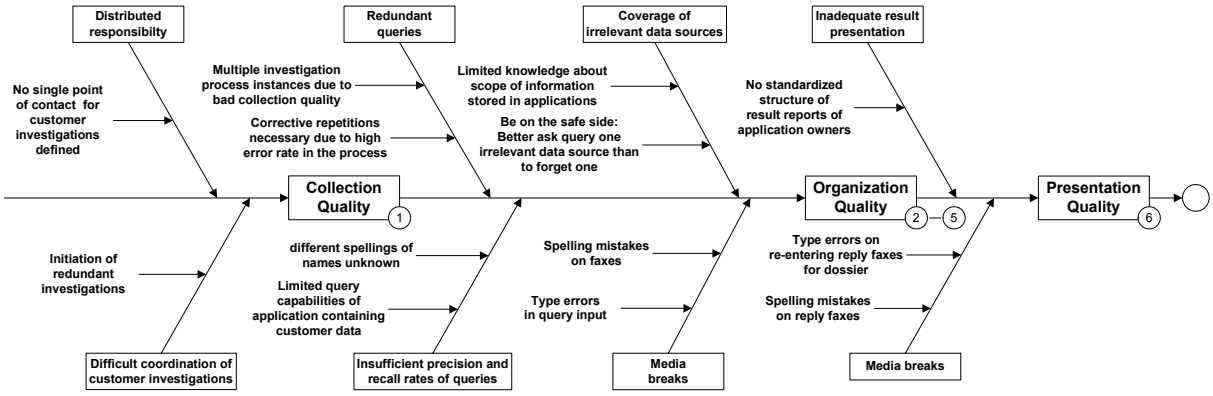

Figure 4: Potential causes of information quality problems and effects on information quality types in the request phase of the customer investigation process

During the acceptance of an inquiry (step 1, *collection quality*) the main problems were caused by *distributed responsibility* and the *coordination difficulties* of inquiries. To be on the safe side an inquirer tried to identify one or more consignees for his information request. That did lead to *unnecessary redundant instances* of the process, since more than one department started the investigation and was searching for requested information, often resulting in querying the same legacy applications.

Each investigation receiver then tried to identify potential departments who might own applications or archives containing information relevant concerning the inquiry and distributed a request to these application owners (step 2 and 3, *organization quality*). Redundant instances of the CIP and corrective repetitions of the process to ensure sufficient information quality lead to an unnecessary high workload of the application owners. Since each investigation receiver could only guess who might have information concerning the inquiry and only had a limited knowledge about the scope of information stored in the different applications, also unnecessary requests were sent to application or information owners. Consequently data sources not relevant for the investigation purpose were covered.

Further causes for potential organization quality issues were the many *media breaks* in the process. Faxes containing the request to the application owners may have contained spelling errors of names. Especially terror-related lists contained considerable spelling mistakes or incomplete spellings of names (step 3). Furthermore, limited query capabilities of common IT tools required an in-depth knowledge of the application owner about foreign names and potential name-variants to prepare the query based on the fax request (step 4). To avoid an *insufficient recall rate* in these situations a large effort had to be made by entering various potential spelling versions of the same name (step 5). The manual entry of queries into the applications could cause unrecognized type errors which led to insufficient query results – both by potentially missing entries (bad recall rate) and by producing too many results (bad precision rate) (step 5). Furthermore there might be a gap between the name entered in the bank's legacy system during the account opening process (where an official identification document, such as a passport, is required) and the name sought by the regulator. These names do not necessarily need to be same. If no additional identification element is being supplied such as birth dates, the assessment of candidates can become a difficult task.

Presentation quality in the request phase was mainly affected by the lack of standardization of replies in step 6. Query *results* delivered by the application owners were *heterogeneously structured* in the way the application presented the query result, e.g. in screenshots of applications or print-outs of scanned documents. Additionally, *media breaks* caused by sending the query results by fax to the inquiry receiver deteriorated the quality of the presentation of the results.

All these problems and their effects on overall information quality affect the information evolution lifecycle in the reply phase (cf. Figure 5). Faxes or e-mails with query results may have been hardly interpretable or ambiguous for the inquiry receiver or could contain insufficient information (step 7) which in turn affected *collection quality* in the reply phase. This could either result in a wrong decision at

the end of the process or in a corrective iteration, which could become very costly and time consuming. Again, media breaks could lead to lost results or to information collection errors when re-entering the information delivered into a computer system to prepare the dossier for the inquirer.

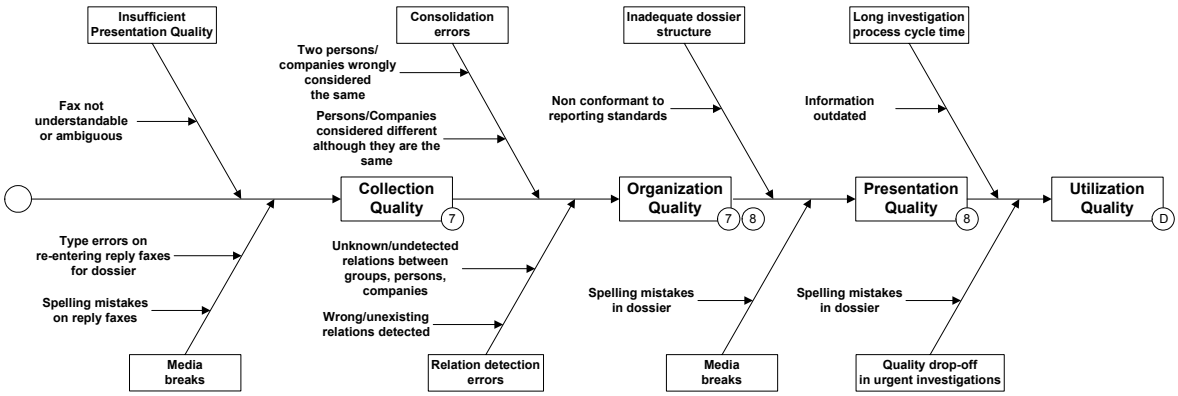

Figure 5: Potential causes of information quality problems and effects on information quality types in the reply phase of the customer investigation process

A major problem when consolidating the information and preparing the dossier (step 7 and 8) concerned *information organization quality*. Information had to be consolidated manually and relations between persons, companies and groups had to be detected by hand. In this task *consolidation errors* as well as *relation detection errors* might have occurred, e.g. when two information records were considered describing the same person or company by mistake or when relations that did exist between persons and companies were not detected.

When preparing the dossier, again *media breaks* could affect *presentation quality* negatively. Also, the structure of the dossier might not have been adequate for the inquirer (step 8).

All the problems and effects on information quality affect *utilization quality*. The inquirer had to take action based on the delivered information. Especially the long cycle time of the CIP could lead to outdated information or to an unacceptable latency for taking action. Investigations that were marked as extremely urgent could only be processed at the cost of quality, even if conducted carefully. Possible utilization errors could be an unnecessary locking of a customer's account or an avoidable investigation of the customer by public authorities potentially resulting in an unsatisfied customer.

REENGINEERING OF THE CUSTOMER INVESTIGATION PROCESS

As outlined in the introduction (cf. Introduction), recent political and regulatory developments led to a higher frequency of customer investigation inquiries and compliance checks. Consequently, a higher information quality was needed to ensure that no inconveniences (e.g. accidentally locked customer accounts) occurred due to quality issues in the CIP. Furthermore, the increasing workload of operational departments and the rising costs of the original process were perceived as unacceptable. Therefore, it was decided to reengineer the process to enable meeting the regulatory demands efficiently. This was done in an evolutionary approach with two phases (cf. Figure 6).

In the *first phase*, the focus was on the effectiveness of the CIP. The major changes of this phase were the centralization and formalization of the organizational process. As a single point of contact for all customer investigation inquiries a new organizational unit "Investigations & Inquiries" (CLFI) was established. This organizational unit became responsible for all activities concerning compliance management and legal inquiries. Therefore, it is also responsible for customer investigations related to legal inquiries. To initiate the centralization and to standardize the investigation process, a comprehensive analysis of the original CIP was conducted. All information resources relevant for customer investigations

were catalogued. For each information resource the following activities were carried out:

- It was documented, which information each resource contains.
- The information request and reply process for each resource was standardized and documented.
- Dependencies between the different information requests were analyzed and documented, e.g. the customer information file (CIF) has to be queried before other databases are queried, since information gathered from CIF is needed for the second step.

Based on this analysis a new process was crafted. The centralization and standardization of the process significantly improved the coordination effort and avoided redundant information requests. Furthermore, the information request and reply processing was improved by replacing fax communication trough e-mail workflows. This eliminated many of the media breaks identified in the original CIP. After the redesigned process had been adopted and had proven to be effective, the focus shifted to make the process more efficient.

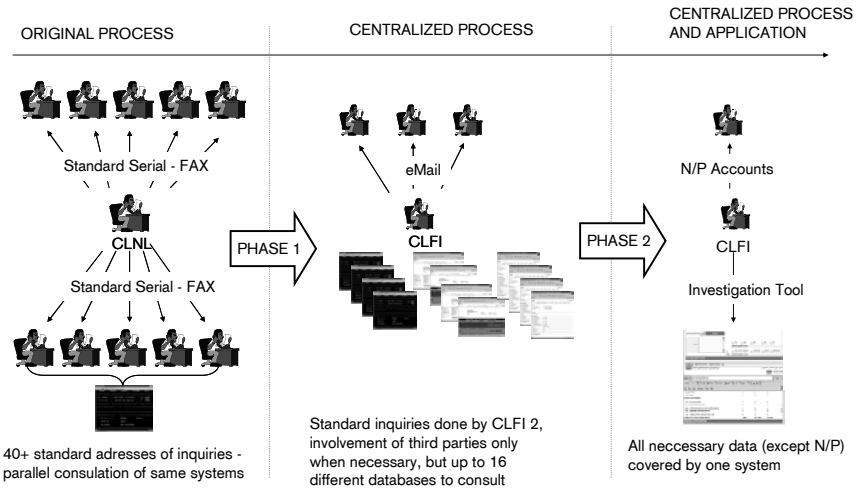

Figure 6: Evolution steps of the Customer Investigation Process

Therefore in the *second phase*, large parts of the CIP were automated by developing a new integrated investigation application. The automation of the process addressed the error-prone tasks of searching, analyzing and consolidating customer information and enabled preventive investigations with a guaranteed high information quality. Furthermore, the use of the application enabled an additional simplification of the organizational process. Though, the process designed in the first phase is the fallback-scenario if technical problems with the investigation application should arise.

Both measures in combination resulted in a simplified and efficient process with a higher quality of customer investigations. The new process and its impact on information quality are described in the subsequent sections.

THE REVISED PROCESS—AUTOMATED BY AN INFORMATION SYSTEM

The following sections illustrate the organizational and technical measures taken by reengineering the customer investigation process and by implementing a data warehouse based investigation application that supports the new process by automating most of the error-prone tasks of the original process.

Workflow of the Revised Customer Investigation Process

The establishment of the central organizational unit “Investigations & Inquiries” and the adoption of the investigation application enabled to process customer investigations completely within the CLFI unit. The reengineered process consists of only a few activities (cf. Figure 7).

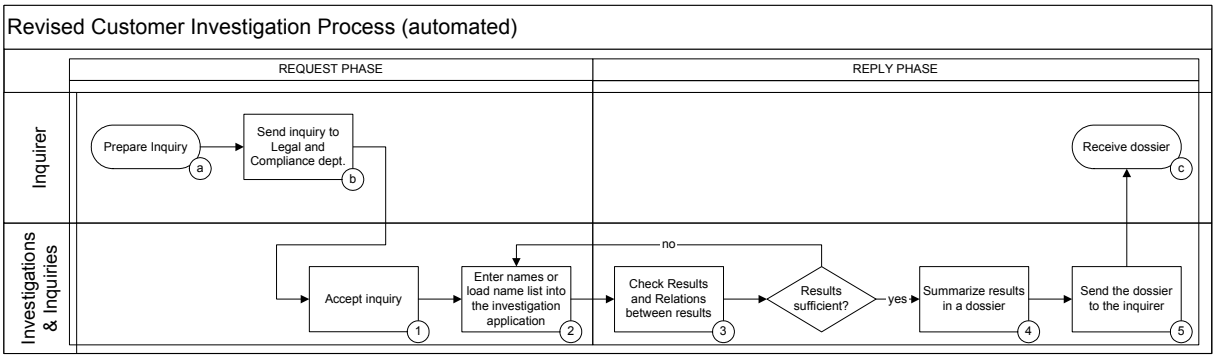

Figure 7: Revised customer investigation process at Credit Suisse

An inquirer (e.g. external/government or internal person who requests that information) sends a prepared inquiry to "Investigations & Inquiries". The organizational unit accepts the request (1) and enters the name of the persons or organizations to be investigated into the investigation application described in the following section of the paper (2). The investigation application delivers all known information about the investigated subjects. This information has to be checked, including relations between subjects (3). If collected information is sufficient to reply to the inquiry, results are summarized in a dossier (4) that is sent to the inquirer (5).

The Investigation Application — A System automating the Investigation Process

The investigation application automates a large part of the CIP by providing two major functions: First, an interactive search function and second, a compliance check function. It was built in collaboration with DeltaVista (http://www.deltavista.com) providing expertise in address management and database search algorithms. The following sections describe functions, architecture and design considerations of the application in detail.

Description of Functions Provided by the Application

Interactive Search Function: The interactive search function enables the user to find all information related to a certain person and to interactively explore that information in a drilldown fashion. Since that person does not need to be an existing customer of Credit Suisse, the only identifying attribute is the person's name. Therefore, the investigation application was specifically designed to only require a person's name as an input, although hit precision can be increased by entering further information like address, birth date etc. The investigation application then automatically searches all selected data pools for this customer name using a fuzzy search algorithm. Depending on the use case, the user may select one or more data pools described in Table 2.

For example, if an inquiry is needed whether or not a certain individual might be on a wanted terrorist search list, the world-check data pool is used or if an anti money laundering investigation has to be conducted, all data pools are selected.

DATA POOL	CONTENT
Internal Data	
Customer Information File	All customers being active or set inactive since 2001, who may have an account, a depot etc.
Power of attorney	Information about who is authorized to sign on the behalf of a certain customer
Unwanted customers	Information about persons, firms and organizations with whom the bank does not want to engage in a relationship due to a high risk
Consumer credits	Customers who have a consumer credit or a leasing contract
Street transactions e.g. currency exchanges	Image Archive of documents, indexed with names of contract partners and/or proxy agents
Safe customer data	Persons who own or did own a safe as well as location and type of the safe

DATA POOL	CONTENT
External Data	
Swiss Addresses and Company Information (DeltaVista)	Address and Company information integrated from a variety of official sources such as yellow pages, phone directories, registration offices and commercial data providers, e.g. Dun&Bradstreet, Orell Fuessli.
Bankruptcies (DeltaVista)	Persons and firms associated with insolvency proceedings
Swiss Official Gazette of Commerce Publications (DeltaVista)	Commercial registry (new entries, mutations, deletions), Bankruptcies, Composition agreements, Debt enforcement, Calls to creditors, Other legal publications, Lost titles, Precious metal control, Marks, Balances, Public procurement, Infoservice, Company publications
World Check Data (World Check)	Names of individuals, firms, organizations, parties and groups that may cause a high risk in a potential customer relationship with the bank, e.g. politically exposed persons, persons associated with terrorist organizations and/or criminal organizations

Table 2: Content of the data pools

The application presents the search result in a list (cf. Figure 8), sorted by hit probability. Each entry of the result set carries the following information: Hit probability in percent (1), name and address of the person or company (2-6), birth date of the person or foundation date of the company (8), an identification number (e.g. customer information file number) if the person has a business relationship with Credit Suisse (8), and an indicator for the data pools (one or more internal and/or external) where the person, group or company was found in (9).

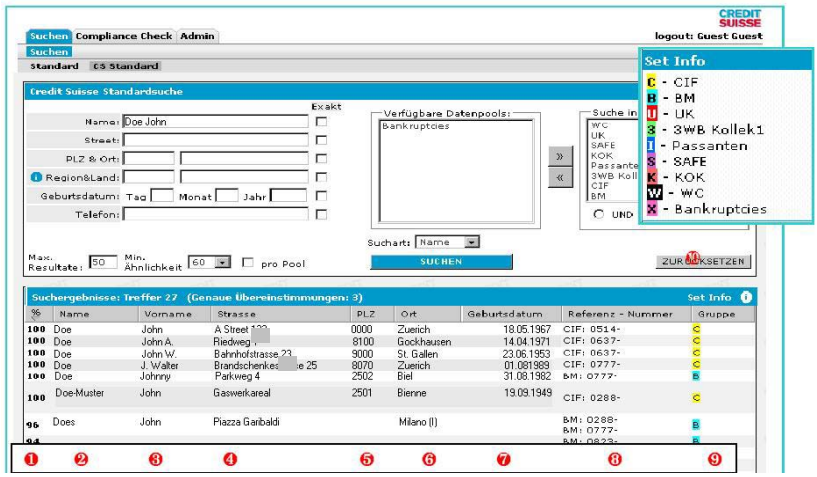

Figure 8: List of search results delivered by the investigation application (fictitious data)

The user may then reorder the list by one of the criteria above, browse the list, select a certain entry, request a report or navigate to related information, e.g. companies, customers or other persons associated with the selected entry. Figure 9 shows a typical screen report for a certain individual. It shows the current address, the move history of the person, whether the address is validated and the data source of the address information (1 and 2). The buttons below the address information allow to be linked directly to the legacy applications containing the source data and thus allowing to investigate real-time data regarding the customer (3). If the person is listed in one of the Credit Suisse data pools, then for each data pool the corresponding information contained about the person is displayed (4). For each data pool entry, a more detailed view can be generated. Furthermore, all available Official Publications about a company shown in (6) as well as all available address details (5), including household (7) and move history (8) can be viewed.

To further investigate the person or company, an address detail view, a specific detail data pool view and a company detail view can be accessed. A special feature of all these detail views is the connection to further associated individuals. For example, the company detail sheet shows all owners or members of the administrative board of the company, the address detail view shows further persons living in the same household etc. This allows an investigation not only of the person or company searched for but also of

persons or companies related to the name entry. Experienced staff from "Legal & Compliance" department of Credit Suisse uses this function to detect typical crime patters of money laundering and to find individuals who are or might become customers of the bank and are associated with criminal groups or persons etc. Interestingly enough, external data provided by DeltaVista (such as a large amount of Swiss addresses and company information) serves as very powerful information to enhance or backup decision making and enhance internal data from Credit Suisse. By matching third-party information to Credit Suisse data, the decision making process can be faster and more accurate. Otherwise not known relationships between clients and companies can be detected automatically—an information that may not be available by using only internal data.

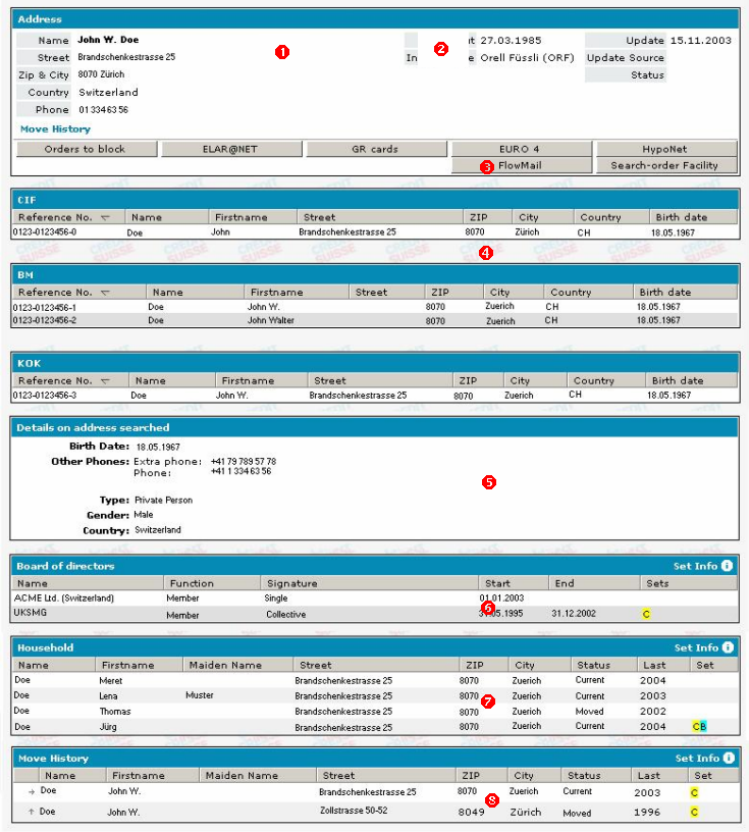

Figure 9: Report screen providing detailed information summary about a result set entry (fictitious data)

Compliance Check Function: The compliance check function allows for testing all data pools against a complete list of names. This function is used in connection with prosecution of terrorists. Government or a regulating body sends a list of names to the Credit Suisse which has to be checked against the customer base. A simple name comparison is not sufficient for this task: Different spellings or aliases of the names have to be detected and also relations between the names on the list and active customers have to be reviewed.

Architecture of the Investigation Application

The investigation application is based on two major subsystems: first, a data mart that is fed by the central Credit Suisse data warehouse, which delivers all the necessary internal data; second, another data mart containing external data as well as the matching and fuzzy search intelligence provided by DeltaVista. Since the data warehouse is an established system within Credit Suisse that daily collects data from over 300 data sources from Credit Suisse's legacy systems, most of the data sources needed for the investigation process where already made available in sufficient quality and timeliness. Therefore only a

few additional external data feeds had to be developed additionally, such as the integration of the registry of politically exposed persons delivered by another third party provider (World-Check). Consequently, the internal data pools were easily integrated into the investigation application, as the implementation project only had to deal with a single data delivery interface. Using the data warehouse as a provider for internal data also ensures accurateness of the data delivered, since the data warehousing process within Credit Suisse contains an established closed loop data quality management. Also, data cleansing and consolidation is done within the extract, transformation and load (ETL) processes that load the data warehouse. A further benefit of using the data warehouse infrastructure is that scalability and stability of the application is high and the application is managed professionally by an operations and batch controlling department. Having standardized and automated secure feeds for both internal and external data, expanding the contents of the system can be performed very effectively in short time. This ensures, that new requirements can be met, if either regulatory requirements change or a new application / instance of the tool for other purposes become necessary.
The external and internal data is integrated during the ETL-process of the dedicated data mart of the investigation application. A core concept used to integrate related data of persons or companies and to build probable relations that exist between the persons is the *Address Universe* described below.
The investigation application itself provided by DeltaVista is based on web technology and runs on an application server that is part of the Credit Suisse Intranet. This ensures an easy access to the application and an intuitive use of the application over an internet browser. Besides the matching and search algorithms, the application provides the necessary navigation and presentation capabilities used to investigate the retrieved information. Due to the sensitivity of the data stored in the application and the ease of retrieving data, the hardware is running in a secure environment. To cope with data-protection issues, access control to the system is regulated and restricted to a low number of specially designated investigators using a security logon certificate.

Core Concepts of the Investigation Application—Address Universe and Fuzzy Search

The investigation application is based on two core concepts contributed by DeltaVista. First, the so called Address Universe is used to link people, customers and companies with each other. Second, a fuzzy search algorithm is used to search for names in the data pool to find an entry point for further browsing the data.
Address Universe: The Address Universe comprises of all addresses stored for investigation purposes. It is used for linking related persons, groups and companies. Furthermore it is the main index for searching the data to investigate. The Address Universe contains about 21 million addresses, from which about 13 million are from external sources. About 6 million older addresses represent the move history of the currently valid addresses. Often, data from internal data sources have an invalid or incomplete address. Therefore, a special consolidation process is used during the ETL-process of the investigation data mart to connect internal addresses with similar external validated addresses. In this process, all new data is compared with existing names and addresses in the Address Universe and a similarity metric is computed. If a certain similarity value is trespassed, the two addresses are considered as equal and get linked with each other. When an address is considered equal to an existing address in the Address Universe the existing address is linked with the new data that is loaded into the investigation data mart. The link is stored in a special “connection-table” to make the connection reversible. Additionally, a special “relation-table” stores all known links between companies, people and groups. (cf. Figure 10) These links may then be navigated by the user using the interactive search function. These relations are also extracted during the ETL-process from operational systems and external data sources. Complementary, also a special “declined-relations-table” exists that is used to store explicit ‘non-links’. This enables marking two persons or companies explicitly as non-related if this is for sure, e.g. if a Credit Suisse customer and a well known terrorist have similar names but are certainly not affiliated with each other.

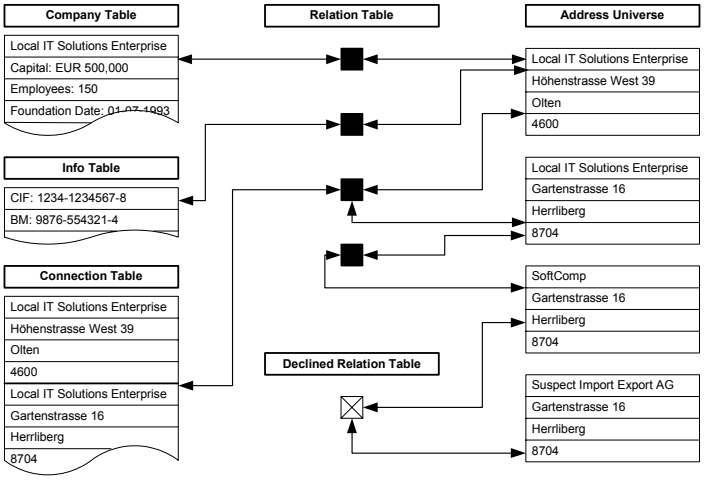

Figure 10: Extract from the data model of the application – tables for linking addresses (fictitious data used)

Fuzzy Search: Names are often slightly misspelled due to typing errors, have different official spellings, e.g. foreign people who carry their original name and a German alias, or because people have aliases to hide their identity, e.g. terrorists. This is a considerable problem in many customer investigations since it decreases recall rate of the retrieved information. To overcome this problem, the investigation application implements a fuzzy search algorithm that is capable to index the name database in a way that information about people or organization with misspelled names is also retrieved. An important feature is that this is of course not only working for German names but also for those in foreign languages or coming from a different cultural context, e.g. Arabic, Russian or Asian names. If in need, the algorithm can also be fine-tuned for a specific language. To leverage search-capabilities a customizable synonym search capability is available as well.

Design concepts and considerations of the investigation application

In the design phase of the user interface, special attention has been brought to the display of search results: Despite all intelligence available in the system, the final decision on whether a result is relevant for the inquiry or not still needs to be made by a human being. Therefore the result-list needs to contain sufficient data to decide. Should a closer look at the details of the customer and further investigation be needed, then as little data as possible is needed to avoid confusion through overloading of the screen. The initial option of having pop-ups to display further information has been quickly abandoned, because it turned out, that the users did not like the nervousness created by them, while following the result list. Colorized icons quickly help to identify records with a link to Credit Suisse data, which is helpful, if the investigator is also considering external data. All result list rows can be sorted: this helps to arrange the initial search results and the selection process for relevant information.

In the so-called *Interactive Search* mode used for the normal inquiries, most of the work results from reproducing documents for a relationship identified to match an inquiry. In the *Compliance Check* module, where lists get processed, an effective clicks-per-result ratio is even more crucial. Lists such as sanctions lists, terrorist lists or even a PEP list, tend to be much less accurate and comprehensive, what additional search criteria's such as birthdates other addresses is concerned, compared to individual inquiries). When processed by the system, an obvious false positive match is the norm; a match that needs a closer look at the exception confirms that rule.

Keeping this in mind, the application has been adopted to have a very small click per false positive rate in order to allow a high output. As a rule of thumb, the slogan "think or start with the false positives first" has been introduced.

Assessment of the Measures—The Value of Improving Information Quality

Compared with the original CIP, the redesigned process is much simpler. The centralized entry point for investigation inquiries prevents redundant executions of the process. It also simplifies the procedure for the inquirer since he is offered a single point of contact delivering consolidated information. The application automates the most error-prone tasks of identifying and querying data sources, consolidating information and detecting relations, eliminates the media breaks and improves presentation quality by providing standardized reports for customer investigation. Table 3 summarizes how the problems affecting information quality were addressed by the process redesign and the investigation application.

QUALITY TYPE	PROC. STEP	PROBLEM CATEGORY IN ORIGINAL PROCESS	MEASURE AND ASSESSMENT OF MEASURE
Request Phase			
Collection Quality	1	Distributed responsibility and difficult coordination of multiple inquiries	Organizational centralization of the CIP effectively prevents unnecessary investigations.
Organization Quality	2-5	Redundant Queries	Redundant queries are prevented; explorative queries for refinement of results do not affect operational systems or people from operational departments.
		Coverage of irrelevant data sources, insufficient recall rate, media breaks	The investigation data mart contains all relevant customer information. If additional sources become relevant, they will be integrated into the data mart and become available for each investigation. Recall rate is improved by a fuzzy search algorithm and by explicitly stored relations between investigated subjects in the investigation data mart. Media breaks are eliminated by the automation of the information retrieval.
Presentation Quality	6	Inadequate result presentation, media breaks	Automation of information retrieval eliminates this process step and the corresponding problems.
Reply Phase			
Collection Quality	7	Insufficient presentation quality, media breaks	Automation of information retrieval eliminates this process step an the corresponding problems
Organization Quality	7-8	Consolidation errors	Automation of information consolidation during the ETL-process of the investigation data mart eliminates most consolidation errors caused by manual consolidation.
		Relation detection errors	Automation of relation detection during the ETL-process of the investigation data mart eliminates most relation detection errors. The table storing relations and explicit non-relations increases reliability and reproducibility of investigation results.
Presentation Quality	8	Inadequate dossier structure	The Investigation application provides standardized reports that can be used for creating the final dossier.
		Media breaks	The complete task of dossier assembly can be done electronically. Media breaks are not necessary.
Utilization Quality	D	Long investigation cycle time	Automation and centralization of investigation know-how shortens process cycle time rapidly (from about 120 to about 20 minutes).
		Quality drop-off in urgent investigations	High information quality is guaranteed though high degree of automation and through data quality assurance during the data warehousing process. Investigation quality in urgent cases is much higher. Though, quality drop-offs may occur when results presented by the investigation application are not interpreted carefully.

Table 3: Assessment of the measures taken to improve information quality

Overall, the automation of the information preparation and retrieval process alone resulted in a cost reduction effect of approx. 3.1 million EUR (net present value calculation over 5 years). Investigation process cycle time could be reduced to one sixth of the original duration. Furthermore, the short cycle time and the high degree of automation enable preventive investigations and can assist Relationship Managers in their customer due diligence procedures as well. Therefore, the measures taken can be considered effective and successfully implemented. The new requirements regarding customer investigations are fulfilled with the redesigned IS-supported process.

CONCLUSIONS

The article depicts how a combination of organizational and technical measures led to a significant information quality improvement in customer investigations at Credit Suisse: First, the organizational centralization of responsibility for customer investigations concentrated know-how necessary to interpret customer information and avoided costly redundant process instances. Additionally the process was simplified for the inquirer as well as for the inquiry receiver. This simplification was enabled by a new application designed for customer investigations. This data warehouse based investigation application has well proven to boost quality and speed of customer investigations. The information quality gain and the cycle time reduction of the process enable preventive investigations and proactive risk management, as it becomes necessary due to recent regulatory developments. It also turned out that both effects of the measures resulted in a significant cost saving.

BIBLIOGRAPHY

[1] U.S. Governance, "To deter and punish terrorist acts in the United States and around the world, to enhance law enforcement investigatory tools, and for other purposes.," in *H.R.3162*, *U.S. Governance*, 2001.

[2] FATF, "Special Recommendations on Terrorist Financing," 2001.

[3] T. Wolfsberg Group, "The Suppression of the Financing of Terrorism - Wolfsberg Statement," 2002.

[4] Basel Commitee on Banking Supervision, "Customer due diligence for banks," Bank for International Settlements, 2001.

[5] Basel Commitee on Banking Supervision, "Overview of The New Basel Capital Accord," Bank for International Settlements, 2003.

[6] T. Wolfsberg Group, "The Wolfsberg AML Principles - Frequently Asked Questions with Regard to Politically Exposed Person," 2003.

[7] U.S. Department of the Treasury - Office of Foreign Assets Control (OFAC), "Specially Designated Nationals and Blocked Persons (OFAC List)," 2004.

[8] K.-T. Huang, Y. W. Lee, and R. Y. Wang, *Quality Information and Knowledge*. Upper Saddle River, NJ: Prentice Hall, 1999.

[9] R. Y. Wang, M. P. Reddy, and H. B. Kon, "Toward quality data: An attribute-based approach," *Decision Support Systems*, vol. 13, pp. 349-372, 1995.

[10] M. J. Eppler and D. Wittig, "Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years," presented at 5th International Conference on Information Quality (ICIQ 2000), Cambridge, MA, 2000.

[11] L. Liu and L. N. Chi, "Evolutional Data Quality: A Theory-Specific View," presented at 7th International Conference on Information Quality (ICIQ 2002), Cambridge, MA, 2002.

[12] H. Miller, "The Multiple Dimensions of Information Quality," *Information Systems Management*, vol. 13, pp. 79-82, 1996.

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

[14] L. P. English, *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. New York, NY: Wiley, 1999.

[15] R. Y. Wang and D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management of Information Systems*, vol. 12, pp. 5-33, 1996.

[16] M. Helfert, G. Zellner, and C. Sousa, "Data Quality Problems and Proactive Data Quality Management in Data-Warehouse-Systems," presented at BITWorld 2002, Guyaquil, Ecuador, 2002.

[17] D. P. Ballou and H. L. Pazer, "Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems," *Management Science*, vol. 31, pp. 150-162, 1985.

[18] D. Te'eni, "Behavioral Aspects of Data Production and Their Impact on Data Quality," *Journal of Database Management*, vol. 4, pp. 30-38, 1993.

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