Total Data Quality Management: The Case of IRI

Rita Kovac	Yang W. Lee	Leo L. Pipino
Information Resources Incorporated	Cambridge Research Group	University of Massachusetts Lowell
Rita.Kovac@infores.com	ylee@crgi.com	pipinol@woods.uml.edu

Abstract

Implementing a Total Data Quality Management (TDQM) program is not a trivial undertaking. Two key steps are (1) to clearly define what an organization means by data quality and (2) to develop metrics that measure data quality dimensions and that are linked to the organization's goals and objectives. This paper presents the case of Information Resources, Inc., which exemplifies how a company can develop a viable TDQM program.

1. Introduction

That quality of data is critical to organizations is a truism. Implementing a Total Data Quality Management (TDQM) program to achieve a state of high data quality, however, is not a trivial undertaking. Organizations from different industries, with disparate goals, and operating in dissimilar environments will each develop their own, specific, custom TDQM program. Regardless of differences, the successful implementation of a viable TDQM will consist of the iterative process of defining, measuring, analyzing, and improving. Within this framework the organization must:

(1) clearly define what the organization means by quality in general and data quality in particular.

(2) develop a set of metrics that measure the important dimensions of data quality for the organization and that can be linked to the organization's general goals and objectives.

The experience of Information Resources, Inc. presents an excellent example of what must be done to properly develop a viable TDQM program. Although, it represents the approach and experience of only one company, the principles used and the lessons learned will prove valuable to any firm contemplating the launch of a TDQM program. In this paper, we use the case of IRI to illustrate a specific instance of a firm addressing the two needs above and developing a viable TDQM program.

2. Industry and Company background

Information Resources, Inc. (IRI) is the leading provider of business solutions based on electronic Point of Sale (POS) purchase data to the world wide Consumer Package Goods (CPG) industry. The company creates and maintains proprietary databases and analytical software designed for this industry. The focus of IRI's business solutions is on the functional areas of sales, marketing, supply chain and retail operations. Figure 1 illustrates IRI's position in the process of delivering data to the customer.

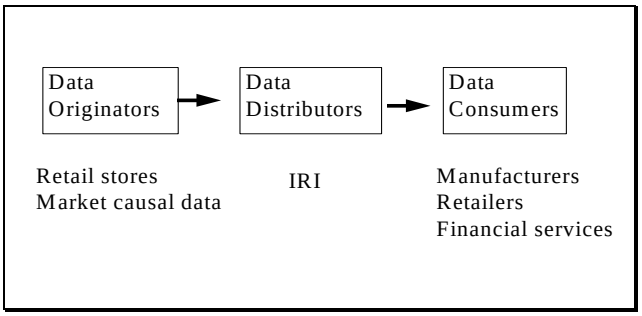

Figure 1: The Delivery Process

IRI has experienced rapid revenue growth due to market driven products. The company has a strong emphasis on innovation and product development. Internal resources are externally focused: meeting the emerging needs of customers. Despite this external focus, IRI has a strong historical track record in investing in quality initiatives for internal processes. This includes investing in scanners for data collection, development of applications in the technology of Artificial Intelligence and Expert Systems to identify and correct data inconsistencies, and continued improvement of sampling and projection systems to better estimate the retail universe.

During the early 1990's, expectations of CPG clients, including both manufacturers and retailers, began to change. Clients began to demand more complicated data delivery, with reduced cycle times. IRI products were now being used for 'mission critical' functions within sales, marketing, logistics and production planning. The net result was increased demand for quality and reliability. This required a shift from project based quality efforts to a total quality management program.

This shift to a total quality management program, shown in Figure 2, incorporated three integrated components: Technology, Work Process, and People. Within the technology components, a comprehensive re-engineering program was initiated. The

goal of this program, referred to as Project OMEGA, was to move from mainframe processing of data to RISC-based production. Expected benefits included lower data processing costs, greater flexibility and improved quality through automation of manual procedures.

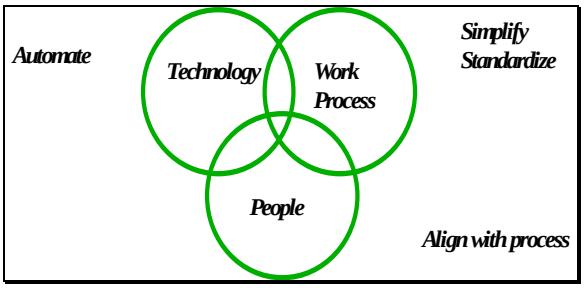

Figure 2: Total Quality Management Program

Within the work process area, standard Work Process Change techniques were used. The goal was to simplify tasks, and then standardize best practices. A full-time, on-going team dedicated to Total Quality Management was formed, reporting to the Chief Executive Officer. This team consisted of senior executives from the operations/production areas and the sales/servicing organization. The Quality Team worked closely with cross functional experts to ensure that all elements of the delivery chain were evaluated.

Finally, results of the technology changes and the Work Process Change efforts were incorporated into organizational re-engineering. Specifically, the existing organization required realignment to meet the new streamlined production process. An outside consulting firm was hired to help with the facilitation and implementation of changes. This effort was referred to as Project ImPACT (Figure 3).

Project ImPACT was an aggressive undertaking. A substantial commitment, in terms of both internal staff time and consulting fees, was made to this project. To support this investment, aggressive goals were clearly defined:

- Eliminate 80% of IRI induced errors
- Reduce average assembly and delivery time of client deliveries to five days after raw data loaded
- Eliminate re-work and re-runs

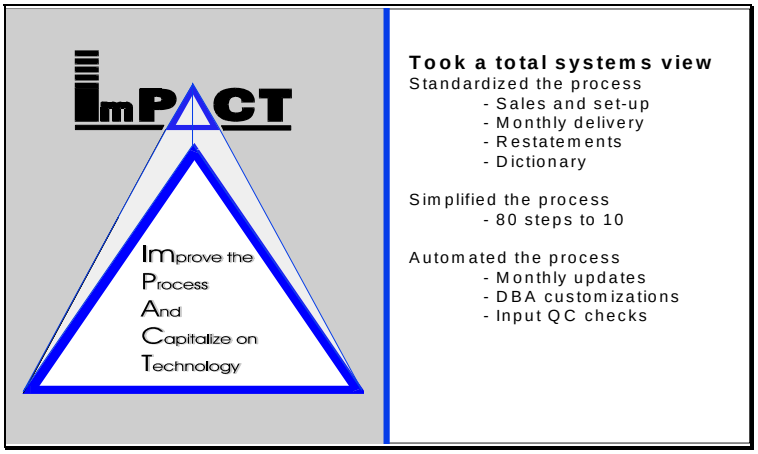

Figure 3: Project ImPACT

Expected benefits and timing were also clearly defined to both senior management and to all groups involved in the delivery process. A summary of objectives and associated benefits is presented in Figure 4.

	Objectives			Benefits			
Solution Areas	Reduce Data Errors	Eliminate Reruns & Rework	Increase Speed & Consist-ency of Delivery	Improve Client Satisfac-tion	Increase Margin	Reduce CPU Usage	Reduce Costs
New Account Sales & Ad Hoc Sales			√	√	√		√
Monthly Delivery	√	√	√	√		√	√
Dictionary	√	√	√	√		√	√
Setup, Renewals, & Restate-ments	√	√	√	√	√	√	√

Figure 4: Objectives and Benefits

3. Basic Framework

Implementation of a Total Data Quality Management program first requires a definition of quality. Each company must choose a definition that is appropriate to its goals, its industry, and its internal culture. Wang et al. [3] have argued that information should be treated as a product - a product delivered to a consumer. This perspective emphasizes the customer and the manufacturing process that produces the information for

the customer. The working definition of quality at IRI emphasizes this importance of customers and partners: IRI defined *Quality* as *conformance to legitimate customer requirements*. This definition requires that we understand what our customers need and what is legitimate.

In understanding customer requirements, IRI uses a hierarchical model, called the Customer Hierarchy of Needs. This model was originally developed with a focus on external customers but is now also applied to internal customer/supplier relationships within the IRI delivery chain. It uses the framework developed by Wang and Strong [4].

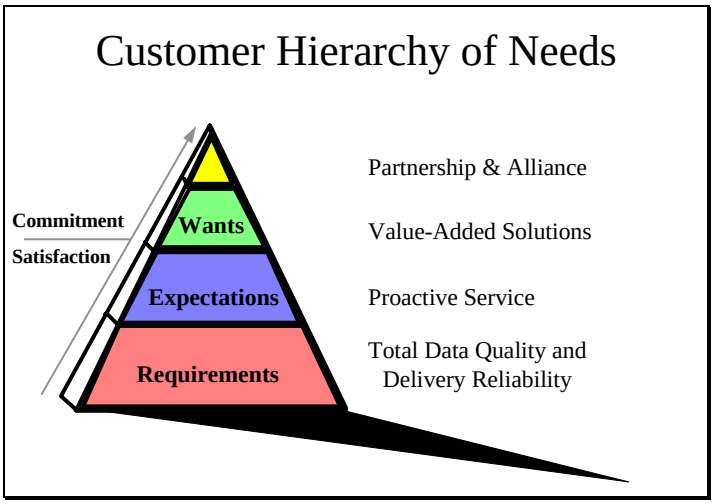

Figure 5: Customer Hierarchy of Needs

This model stresses that Total Data quality is a necessary condition and the foundation of any customer/supplier relationship. Highlighted within IRI's Vision Statement is the following: *"Recognizing that clients use our services in critical decision-making processes, we are committed to fully meeting their expectations of timeliness, reliability and accuracy."*

Movement up the hierarchy to committed partnerships is not feasible unless this basic foundation has been laid. Obviously, understanding the customer's requirements becomes a critical path step towards the goal of creating a committed partnership between customers and suppliers. A process for collecting and organizing customer requirements is discussed later in this paper.

These requirements must, however, be legitimate. That is, a requirement must be feasible to accomplish and it must be valued by clients at a profitable price for the supplier. Both customer and supplier must achieve an attractive Return on Investment.

As pointed out earlier the Hierarchy of Needs model is based on a series of comprehensive empirical studies [1, 2, 4], which led to a taxonomy with four information quality (IQ) categories (Table 1). Intrinsic IQ denotes that information has quality in its own right. *Accuracy* is merely one of the four dimensions underlying this category. Contextual IQ highlights the requirement that information quality must be considered within the context of the task at hand; i.e., information must be relevant, timely, complete, and appropriate in terms of amount so as to add value. Representational IQ and accessibility IQ emphasize the importance of the role of systems; i.e., the system must be accessible but secure, and the system must present information in such a way that it is interpretable, easy to understand, and represented concisely and consistently.

Table 1: IQ Categories and Dimensions

IQ Category	IQ Dimensions
Intrinsic IQ	Accuracy, Objectivity, Believability, Reputation
Contextual IQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of information
Representational IQ	Interpretability, Ease of understanding, Concise representation, Consistent representation
Accessibility IQ	Access, Security

IRI has made effective use of this framework. In conjunction with understanding customer requirements, IRI implemented a formal Work Process Change process to improve the current process and organization and to develop metrics to evaluate the process from the customer's perspective.

The first step was to build a team of functional experts. First, functional experts were identified and commitments of time and effort were agreed to. Next, a formal charter document for this expert team was written and agreed to by all team members. This document clearly specified the current state, the desired end state, the benefits, the project scope of what is and is not included, project output and a timetable of effort required.

The second critical step in this integrated process was to understand the customer requirements. Previous work was done on external customer requirements, using the structured methodology described earlier. The focus at this stage was on internal customer/supplier relationships. Internal customer/supplier relationships were first identified. Then, each customer/supplier team had to agree to legitimate customer

requirements and to develop objective measurements of customer requirements. Linking back to the corporate vision statement and to external customer needs, these measurements were focused on database accuracy and delivery timeliness.

4. Development of Metrics

The two fundamental variables or data quality dimensions to be measured were database accuracy and delivery timeliness. A multiple set of metrics were developed to assess these dimensions. The metrics linked to the corporate vision statement and to the external customer needs. To identify the customer/supplier relationships and to define metrics, an iterative process named RUMBA was developed. The process was based on five criteria: is the metric Reasonable, Understandable, Measurable, Believable, Achievable (Table 2).

Table 2: Criteria for Metric Assessment - Rumba

R	**Reasonable**
U	**Understandable**
M	**Measurable**
B	**Believable**
A	**Achievable**

Functional experts first used RUMBA to identify customer and supplier relationships within the internal delivery chain (Table 3).

Table 3: Customer and Supplier Relationships

Supplier	Customer 1	Customer 2	Customer 3	Customer 4
Retail Data Acquisition	Field Data Collection	Sampling & Projection	Data Loading	
Field Data Collection	Retail Data Acquisition	Sampling & Projection	Data Loading	Item Identification
Sampling & Projection	Data Loading	CDS		

For each of the customer/supplier relationships defined, metrics were identified. Consistent with the corporate vision and the external client requirements, these metrics were developed to address timeliness of delivery and accuracy of delivery. Table 4 illustrates the metrics developed and agreed to by some of the customer/supplier teams within the delivery process.

Table 4: Metrics for Data Quality Dimensions

Supplier	Customer 1	Customer 2	Customer 3
Retail Data Acquisition	**Field** Accuracy: number of store authorization letters outside of schedule Timeliness: no separate measures	**Sampling & Projection** Accuracy: number of stores signed up relative to number authorized to request Timeliness 1) Number of stores add to sample vs. goal. 2) Required support information received to on agreed to schedule	**Data Loading** Accuracy % of stores with data problems Timeliness % of stores received on agreed to schedule

Once these customer and supplier relationships were identified, a high level, total process map was developed (Figure 6). The process ranges from receipt of raw material (retailer store movement data) to delivery to client. Groups responsible for each step of the process were identified. Particular focus was placed on hand-offs or transitions across functional groups. These represented the internal customer supplier relationships. It is critical that specific metrics be developed to measure the hand-off from one group to the next in the internal delivery chain.

This simple process map has had several concrete benefits. First, it has been used successfully with external clients to demonstrate the complexity of client deliverables. Second, this process map has been used internally to educate all members of the delivery chain on the process required to deliver quality databases to external clients.

5. Collection and Reporting of Data Quality Metrics

Three levels of metrics were used: (1) process indicators, (2) quality indicators, and (3) total system measures. These are shown in Table 5. Based on the RUMBA process, each customer/supplier team within the delivery process defined and then collected and reported objective metrics on performance relative to timeliness of delivery and accuracy of delivery. A sample of results is shown in Table 6.

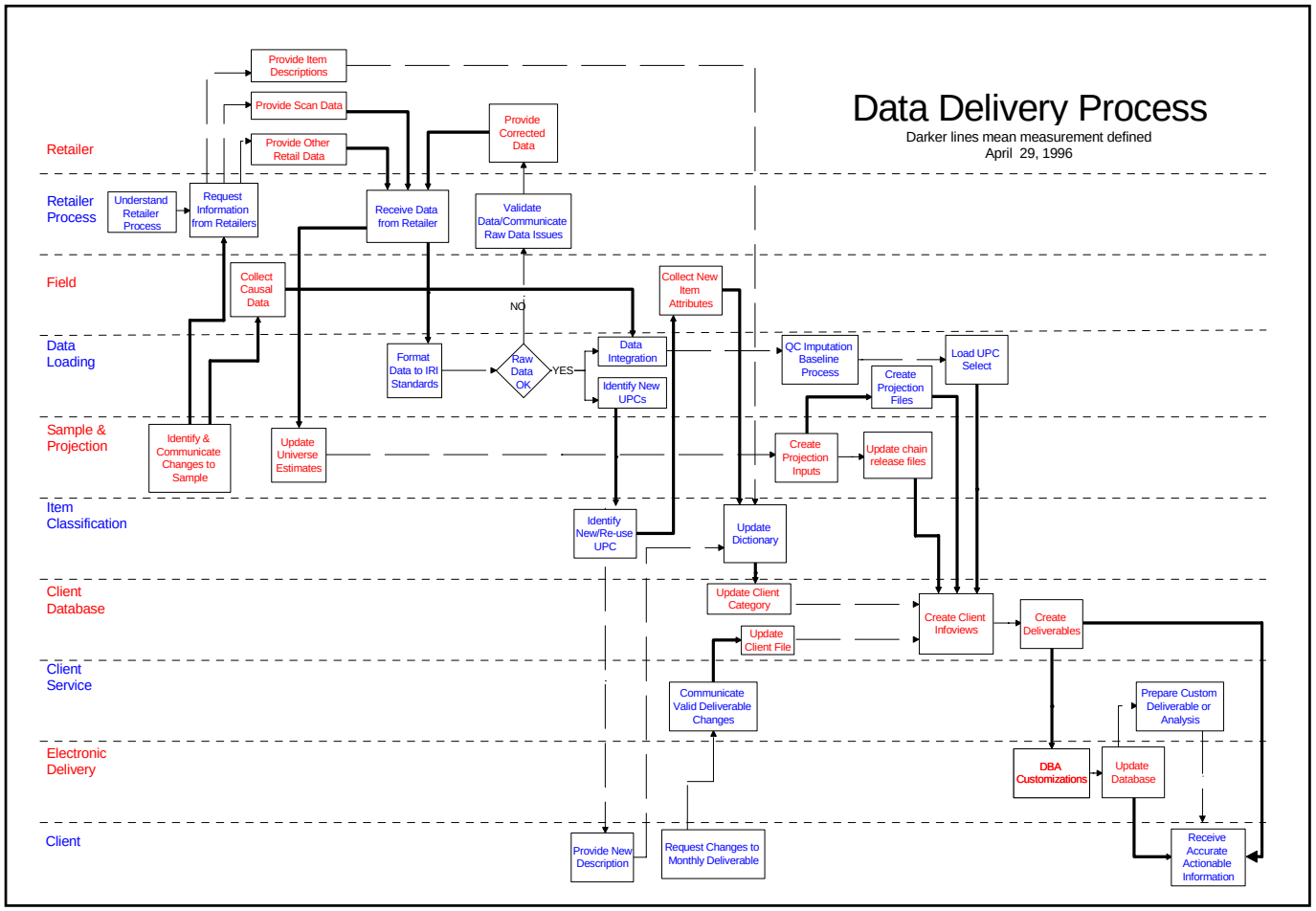

Figure 6: Data Delivery Process Map

Table 5: Data Quality Metrics

Type of Metric	Description
Process Indicators	• Used within department or process step • Diagnostic or early warning system; time available for corrective action
Quality Indicators	• Department or process step end product • Focus on hand-off of supplier to customer • Primarily report card for process • Some measures weekly, reported by period
Total System Measures	• Client perspective • Total delivery chain • IRI report card & diagnostic • Reported by IRI period

Table 6: Example Results

1996 RUMBA Results for Sampling & Projection				
SERVICE	**Qtr. 1**	**Qtr. 2**	**Qtr. 3**	**Qtr. 4**
	1996	**1996**	**1996**	**1996**
TOTAL SCORE				
- Inputs Timeliness	99	97	99	99
- Inputs Accuracy	85	96	91	97
- Key Acct Availability	100	100	100	100
RUMBA RESULTS				
- Avg Timely/Accurate	92.3	96.3	95.3	98.2
- Q4 Target:				96.1

In addition to this internal measurement system, a total measurement system was developed to measure performance of the total delivery system from the client's point of view. This system is fully integrated within the production process and is referred to as TRAQ: ***Timeliness + Reliability + Accuracy = Quality***

The TRAQ system has two major objectives. First, it must provide objective, consistent measurement of data quality and delivery reliability. It reflects the total delivery process and the external client view of delivery performance. This allows management to evaluate performance for a specific client or group of clients such as sales region.

The second objective of TRAQ is more important from a total quality management perspective. This second objective is to provide continuous improvement to the delivery process. The system is designed to provide continuous improvement to the delivery process. This requires specification of where the problem occurred and why it occurred. This information on process failure and root causes allows functional experts to develop solutions that will prevent future occurrences of this problem.

Some benefits of TRAQ include improving client confidence, prioritization of resources to the most critical problems, improvement of margins through reduction of re-work, increased revenue as basic client hierarchy of needs met, and simplification of job tasks and responsibilities.

The TRAQ development process was based on Work Process Change techniques. Six simple phases were completed:

- *Document* delivery process
- *Track* problems
- *Link* problem to process
- Identify *root cause*
- Develop and implement *solutions*
- *Report* performance

To track and identify problems, a custom application was initially built using Microsoft Access. This application currently uses an ORACLE database with a Visual Basic front end. A simple, easy to use system of menus and screens has been developed to log and classify all problems and their associated root causes and solutions. Some sample screens are shown in Figures 7-8.

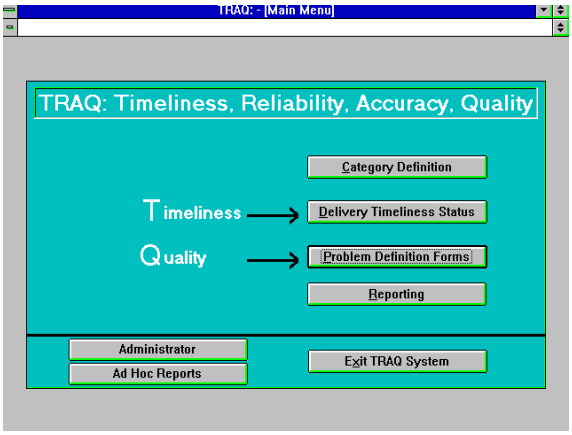

Figure 7: TRAQ Menu

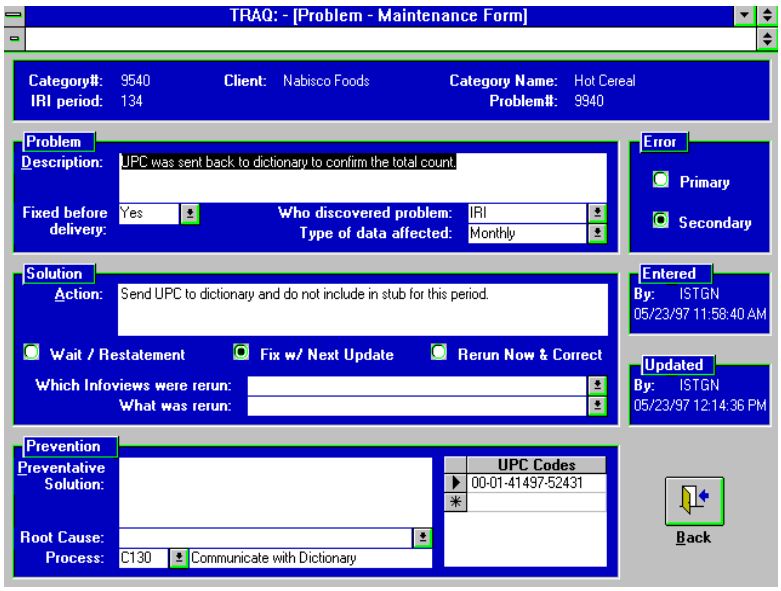

Figure 8: TRAQ - Problem Definition

After each delivery period, each account team updates TRAQ to include information on the description and causes of any errors and the status of delivery relative to timeliness. This information is then used in a series of reports described in the next section.

4. TRAQ Collection and Reporting of Metrics

The overall process for collecting and reporting of TRAQ information is described in Figure 9.

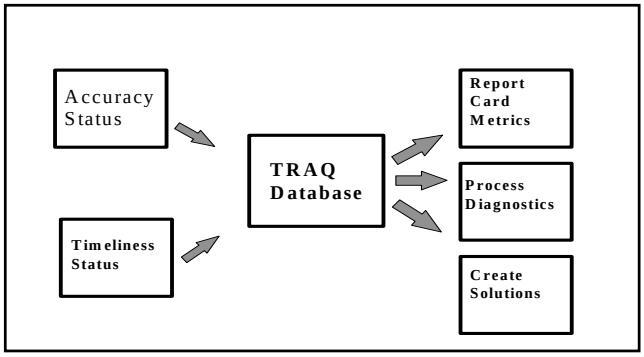

Figure 9: TRAQ Data Collection and Reporting

There are two components to the TRAQ scores reported. First, was the database delivered on time? If yes, the database score is 100, if not, the score is 0. Second, was the database accurate and complete? To measure this, an accuracy score based on the number and severity of each error is considered. The goal is to reflect the external client view of delivery accuracy. The accuracy score starts at 100 (no errors) and declines based on the following factors (Figure 10).

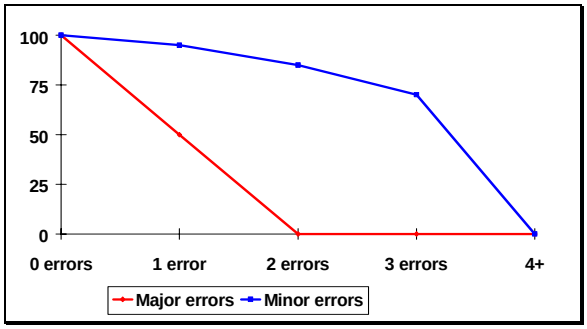

Figure 10: TRAQ Error Scores

TRAQ scores are routinely used as a report card on basic elements of delivery performance. This includes review of the number of errors by sales region, by client and by client deliverable. Information on the number of errors and the accuracy and timeliness scores described earlier are provided each month (Tables 7-8).

Table 7: Example of TRAQ scores: Review across time

Period	Client Visible	Fixed	Major	Minor	Accuracy Score	On-time Score
132	44	68	36	76	92	97
133	63	62	44	81	89	96
134	32	70	22	80	94	96

Table 8: Example of TRAQ scores: Review across Regions

Region	Client Visible	Fixed	Major	Minor	Accuracy Score	On-time Score
Central	34	68	21	81	94	97
Cincinnati	13	49	14	48	92	98
NE	13	17	9	21	93	97
NJN	8	18	2	24	99	95
NJS	11	54	16	49	97	93
West	43	33	39	37	87	94

In addition to the report card metrics provided for sales regions, clients and categories, diagnostic information is also provided. This reporting is designed to address the second stated objective for the TRAQ system: to provide feedback for continuous improvement to prevent problems from re-occurring. This information is reviewed by senior management as well as functional experts responsible for each step of the delivery process. A sample graph is shown in Figure 11.

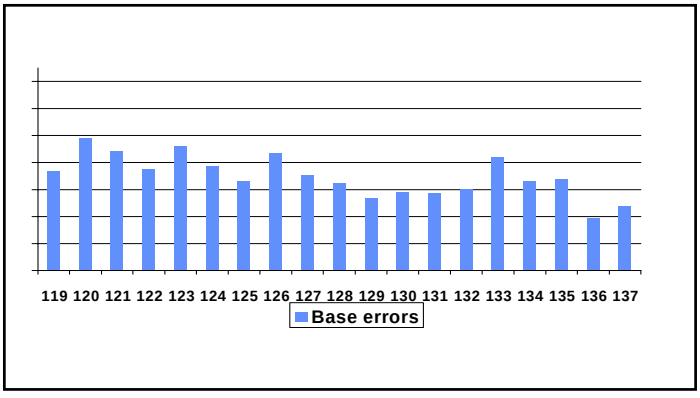

Figure 11: Example of Senior Management Graphs

Error trend data from TRAQ system is used to report back to clients when, where and why errors occurred. This open interaction helps create the foundation for partnership and alliance described within the Customer Hierarchy of Needs. Figure 12 is an example of error trend information provided to a client.

TRAQ reports also include information on which process areas are most error-prone. Functional experts review this information to identify the most error-prone processes and then to prioritize internal development resources towards solving these problems. In the chart shown in Figure 13, it is apparent that six processes are contributing about 80% of total errors. Based on this information, internal development resources were focused initially on this six process areas.

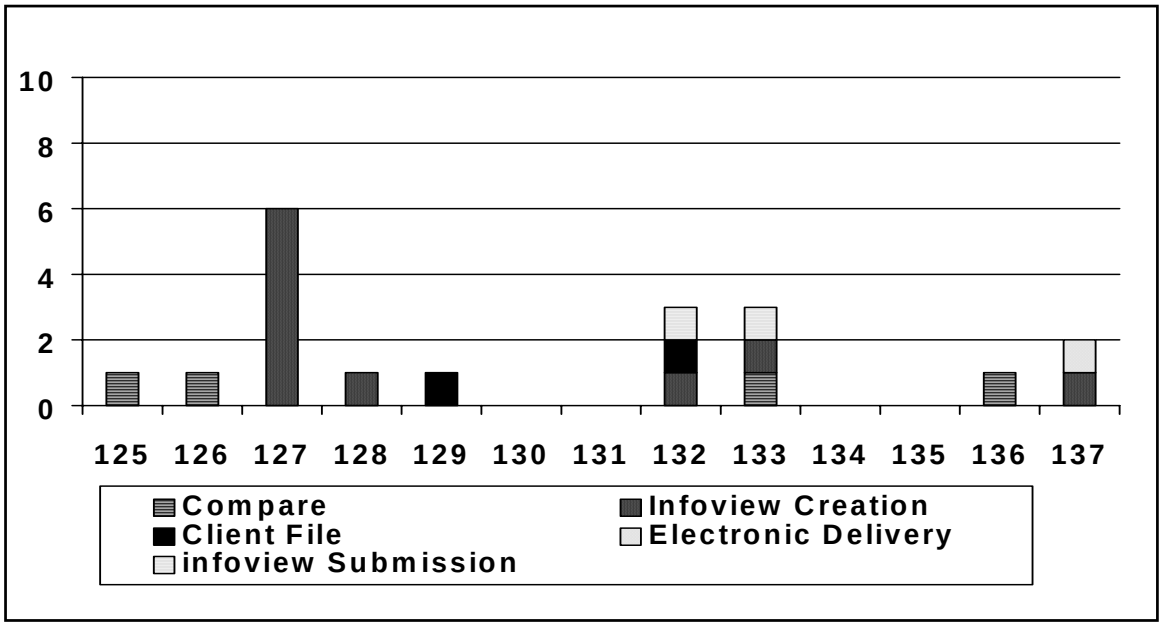

Figure 12: Client Error Trend Information

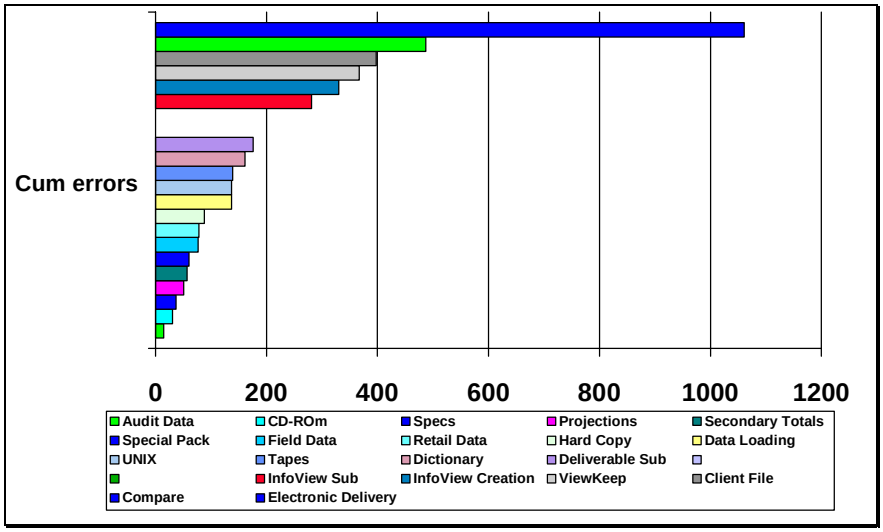

Figure 13: Source of Errors

Two types of solution response were made based on this information on the source and frequency of errors. First, there was a local area, or iterative response to error elimination. These solutions tended to be smaller in scope, confined within one area, and shorter term to implement. A trend of error levels within a specific process area shows the positive impact of this type of iterative response (Figure 14).

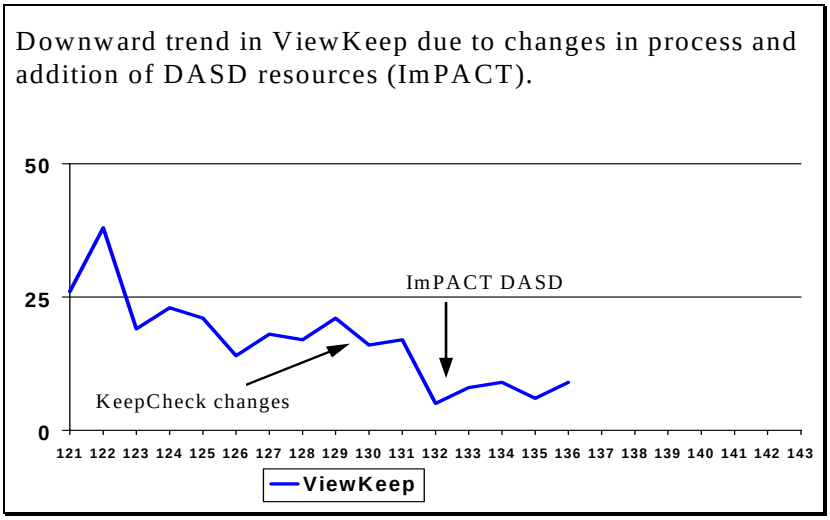

Figure 14: Solutions and Error Level

The second type of solution response focuses on systemic or breakthrough elimination of errors. These solutions tend to be larger scale, cross several organizational or functional boundaries, and require significant investment of internal and external

resources. Project ImPACT efforts would be classified as these longer-term, break through solutions (Figure 15).

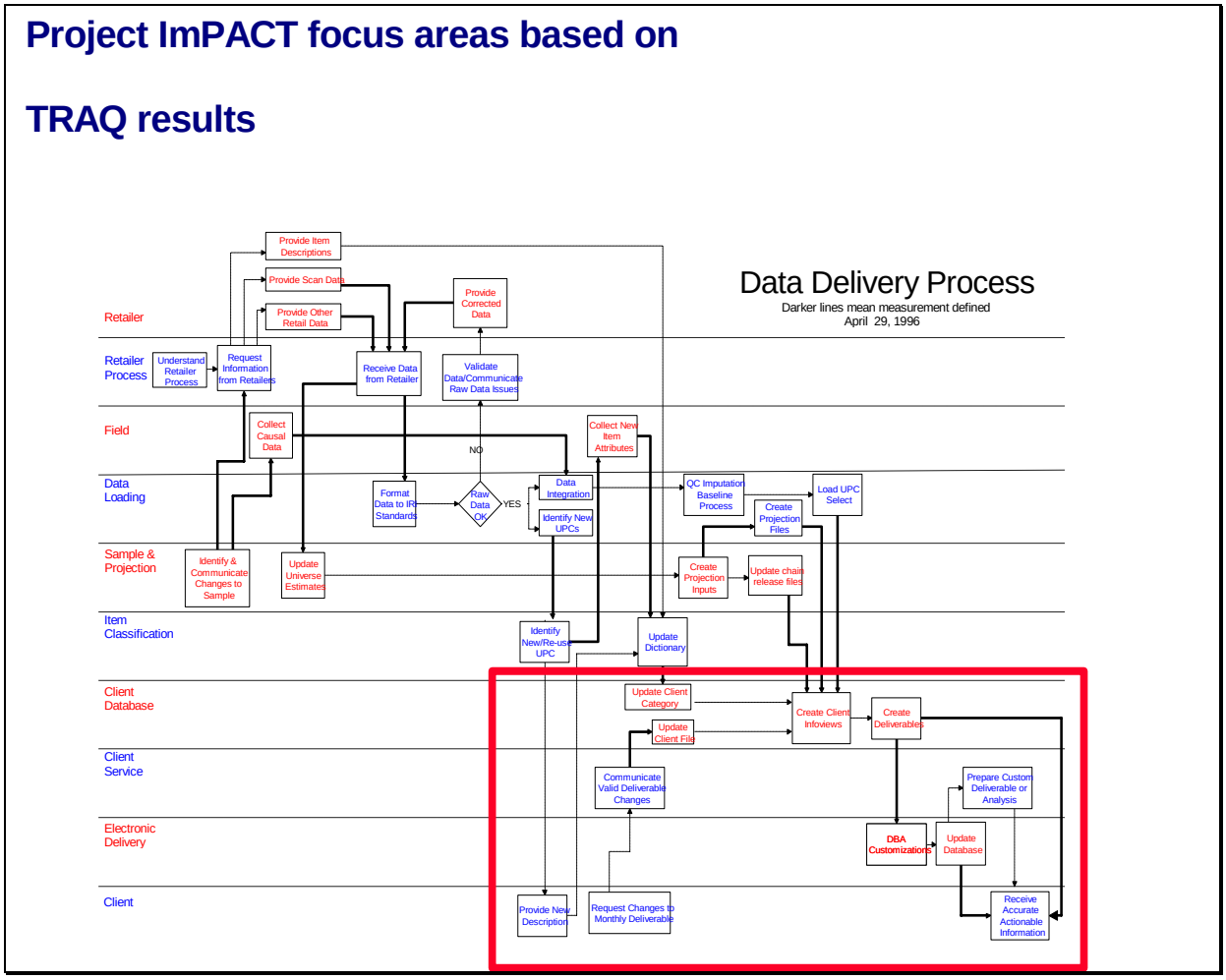

Figure 15: Project ImPACT Focus Areas

5. Conclusion

Implementation of a Total Data Quality Management program has provided substantial benefit to IRI. Specific integration of metrics across all levels and processes within the delivery chain has allowed IRI to: (1) Focus on the customer, (2) Improve quality through process simplification and standardization, and (3) Justify resource allocation for re-engineering efforts.

Without metrics, IRI could not have identified problem processes, prioritized solutions, secured corporate resources of several million dollars for re-engineering projects, or determined the effectiveness of the Total Quality re-engineering efforts.

With metrics consistently reported and evaluated throughout the organization, we have been able to focus on facts, not anecdotes, increase confidence among employees, clients and senior management, and to quantify improvements implemented.

6. References

[1] CRG, *Information Quality Survey: Administrator's Guide.* Cambridge Research Group, Cambridge, MA, 1997.

[2] Strong, D. M., Y. W. Lee and R. Y. Wang, Data Quality in Context. Communications of the ACM, *40*(5) 1997, pp. 103-110.

[3] Wang, R. Y., Y. L. Lee, L. Pipino and D. M. Strong (1997). *Manage Your Information as Product: The Keystone to Quality Information.* (No. TDQM-97-01). Total Data Quality Management (TDQM) Research Program, MIT Sloan School of Management.

[4] 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.