Institutionalizing Information Quality Practice:
The S. C. Johnson Wax Case

James D. Funk
S. C. Johnson Wax

JDFunk@scj.com

Yang W. Lee
Northeastern University
Mass. Institute of Tech.
YWLee@cba.neu.edu
ylee@mit.edu

Richard Y. Wang
Boston University
Mass. Institute of Tech.
rwang@bu.edu
rwang@mit.edu

Abstract Institutionalizing a total data quality management (TDQM) program that will sustain long-term benefits is a challenging endeavor and an area that has not been systematically approached. In this paper, we present how S. C. Johnson Wax starts to institutionalize their TDQM. We illustrate the effort in the context of a data warehouse project and show how information quality tools are deployed to facilitate these processes.

1. Introduction

Planning and implementing a program to achieve a state of high information quality (IQ) for short-term results and long-term benefits is a non-trivial undertaking [3]. Organizations from different industries, with disparate business goals and operating environment, call for developing an IQ management program that fits their purposes. The experience of S.C. Johnson Wax (SCJ) presents an example of what can be done to institutionalize a Total Data Quality Management (TDQM) program [7, 9]. The SCJ experience is a valuable input to other firms contemplating on launching a TDQM program although it represents the approach of only one company.

The business drivers for the TDQM initiative at SCJ include two business mandates: (1) SCJ needs to operate on a global basis instead of a multinational basis; (2) SCJ needs to reduce the overall costs for managing the firm's information assets. Based on these mandates, a clear goal for the TDQM emerged: *Deliver the right business information to the right person at the right time.* This explicit goal entails improved customer information, improved product information, and improved vendor information. Toward this goal, SCJ launched a project to develop a global data warehouse [2].

A key decision was made early on in the data warehouse project to ensure the quality of data before data are loaded into the data warehouse. Thus, a target for this initiative is to institutionalize IQ processes throughout SCJ that will form the basis for continuous IQ improvement. Additionally, the TDQM objectives are to develop SCJ IQ measures, a global reporting capability, and continuous IQ improvement processes.

2. Company Background

S. C. Johnson Wax (SCJ), often referred to as Johnson Wax, is a leading provider of household products. The company has been in business for over 110 years in manufacturing consumer goods and has operations in over 50 countries around the world [4]. Its newest countries of operations include Slovenia, Slovakia, and Kazahkistan.

It is a leading provider of household consumer package products that provide solutions for consumers in the areas of air care, insect control, personal care, and specialty cleaners for a variety of tasks within the home. It also provides consumer packaging solutions with its recent acquisition of the consumer products division of DOW Chemical Company. These products include the ZIPLOC plastic bags that can be used to solve a variety of home storage issues. The leading brands sold by the company are Glade, Pledge, Windex, Raid, and Edge.

The focus of SCJ's business is on meeting the needs of the consumer to solve their problems relating to air quality, cleaning needs and insect problems within the home. It is aggressively realigning its operations to become a global operation as opposed to a series of multinational operations. SCJ has experienced rapid revenue growth due to strong marketing initiatives for its core brands and two major acquisitions that strengthened its competitive position by providing strong consumer products in new product categories in the consumer package good aisles within mass merchandising, drug, and grocery segments of the retail business.

The company has a strong commitment to innovation and product development and delivering products with superior solutions or what they call "product plus." Internal resources are externally focused: meeting the emerging needs of customers. Despite this external focus, SCJ has a strong historical track record in investing in quality initiatives for internal processes. It has a strong TQM function within the manufacturing process and the TDQM initiative has drawn from the lessons learned within the company as it developed and improved quality processes associated with its manufactured products. This includes investing in scanners for use within the warehouse operations, extensive use of EDI transactions in dealing with business partners, and extensive analysis of internal and external sales information to better meet the needs of customers and consumers.

Recently, there has been a great emphasis within the company to analyze sales information to both SCJ's customers and to the ultimate consumers, users of SCJ products. In that regard, SCJ has initiated the assessment of the quality of the information that they are maintaining for both their internally generated data and the data that they receive from external sources concerning the movement and sale of their products. The external data come from both SCJ's own customers and third party sources that collect and provide scanner data at the retail location. SCJ now needs to have a higher level of quality for the data it uses to enable the successful analysis of customer and consumer purchase patterns, which will enable them to better serve the consumer. The net result is increased demand for data quality and reliability. This requires a shift from project based or functionally focused quality efforts to a company-wide total quality management program.

This shift to a total quality management program incorporates the need for a process to continually assess the perception of the internal consumers of the data and a process to continuously improve the actual quality of the data.

3.TDQM at SCJ

Implementing the TDQM at SCJ is based on the following two major assessments. *First, the function view of information must be transformed into that of the enterprise view.* Functional design of information hinders global business initiative. The legacy of stove-piped functional development of information systems creates islands of automation, resulting in disparate IQ between the functional and enterprise levels. *Second, piece-meal and ad hoc projects must be transformed into long-term and institutionalized IQ practice.* Every local unit of SCJ was striving for business excellence with quality data. Without reflecting on what multiple departments in global areas produce and how they impact each other, however, all the efforts for productivity and efficiency generates long-term negatives for business.

Reviewing and understanding information flow, business processes, and associated information technologies across functional areas set the basis for launching the TDQM initiatives. Based on the above assessments, a TDQM program is being instituted with the explicit TDQM cycle incorporated. As shown in Figure 1, The *definition* component of the TDQM cycle identifies important IQ dimensions [10] and the corresponding IQ requirements. The *measurement* component produces IQ metrics. The *analysis* component identifies root causes for IQ problems and calculates the impacts of poor quality information. Finally, the *improvement* component provides techniques for improving IQ [9].
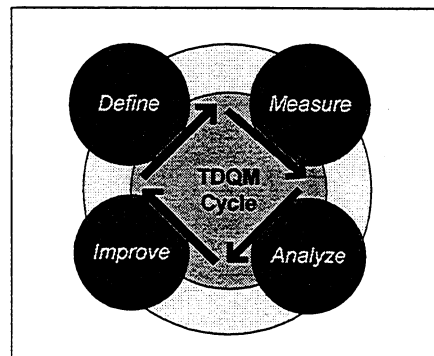

Figure 1: Components of the TDQM Cycle

4. Initial IQ Assessment

Implementation of a TDQM program first requires an initial assessment of the perception that employees have concerning the data that they use. In 1997, SCJ initiated a process whereby the authors conducted an assessment of key data products with the analysts within the company. The analysts were given a survey instrument called IQA that collected information about the subject's assessment of a data set [3, Ch. 4-5, 9]. The collected information was analyzed.

As shown in Table 1, the analysis indicated that overall, most databases were assessed as highly relevant and adding value to SCJ's business but their IQ needs to be improved in the specific IQ dimensions such as *consistency, concise representation,* and *ease of manipulation.* In particular, *consistency* and *concise representation* were rated low by finance/accounting and information systems departments. *Ease-of-manipulation* was rated low by marketing & sales, production & manufacturing, and finance & accounting [5].

Based on the results, SCJ identified areas for further work in IQ improvement. SCJ also identified the need to expand the scope of the assessment and to develop an ongoing process to continually assess these quality perceptions.

Table 1: Information quality by Department

Department	High-quality IQ Dimensions	Low-quality IQ Dimensions
Finance & Accounting	H1. Relevancy H2. Value added H3. Interpretability	L1. Consistent representation L2. Ease of manipulation L3. Concise representation
IS Department	H1. Relevancy, Value added H2. Objectivity H3. Timeliness	L1. Security L2. Appropriate amount of Info. L3. Concise representation
Marketing & Sales	H1. Relevancy H2. Value added H3. Timeliness	L1. Ease of manipulation L2. Interpretability L3. Understandability
Production & Manufacturing	H1. Relevancy H2. Value-added H3. Objectivity	L1. Security L2. Ease of manipulation L3. Free of Error

IQ seminars were also given to these data analysts to expose them to the state-of-the-art of the IQ field and the TDQM. These training seminars and

discussions with participants provided opportunities for the participants to understand and evaluate the IQ environment and the IQ status at SCJ.

5. Barriers to High IQ

Further investigation of root causes for IQ problems at SCJ reveals factors such as different definitions, duplicate identifiers, complex coding structures, and disparate information quality at the functional and enterprise levels. The impact of these barriers is estimated at 60% of project time spent cleansing data for IQ problems.

Different Definitions of a term cause different interpretation and understanding of some business results, in part due to cultural differences. The problem is exacerbated by SCJ's need to become a global company and the development of a global data warehouse. Different understandings of terms necessitate SCJ to develop a global data warehouse structure and global data mart structures with dimensions such as the customer, product, sales organization, and time, as will be described later in the star schema of the data warehouse.

Duplicate identifiers include the same identifier for different items. For example, Code 123456 could mean a finished product in North America, a chemical formula in Europe, and a packaging component in Asia.

Complex coding structure is another area that resulted in poor IQ. For example, raw materials have intelligence in the first five digits. As a result, the two largest categories of raw materials were found to be "To be Determined" and "Miscellaneous." Embedded intelligence codes for the RD&E project number also created ranges for combinations of project type, laboratory location, and product category.

A subtler root cause of IQ problems arose from the need to aggregate data at the functional levels to the enterprise level. Although IQ might be high at the functional level, the quality of aggregated data at the enterprise level could be low. For example, sales revenue at the functional levels was correctly

recorded, but sometimes incomplete. When these revenue data are aggregated into the enterprise financial book, the reports were incorrect because estimations were made for those incomplete entries. Further efforts were needed to collect, aggregate and present the needed data.

A primary goal of SCJ's TDQM program is the development of a global data warehouse and global reporting system with associated data marts that will provide the needed information to the right person within the company. This initiative has strong functional support and leadership, which has been the key driver for improving IQ within the company. As the data warehouse was initiated, SCJ discovered IQ issues that constrained the capability of the data warehouse and associated data marts to deliver the appropriate information to the appropriate people in a timely manner.

The first effort at collecting data on a global basis uncovered the existence of data problems that delayed the delivery of needed information by months. The problems included duplicate keys for data items and inconsistent representation of key characteristics of the data that complicated the ability to aggregate and analyze this information using those key characteristics. An example of the effort needed to prepare data for use in the data warehouse was the need to resolve duplicate keys for packaging components and to develop global categories for key attributes of these packaging components on a global basis.

Prior to this effort information was maintained on a local basis and when requests were made on an infrequent basis to aggregate purchases of these packaging components, each local company spent weeks of effort simply to gather and format the data. Then an additional effort was made at corporate headquarters to further clean and aggregate the data sent by the local companies. This process typically would take several months. The first time SCJ tried to aggregate information about purchases of cans and aerosol valves on a global basis, it took three months for a team of business analysts and IT analysts to format the data for entry into the data warehouse.

6. A Star Schema for the Global Data Warehouse

To further our discussion, we present the star schema for a portion of the global data warehouse. A star schema consists of facts that are stored in fact tables and dimensions that are stored in dimension tables. Each dimension table contains attributes that provide descriptive characteristics about the facts. Certain attributes within a dimension are ordered in an attribute hierarchy. The attribute hierarchy yields a top-down data organization that permits both aggregation and drill-down or roll-up data analysis [8].

There are many conceptual dimensions in SCJ's business operations such as product, customer, and sales organization, as shown in Figures 2a-2c. The time dimension forms an attribute hierarchy from fiscal year, fiscal semester, fiscal quarter, fiscal month, to fiscal week, based on the SCJ fiscal calendar,
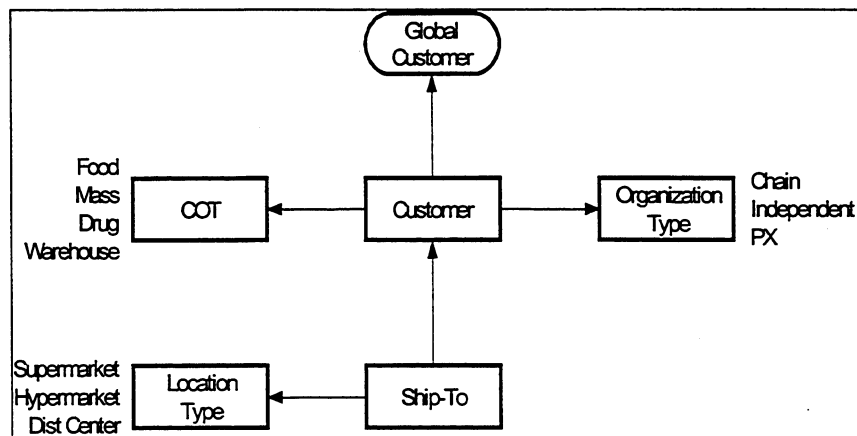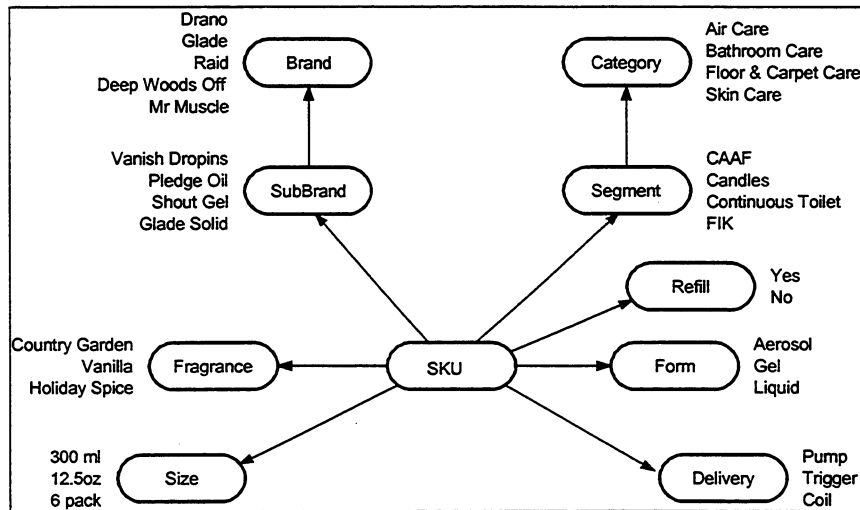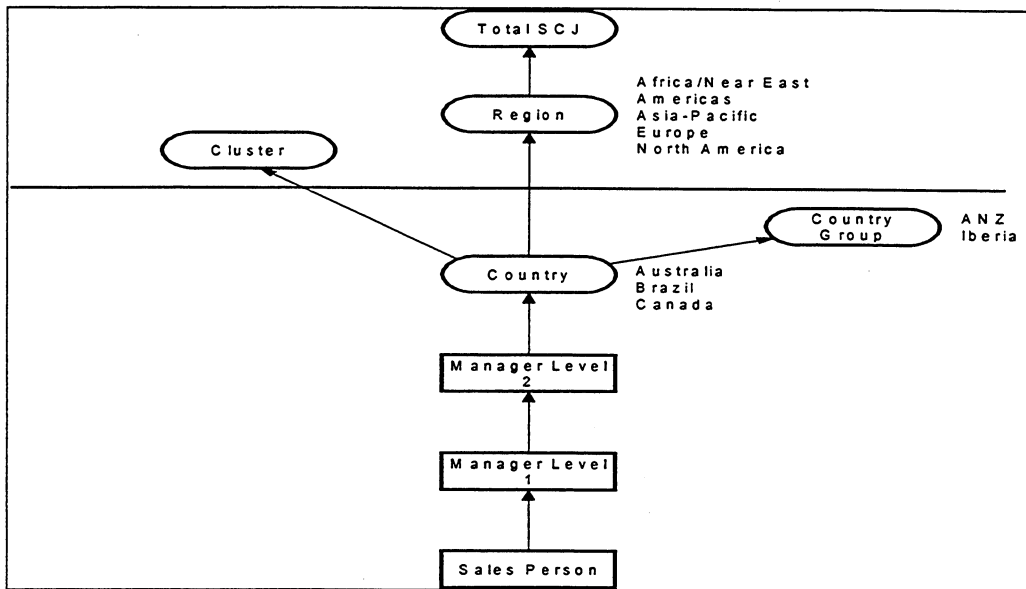
Figure 2a - 2c: Illustrative SCJ Business Dimensions

We illustrate the shipment data that is a portion of SCJ's overall data warehouse. As shown in Figure 3a, central to the star schema for the shipment data are three fact tables: the Ship_to_cust_fact, the Product_master_fact, and the Product_report_fact. Each fact table reflects a different view of the shipment data in the SCJ data warehouse regarding their products. The Ship_to_cust_fact table tracks how many cases of a certain product (and the dollar amounts) are shipped to a customer's location in an area for a certain week. It includes the number of cases (and their dollar amounts) shipped to the location, booked by SCJ, and returned to SCJ for a variety of reasons. The Product_master_fact table tracks how many cases (and the dollar amounts) of a certain product are shipped to a customer in the master customer file in a certain week. The Product_report_fact table tracks how many cases of a certain product have been reported based on the grouping as required by the sales department during a certain week.
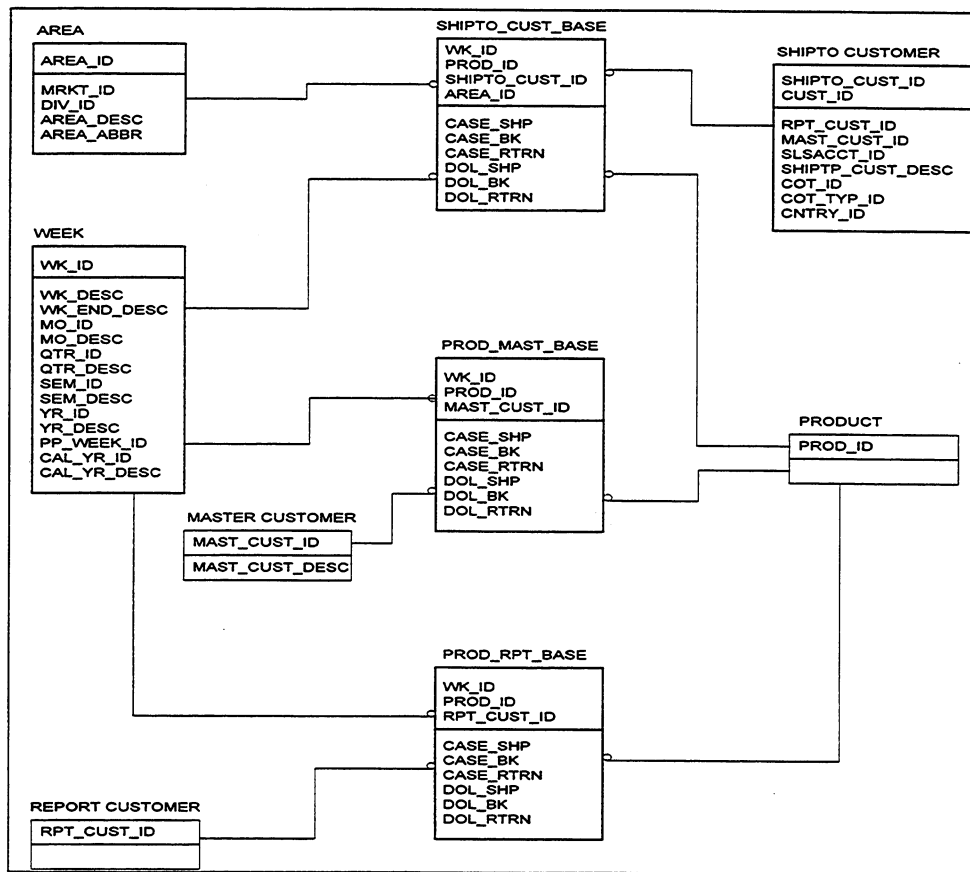
Figure 3a: A Star Schema for Shipment

Corresponding to these three fact tables is dimensions and attributes peripheral to the fact tables. As shown in Figure 3b, A division (e.g., East or West Ontario in Canada) contains many markets (e.g., Eastern Ontario) which in turn contains many areas (e.g., Middleton), yielding an attribute hierarchy. Similarly, A business group (e.g., Home Care) contains many categories (e.g., Total Raid); each category contains many brands (e.g., Raid Crack & Crevice) which in turn contains many products (e.g., Raid Crack & Crevice 350 GR12), yielding another attribute hierarchy. A Master Customer contains many intermediate grouping by sales that are referred to as report_customer (e.g., Kmart), each of them contains many ship_to_locations (e.g., Kmart Limited). Each Ship_to_location constitutes a fact table with other dimensions such as country, sales_account, and class_of_trade.
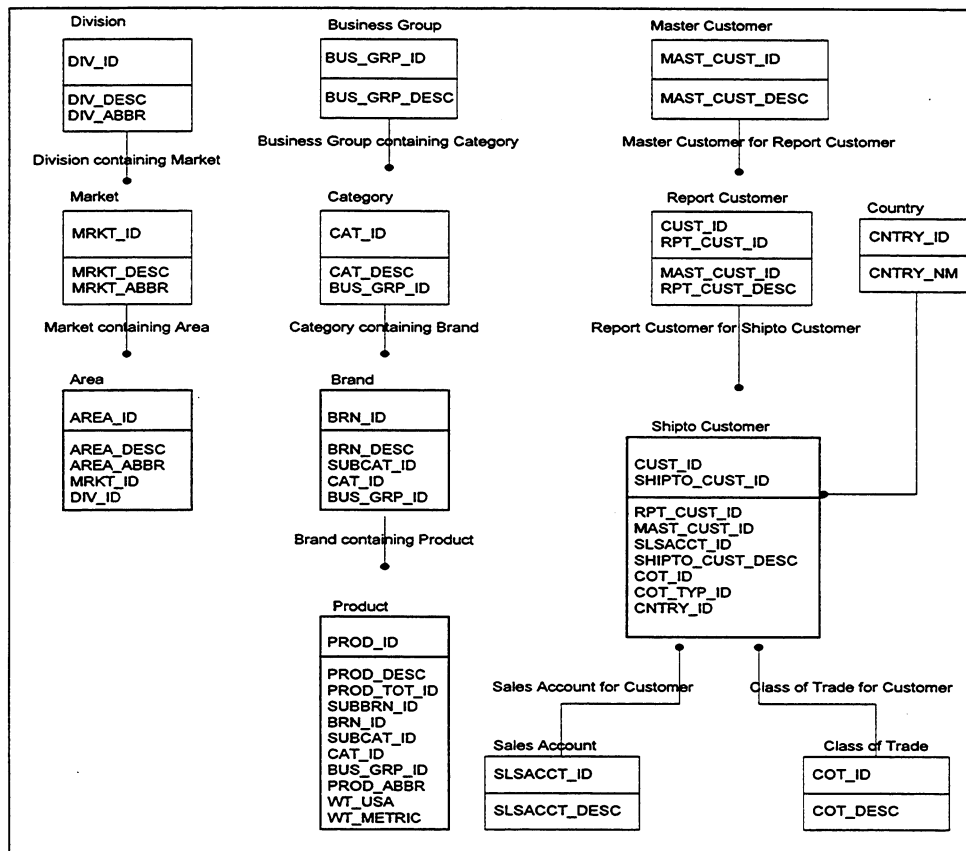
10

Division

DIV_ID

DIV_DESC
DIV_ABBR

Division containing Market

Business Group

BUS_GRP_ID

BUS_GRP_DESC

Business Group containing Category

Master Customer

MAST_CUST_ID

MAST_CUST_DESC

Master Customer for Report Customer

Market

MRKT_ID

MRKT_DESC
MRKT_ABBR

Market containing Area

Category

CAT_ID

CAT_DESC
BUS_GRP_ID

Category containing Brand

Report Customer

CUST_ID
RPT_CUST_ID

MAST_CUST_ID
RPT_CUST_DESC

Report Customer for Shipto Customer

Country

CNTRY_ID

CNTRY_NM

Area

AREA_ID

AREA_DESC
AREA_ABBR
MRKT_ID
DIV_ID

Brand

BRN_ID

BRN_DESC
SUBCAT_ID
CAT_ID
BUS_GRP_ID

Brand containing Product

Shipto Customer

CUST_ID
SHIPTO_CUST_ID

RPT_CUST_ID
MAST_CUST_ID
SLSACCT_ID
SHIPTO_CUST_DESC
COT_ID
COT_TYP_ID
CNTRY_ID

Product

PROD_ID

PROD_DESC
PROD_TOT_ID
SUBBRN_ID
BRN_ID
SUBCAT_ID
CAT_ID
BUS_GRP_ID
PROD_ABBR
WT_USA
WT_METRIC

Sales Account for Customer

Sales Account

SLSACCT_ID

SLSACCT_DESC

Class of Trade for Customer

Class of Trade

COT_ID

COT_DESC

Figure 3b: A Star Schema for Shipment (Continued)

7. Key TDQM Tasks

Two key tasks are being performed as part of SCJ's on-going TDQM institutionalization effort: (1) IQ auditing of data to be loaded into the data warehouse and (2) IQ assessments by data warehouse participants periodically. IQ auditing is performed, in part, by using Integrity Analyzer [1, 6] against extract files before these extracts are loaded into the global data warehouse; IQ assessments are conducted using the IQA instrument.

IQ Auditing

The Integrity Analyzer is used against extract files to ensure their data integrity. When errors are found, corrections are made in the source system that includes source data and source processes. Erroneous data are not loaded into SCJ's global data warehouse until resolutions are identified. The Integrity Analyzer is also used against source reference files. The results of using

Integrity Analyzer include the followings: duplicate identifiers, inconsistent business logic, and incorrect categorization, as presented earlier.

Since the Integrity Analyzer provides a project file that keeps track of the data integrity rules specified by the IQ manager, each of these specified rules is referred to as an *audit item*. These audit items, in turn, are used as measures for tracking information quality, providing a set of IQ metric or index. As such, the Integrity Analyzer facilitates the development of an SCJ IQ Index, and the continuous monitoring of the IQ index over time.

SCJ is also poised to apply the Integrity Analyzer for data integrity check of the key databases within the organization. The use of Integrity Analyzer is to develop base measures for the current accuracy of key characteristics of the data, to develop benchmarks to be used in continuous improvement efforts and to prioritize the areas for IQ improvement initiatives. For example, SCJ have used the Integrity Analyzer to analyze the data concerning global purchases of raw materials for the last three years and currently are analyzing shipment information in North America. Using the User-Defined Integrity facility, SCJ has found certain number of instances, as shown in Figure 4, has negative number of cases booked (e.g., -17 cases of Glade were booked for the week of April 1, 1998) and the dollar amount of revenue recorded in the financial book is zero.

Results such as these have prompted the IQ manager to investigate the root causes of the problem. In this simple case, it was determined that the negative number was a result of accounting adjustment. One benefit of using the Integrity analyzer is that the results are strictly based on the facts in the underlying databases that are repeatable. As one joked, "Now we can do fact-based whining."

User-Defined Integrity - Records Meeting Condition

Change Fields

CASE_BK	DOL_BK	AST_CUST_ID	PROD_ID	WK_ID
-17	0	69000	90095	199841
-7	0	69000	90096	199841
-6	0	155450	1943	199738
-1	0	348000	90097	199643
-12	0	465600	90095	199633
-49	0	541800	90095	199710
-2	0	541800	90097	199645
-1	0	562800	90917	199815
-4	0	562800	90917	199821
-5	0	607500	90097	199648
-42	0	763800	90095	199745
-12	0	763800	12725	199741
-12	0	763800	12725	199744
-10	0	763800	32726	199744
-7	0	763800	52726	199744
-2	0	786028	2095	199709
-28	0	805500	90096	199650
-28	0	805500	90097	199650
-1	0	960432	90095	199639
-1	0	960432	90096	199639
-2	0	960667	90917	199812
-12	0	960706	90917	199820
-57	0	999999	90095	199737
-1	0	999999	90095	199833
-1	0	999999	90097	199651
-1	0	999999	90917	199819

Figure 4: IA Detected a Negative Number of Cases

IQ Assessment (IQA)

A process is now in place to expand the number of employees taking the IQA and to do it in a manner that matches the perceptions of consumers, providers, and custodians of key information [9]. The new study is concentrated in the area of the quality of information that they maintain concerning their products and the customer orders that they process. This involves giving the IQA instrument to functional personnel in the customer service and order processing areas as well as employees in the supporting Information technology areas. The intent is to develop a process that is repeatable and provides information about the perceptions of employees about information quality over a long period of time. This will be one of the foundations for improving the quality of information used within the company and will provide a foundation for better serving their

customers. It is important to determine perceptions of IQ in Organization because in many cases perception is the reality. After all, a data set that nobody uses should be considered as of poor quality no matter how accurate it is.

SCJ is also developing a process to repeat the assessments on a periodic basis to provide one IQ measure within the company. The IQA survey results also help to Identify areas for improvement. SCJ plans to conduct the IQA survey periodically and upon completion of major IQ initiatives.

To facilitate the process, a diskette version of the IQA instrument has been deployed to collect data with success. Using a diskette that the subject can answer via a desktop/laptop PC has proven to be much more effective than that of a paper instrument. The diskette version has enabled SCJ to reduce the overall assessment time (from giving the instruction, hand out the survey, administer the survey process, enter data into the IQA database, to complete the analysis) from a couple of months to within a week.

Because of the anticipation of a firm-wide participation in the IQ assessment, an Internet-based IQA instrument has been developed. Hosted in an NT 4.0 server that runs an MS/SQL server, the Web IQA allows a subject to assess the quality of data for a specific IQ project via the Internet. It also allows an IQ manger to set up login IDs and passwords for the subjects participating in the IQ assessment study. As one of the SCJ managers commented, "We did think this was a good addition to the capability for those people who have access to an Internet browser. Not every person in the organization still has such access although it is increasing rapidly with the proliferation of Intranets."

8. Managing Competing IQ Requirements

IQ managers often face the need to make high-level decisions regarding the competing needs across IQ dimensions and functional needs. For example, for a set of data to be entered to warehouse, business at times requires timely data that may not be 100% accurate but 100% timely and complete, despite the IQ managers efforts to strive to make the accuracy 100%. It is important to

understand the priority among the dimensions and make it sure to have strategies for reviewing this issue.

Another example for timeliness concerns the differing needs across functional areas. Timely data for sales department entails data available for sales managers within 24 hours after sale transaction is executed. Timely data for finance department means data available for finance managers before the 5th working day of every month. For the same set of data, different functional areas require different time frame of timeliness. IQ managers must aware of the relevant priorities and scales of IQ that is based on business needs (short-term results, at times), and must act accordingly without jeopardizing the on-going efforts for long-term IQ improvement. This juggling situation needs to be addressed in advance and one must develop appropriate IQ strategies. Typical strategies include source-tagging, monitoring adjustments, and developing supplemental IQ index.

9. Starting a Long Journey

It cannot be over-emphasized that there is no magic bullet in the endeavor to attain a high-level of information quality in organizations. The S.C. Johnson Wax (SCJ) case presents a concrete example of the commitment management must make and the time horizon that the management must allow for a total data quality management program (TDQM) to bear fruit. As a senior manager quipped, "We have just started a long journey!"

There are many tasks that must be performed in institutionalizing SCJ's TDQM program. In continuing on the information quality audit and information quality assessment initiatives that SCJ is undertaking, SCJ must also develop an SCJ IQ Index for monitoring the progress of SCJ's TDQM program over the long term. The SCJ Index could come at two levels. At the higher level for senior management, a single IQ measure (similar to that of the Dow Jones Industrial Average) or a simple set of easy-to-understand measures will be useful. At the lower level, SCJ is developing IQ measures along the dimensions developed at

MIT TDQM Research Program. Together, these IQ measures would provide a systematical approach for monitoring IQ at SCJ.

Another major task that SCJ is working on is how to manage their information asset as products, and offer these information products to both external and internal customers. Furthermore, SCJ must balance the delicate interplay among technical, organizational, and strategic/business perspectives in their institutionalization of a long-term TDQM program.

It is often said that the 21st century will witness only two kinds of companies: those that exploit IT and those that are out of business. But exploiting IT for what purpose? *To deliver quality information.* Until firms treat information as product and manage it accordingly, quality information will not be delivered to consumers consistently and reliably. The results will be lower margins, missed opportunities, and tarnished images. Much is at stake. In an increasingly networked world, information of different quality levels is being aggregated for business use. In this new environment, information quality must be a paramount concern. SCJ is making steady progress in delivering quality information to their internal and external customers.

References

[1]   CRG, *Integrity Analyzer: Getting Started*. Cambridge Research Group, Cambridge, MA, 1998.

[2]   Funk, J. D. , *MIT Summer TDQM Course Lecture Notes* (No. TDQM-97-1556s-SCJ). Massachusetts Institute of Technology, 1998.

[3]   Huang, K. T., Y. W. Lee and R. Y. Wang, *Quality Information and Knowledge*. Prentice Hall, Englewood Cliffs, NJ, 1998.

[4]   Johnson, S. C., *The Essence of a Family Enterprise: Doing Business the Johnson Way*. The Curtis Publishing company, Indianapolis, Indiana, 1988.

[5]   Lee, Y. W. , *IQ Report for S.C. Johnson Wax* (No. CRG-97-11). Cambridge Research Group, 1997.

[6]   Lee, Y. W., D. M. Strong, L. Pipino and R. Y. Wang , *A Methodology-based Software Tool for Data Quality Management* (No. TDQM-97-02). MIT TDQM Research Program, 1997.

[7]   Madnick, S. and R. Y. Wang , *Introduction to Total Data Quality Management (TDQM) Research Program* (No. TDQM-92-01). Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management, 1992.

[8]   Rob, P. and C. Coronel, *Database Systems: Design, Implementation, and Management*. 3rd ed. Course Technology, Boston, 1997.

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

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