

Organizational Realism Meets Information Quality Idealism: The Challenges of Keeping an Information Quality Initiative Going

Beverly K. Kahn

Suffolk University
Sawyer School of Management
8 Ashburton Place
Boston, MA 02108
Tel: (617) 573-8642
Fax: (617) 573-8345
bkahn@acad.suffolk.edu

Raïssa Katz-Haas

Ingenix
450 Columbus Blvd
Hartford, CT 06115
Tel: (860) 702-8721
rkatzh@uhc.com

Diane M. Strong

Worcester Polytechnic Institute
Management Department
100 Institute Road
Worcester, MA 01609
Tel: (508) 831-5573
Fax: (508) 831-5720
dstrong@wpi.edu

Abstract

While many seem to understand, in theory, the value of high-quality information, the realities of daily organizational dynamics may cause information quality projects to falter. In our conference paper last year, we described how Ingenix moved from ‘theory’ to an established Information Quality initiative, and how it laid the groundwork for its first and future information quality improvement projects. This paper describes the challenges and successes of keeping information quality initiatives going in the presence of development deadlines, the call of daily work activities, organizational dynamics, etc. The lessons learned from Ingenix provide insights for other organizations as they seek to keep information quality initiatives strong and healthy. One of the lessons/insights is that multiple, even dissimilar IQ (*See acronyms, p. 12*) efforts, carried out simultaneously can be as effective as a single overarching initiative.

*This paper is dedicated to the memory of Tim B. Kaufman
Galaxy Architect. Mentor. Friend.*

INTRODUCTION

In our IQ 2000 conference paper, "How to Get an Information Quality Program Started: The Ingenix Approach", we described one company's approach to starting an information quality improvement program [Kahn, Katz-Haas, and Strong, 2000]. While the essential first step of an information quality (IQ) improvement endeavor—getting the project started—is more difficult than one might expect, keeping it going is just as challenging. In this paper, we look at a year in the life of what began as a corporate-sponsored information quality initiative.

What we see are actually several information quality efforts operating simultaneously and interacting with each other. Some of the efforts operate at a local level, others under corporate sponsorship, some as part of information systems development projects, and some as more independent information quality efforts. As is typical of the dynamics of any organization, these efforts start and then may flounder, and then re-gain momentum in a different form—or several forms.

OVERVIEW OF INGENIX

Ingenix is a wholly owned subsidiary of UnitedHealth Group (UHG), a multi-billion dollar diversified health and wellness company headquartered outside Minneapolis. Ingenix is one of the six business segments of UHG. (For more information on UnitedHealth Group, see <http://www.unitedhealthgroup.com>.) Ingenix has become one of the largest health care information and research companies. At the UHG web site, Ingenix is described as follows:

“ Improved knowledge and information are key to achieving improved health and well-being. Access to accurate, unbiased research and information helps improve the effectiveness of care by supporting fact-based clinical and financial decisions. Ingenix is uniquely positioned to fulfill this need.”
(<http://www.unitedhealthgroup.com/ingenix/index.html>)

Ingenix has become one of the largest health care information and research companies in the industry, providing a comprehensive line of products and services—many of which are rooted in proprietary databases. Therefore high quality data is essential to Ingenix's continued growth, and Ingenix is well aware of this. To quote from the CEO of Harvard Pilgrim Health Plan,

“ If you are in the health [care] business, your data better be damn good because you're using that data to make decisions that are going to impact your company six, 12, and 18 months from now. If you don't have full faith in the comprehensiveness and precision of the data you're using, you're taking some huge and unacceptable risks.”

[McCue, 2001]

THE SHARED DATA WAREHOUSE

The Shared Data Warehouse (SDW) is a department within Ingenix's Information Systems (IS) unit. The SDW is made up of six interdependent groups: 1) Information Quality & Strategies (IQS), 2) Business Analysis, 3) Software Engineering, 4) Training and Support, 5) Project Management, and 6) Applications Development.

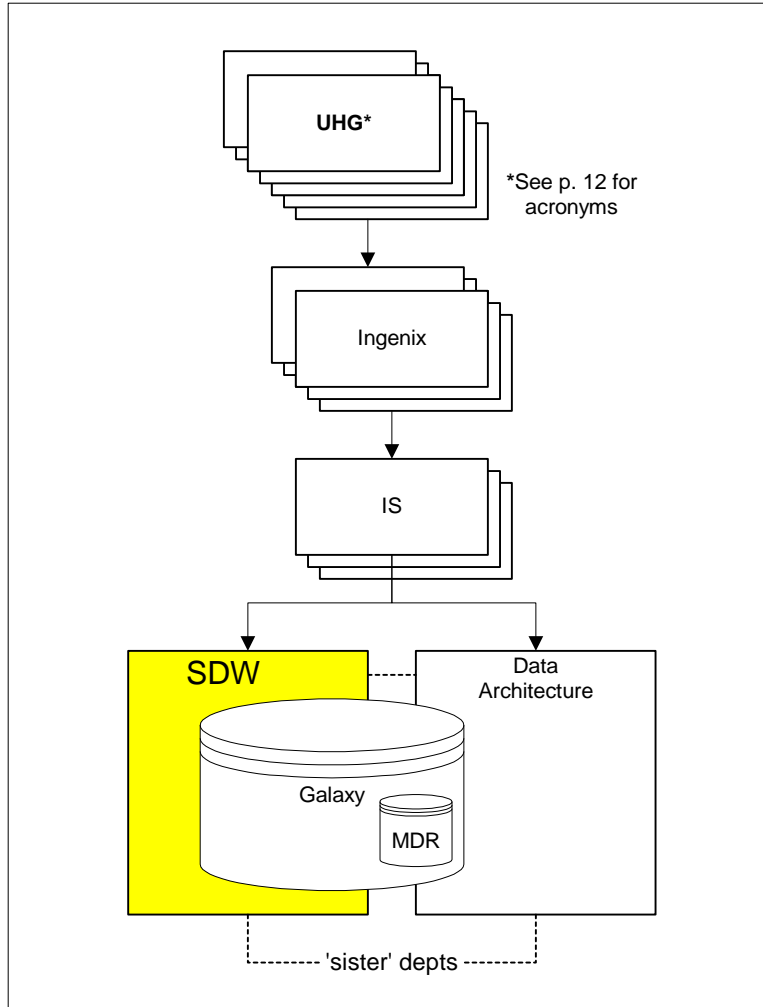


Figure 1 SDW Relationships

The SDW works closely with another IS department (which, up until a few months ago was part of the SDW), the Data Architecture Department. This department includes six groups: 1) Metadata Repository (MDR) group, 2) Data Administration group, 3) Modeling group, 4) Database Administration group, 5) the ETL (extraction, transformation, and load) group, and 6) the Technical Infrastructure group.

In the summer of 1998, UnitedHealth Group charged the SDW with developing a new data warehouse (now known as Galaxy) to integrate data from UnitedHealth Group's two major data stores (MARS and DSS) into one repository. This would make reporting across the enterprise easier for customers/users. As part of the Galaxy charter and to support the needs of its direct and indirect users, the Shared Data Warehouse committed to deliver data of equal or better quality as MARS/DSS data. In addition, Galaxy was purposefully architected to minimize data quality problems. Thus from the beginning of the Galaxy project, Ingenix recognized that information quality issues are an integral part of data warehouse development.

QUALITY DATA SOURCES FOR THE DATA WAREHOUSE

The Galaxy data warehouse is a significant project, which was completed on schedule on August 1, 2001. (Completion means that all the planned subject areas are in the data warehouse and accessible to its users). Each Subject Area team decided what data to include and from which sources. Data quality issues were among the considerations in selecting data and their sources. This enabled subject area teams to make best choices about what sources to use for the warehouse.

GALAXY INFORMATION AND METADATA QUALITY STRATEGY

SDW Management and IQS developed Galaxy's data quality strategy in December of 1999. The strategy included eventual corporate sponsorship, the not yet formed IQPI initiative/group, continuous process improvement (using a Six Sigma methodology [Pande, Neuman, and Cavanaugh, 2000]) and the ultimate goal: prevention of data and metadata quality problems.

Very little cleansing or editing of source data is done for the data warehouse. This is by design. IQS and SDW management decided that the best way to improve the quality of data over the long run was to use most data 'as is' rather than attempting to hide or cover up every data quality problem by cleansing and filtering/editing. This way, root causes of data problems can be located, analyzed, and remedied, preventing future problems. Additionally, an overabundance of cleansing/editing routines degrades warehouse performance, a highly significant data quality issue to customers.

Users need high quality metadata as well as quality data. The metadata repository (MDR) group and the modelers within the Data Architecture department built Galaxy's metadata repository based on business requirements gathered by IQS. IQS and the MDR group work closely to ensure high quality metadata for Galaxy's customers and SDW staff.

The MDR group maintains the metadata and ensures that metadata is accessible by the SDW staff for various purposes. One of these purposes is to allow extraction of appropriate metadata for delivery to Galaxy's customers. The 'web team' (an extension of IQS) publishes the extracted metadata to the web for easy access by users. Below is an example of metadata as the user sees it. This is from a page of the Galaxy Data Dictionary:

The screenshot shows a web interface for a data dictionary. At the top, it displays 'Table Name: Diagnosis Code' and 'DB2 Name: DIAGNOSIS_CODE'. Below this are navigation links for 'Table Descriptions', 'Columns', 'Indexes', and 'Joins', along with a search icon and a 'Dictionary Home' link. A 'Skip to' menu lists letters A, R, G, L, M, U. The main content area is titled 'A' and lists the following details for the 'Diagnosis Code' table:

Business Name: AHRQ Diagnosis Detail Category	Data Type: CHARACTER
Code	Length: 3
DB2 Name: AHRQ_DIAG_DTL_CATGY_CD	
Identifies the AHRQ (Agency for Healthcare Research and Quality) detailed grouping of diagnosis codes that describe a disease (e.g. tuberculosis).	

Figure 2. Meta Data Example

Metadata quality improvement is a core part of building and operating the data warehouse at Ingenix. Its business is to supply healthcare data and information to other segments within UHG and to external customers. For Ingenix to provide a high quality product to its customers, it must ensure both high quality metadata and data.

CORPORATE HEADQUARTERS TAKES ON INFORMATION QUALITY

To dramatically improve information quality in a company, improvement initiatives must go beyond its IS group(s) because IS groups do not 'own' business processes that create, capture, and gather data for data entry, processing, or other manipulation of data. There are many reasons for particular quality problems. For instance, many information quality problems in organizations are introduced at the original source (i.e., first business event/transaction). Others are introduced as a side effect of policy decisions about the business. Thus, corporate-wide information quality initiatives, sponsored or lead by Senior Management, are critical to the success of IQ initiatives.

In our IQ 2000 conference paper, we described the beginnings of the Information Quality Process Improvement (IQPI) initiative, a UHG-wide information quality initiative [Kahn, Katz-Haas, and Strong, 2000]. Planning and discussion of information quality began in earnest mid-1999. Ingenix sponsored an enterprise-wide IQ seminar that took place about a year later in June 2000. From this seminar, an initiative and program group consisting of fifteen members from all major business segments of UHG were formed in July. The name of the group and initiative was IQPI (Information Quality Process Improvement).

At the time of last year's IQ conference in October, this corporate-wide information quality initiative was off to a good start. It had support from the President of UHG, involvement from the appropriate groups, and a list of information quality issues that were the focus of the group's work. Unfortunately, the momentum was not maintained, and the last time the IQPI group met was in December 2000. *However*, consciousness was raised across the enterprise of

the criticality of good quality data. Other IQ efforts emerged as a result of this and of the IQ seminar and IQPI. In addition, the UnitedHealth Group president created a position, which is wholly dedicated to data quality across the company with a focus on improving the quality of the operational databases and their data. The position is high enough in the organization to carry the clout needed to work across segments/divisions. The President filled this position in the summer of 2001.

CORPORATE QUALITY IMPROVEMENT INITIATIVE

Planning for a corporate-wide focus on a ‘TQM-type’ quality and process improvement initiative began at UHG corporate headquarters in January 2000. The plan is to phase in this quality program, which we will call the Quality Excellence (QUE) program (a fictitious name), to all of UHG within three years. The QUE program has several general purposes: 1) support UHG’s business strategy, 2) improve results, 3) exceed customer expectations, and 4) achieve performance excellence.

Many business segments have begun at least one QUE project. Ingenix selected Galaxy’s IQ as one of its QUE projects and provided two sponsors for assistance, if needed. (These sponsors also report project progress to corporate sponsors.)

The QUE initiative is similar to TQM, CPI, and other quality/improvement programs. QUE essentially follows a Six Sigma methodology, including SPC (Statistical Process Control) and the DMAIC model. DMAIC is an essential part of Six Sigma and stands for **Define, Measure, Analyze, Improve, and Control**. Defining what “quality” means, according to the customers, was part of the 2nd quarterly QUE report. For its project, IQS provided the following definition:

“The quality of data = its fitness for use. In addition, the quality of the data meets or exceeds customer requirements.”

MEMBERSHIP DATA QUALITY IMPROVEMENT INITIATIVE

IQS had already been using the DMAIC model in its focus on information quality. Its first project involved looking at, and improving membership data. For many businesses, membership data is often difficult to improve and is an enormous task. Clearly this was not a one-person job. IQS formed a small, focused QUE team consisting of:

- representatives from four business areas (customers)
- two source system representatives
- four staff members from the SDW

Working in small teams is part of the DMAIC methodology. What is a little unusual about this team is that it is cross-functional and must work through a variety of agendas, locations, and schedules. However the advantages of a cross-functional team far outweigh any disadvantages.

Membership data refers to data about the people enrolled in a healthcare plan and who are eligible for certain benefits. Membership data passes through many business processes, paper forms, eligibility systems, claims processing systems, and the all business processes involved in installing, canceling, and re-installing policies and their attendant membership. Improving the reliability of membership data is a huge undertaking due to the number of members, the number of systems (mostly 'stovepipe'), the number and complexity of business processes involved (also often stovepipe), and the interrelationships between all of the above.

To understand why an accurate member count is difficult, consider the following:

- When policies are cancelled, members' claims, which were incurred before policy termination typically remain eligible for a period of time called 'runout'
- When a member dies, coverage may continue for the spouse. A business decision was made that, out of consideration for the spouse, coverage would continue under the social security number of the deceased. As a result, there remains a record in the data warehouse for the deceased, as well as a record for the surviving spouse.
- Contract cancellation dates that are not in proper relationship to service dates cause innumerable problems. These problems (i.e., symptoms) are fairly easy to spot, but finding their root causes is complex and difficult.
- Most significantly, membership data are housed redundantly in numerous systems across the enterprise (See Figure 3.)

These are the types of issues that can skew member counts, causing departments to develop expensive workarounds among other consequences.

The QUE team started with a *relatively* small project: finding out why there are records for 'active' members whose policies have been cancelled and how to remedy this situation.

Thus far, the QUE team has conducted over 40 interviews, exchanged 130+ emails, had numerous 'phone meetings' across the company and have begun data and document collection. The purpose of all this data collection is to understand one membership process: the Case Cancellation process.

The team has mapped data flows and business and systems process related to the Case Cancellation process. These data flows are shown at a high level in Figure 3. The complexity of the data and systems relationships is apparent.

Case Cancellation: high-level data flow

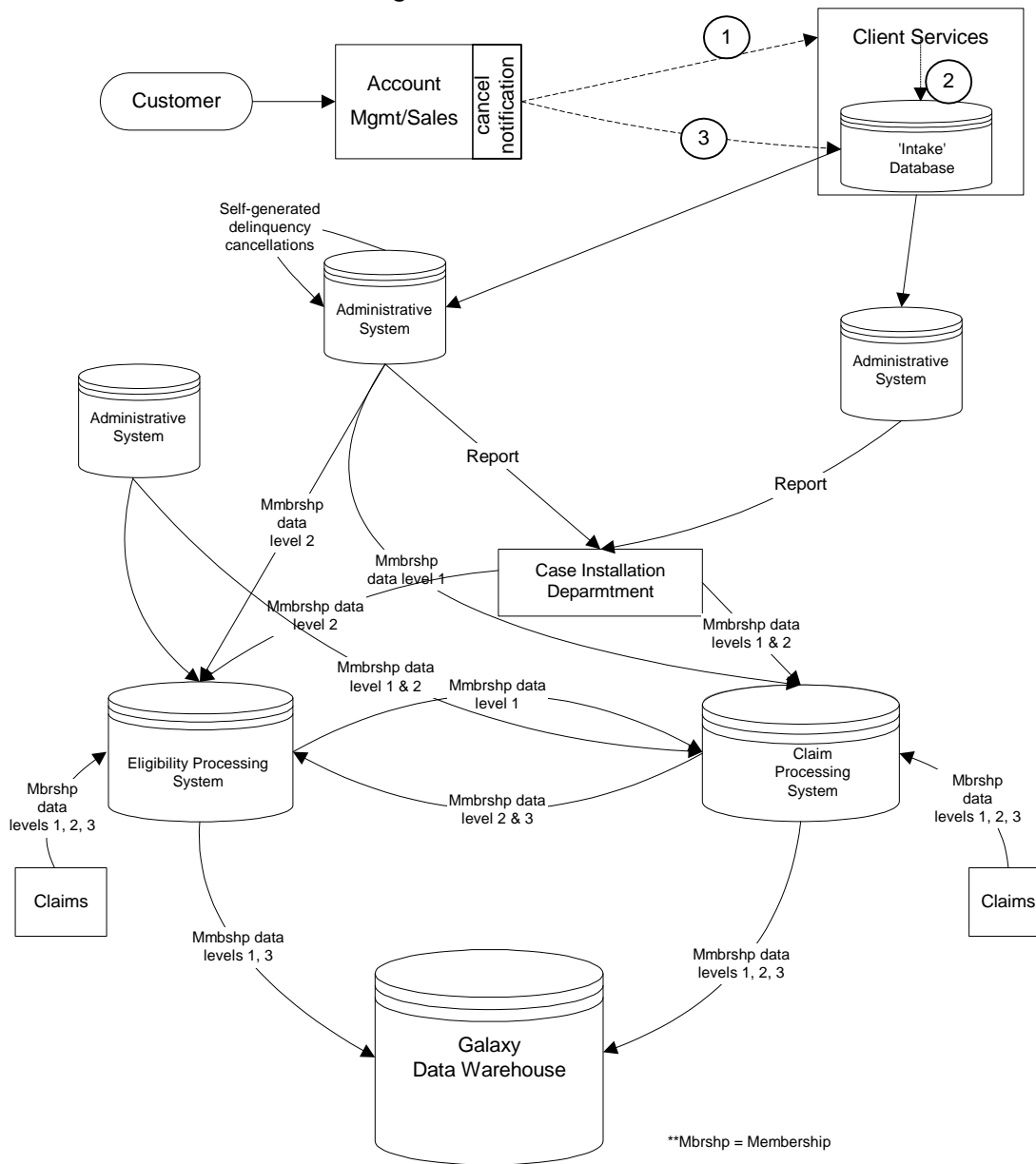


Figure 3 Case Cancellation Process

The team has also put a measurement system in place and looked at past data behaviors. For example, the control chart in Figure 4 plots the number of members who show up as active in the Administrative systems (see Figure 3), but whose policies are, in fact canceled according to data in Galaxy. The chart helps the team estimate the extent of the problems as well as some possible causes.

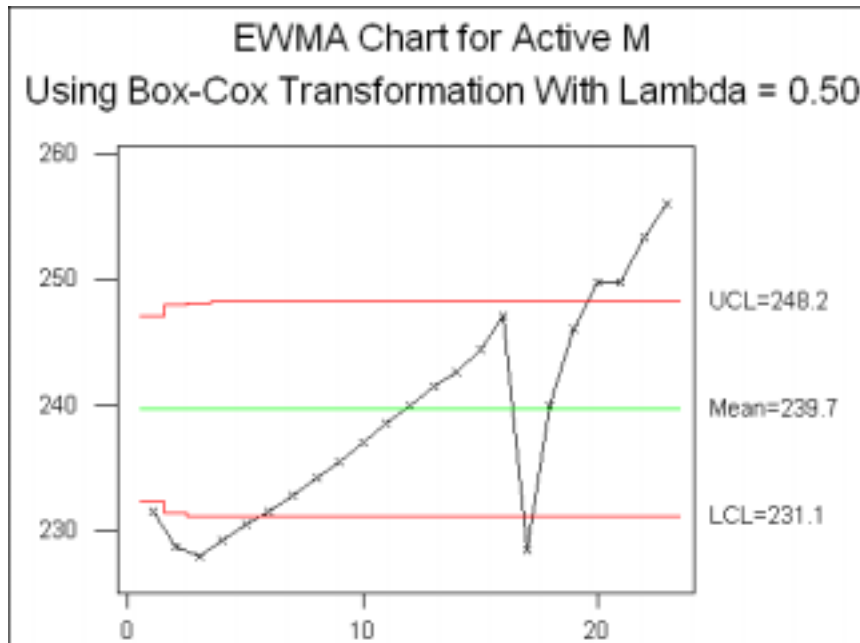


Figure 4 Control Chart indicates that the Case Cancellation process is not in statistical control.

The team has also begun statistical and in-depth root cause analysis. Information from these activities will form the basis/leverage to request changes to error-prone systems and/or business processes. These changes will result in improved quality of membership data and will be monitored with control charts and other measures. Then another improvement process will begin with a new membership-related data quality project until membership data are highly reliable. While this plan seems fairly straightforward, it is not: it consists of many iterations and its implementation is sometimes organizationally difficult.

DISCUSSION

Our purpose in this paper is to learn about how organizations keep data and information quality initiatives going in the presence of real organizational dynamics, rather than to focus on the detailed quality definitions, metrics, analysis techniques, etc. of a particular IQ project. To understand better the information quality initiatives at Ingenix and UnitedHealth Group (UHG), we present the time lines shown in Figure 5.

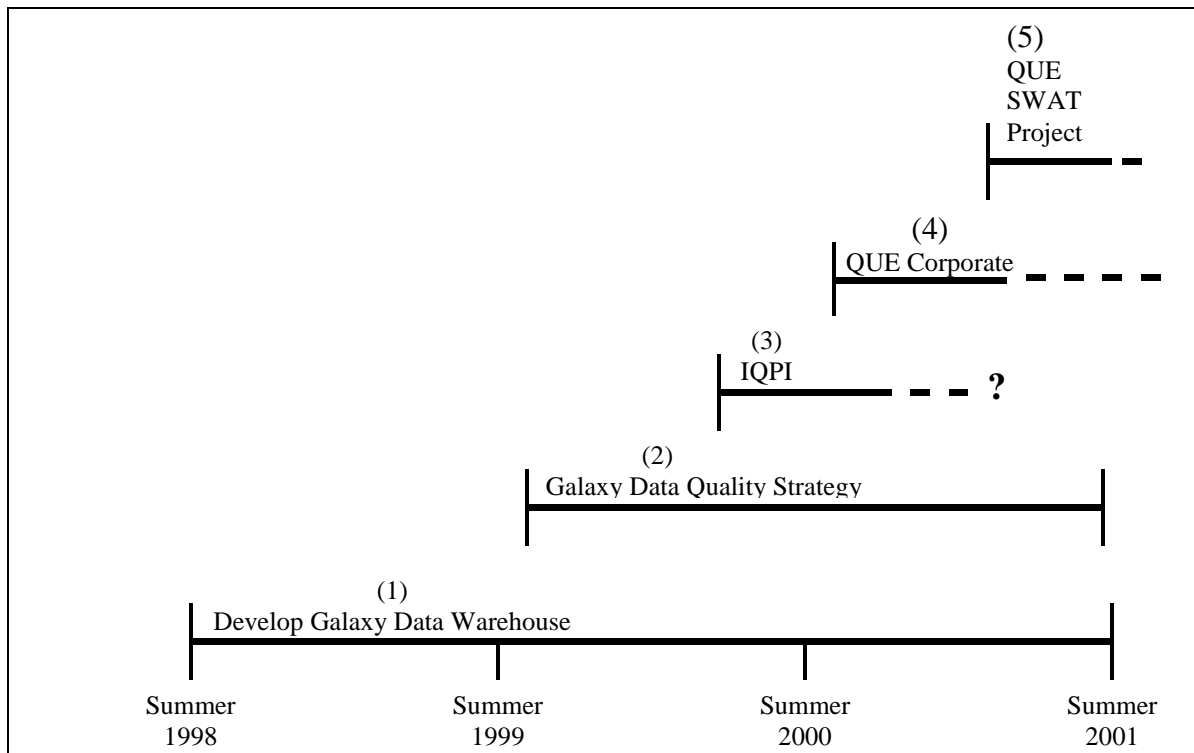


Figure 5 Time Line of Quality Initiatives

In **Figure 5**, there are five IQ projects, each of a different type:

- The baseline project is the development of the Galaxy Data Warehouse, which is a standard IS development project. As part of this project, the quality of the data is improved by carefully selecting the sources of data to use in the data warehouse. In addition, even before development began, Galaxy was designed to minimize data quality problems.
- Meanwhile IQS interviewed and surveyed users, once about their data quality needs, and again about their metadata needs. These surveys underscored the importance of quality metadata as well as quality data to Galaxy's customers. These quality-related surveys led to broader awareness of the importance of information quality to SDW, Ingenix, and eventually to Galaxy's information quality strategy.

- IQPI started within SDW and Ingenix. It was then moved to a more ‘central’ segment. IQPI was continued for about six more months, then seemingly went into hibernation.
- The QUE initiative is a corporate level quality/process improvement initiative, to be adopted by all business segments.
- Ingenix chose Galaxy data quality as one of its QUE projects.

A MORE COMPLEX PICTURE

This is a more complex picture of information quality initiatives than has been reported in the literature. The paper, “Data Quality in Context”, which studied data quality improvement projects in three organizations, reports three contexts in which data and information quality projects are usually initiated and performed [Strong, Lee, and Wang, 1997]:

- As part of an IS development project
- As an information quality project typically initiated by an information quality champion
- As a TQM project initiated as part of the company’s TQM focus

In that study, each organization used only one of these contexts for information quality improvement. At Ingenix, all three contexts are present at the same time:

1. Data quality improvement is taking place as part of the data warehouse development project (Galaxy).
2. IQS and SDW management ensured progress toward both metadata and data quality. The head of IQS is an information quality champion who was instrumental in planning Galaxy’s IQ Strategy, as well as being instrumental in initiating the IQ seminar and the IQPI initiative. This initiative became a corporate sponsored information quality project.
3. The QUE initiative provides an overall TQM-type context in which IQS formed a cross-functional team to work on a membership data quality project.

Maintaining IQ improvement projects in each of the three contexts can be difficult. Data quality initiatives within IS projects can all too easily be put on the back burner in the effort to complete the development project on time. Independent information quality projects may fail to receive adequate support of resources and of managerial attention to improve information quality in a significant way. Even if they are conducted at a corporate level rather than a local level, projects may falter. Corporate TQM-type initiatives may provide excellent contexts for local information quality initiatives--but companies can find such initiatives difficult to sustain over a long time period.

Ingenix is simultaneously maintaining information quality initiatives in all three contexts. It started as part of an IS development project, added a second context, an independent information quality group that initiates information quality projects, and finally a TQM-type context in which local quality projects (including IQ projects) are developed. These efforts demonstrate Ingenix's growing expertise in information quality improvement.

Ingenix's experiences with IQ improvement projects provide lessons to other companies. The most significant one is that IQ improvement projects/initiatives are not always easy, but by initiating projects in more than one context, these projects have a greater opportunity to succeed. Ingenix's view that IQ is important to its business has given it the vision to see opportunities for IQ improvement within IS projects, within corporate initiatives, and as purposeful information quality improvement projects. All of the three contexts can provide support for improving information quality.

CONCLUSION

This paper takes the next step beyond our paper at last year's conference, "How to get an Information Quality Program Started" to focus on how to keep an information quality program going. Although the project reported in last year's paper, the IQPI initiative, has faltered somewhat, Ingenix has made significant progress in keeping its information quality efforts alive by using several approaches, each different than the other, and some interacting with each other.

Inherent in this paper are some larger questions: "Why are 'quality' initiatives so difficult to sustain over time?" "Why was the momentum of the original IQPI initiative not maintained? Why are data quality efforts so difficult?" While there may be many reasons, one explanation lies in the very make-up of organizational systems. Looking at it from a 'systems perspective', *events* often are thought to have had a variety of causes: "This and this happened", "they caused it to happen . . ." and so on. Focusing on *events* leads to reactivity. The explanations may be true, but they are not necessarily useful. Systems thinking is *generative*. The fact that the original IQPI initiative was not maintained intact is a symptom, not a cause. Fixing symptoms doesn't help; in fact, it often worsens situations. Root causes would be difficult to arrive at without modeling the organizational system and the dynamics therein.

According to Peter Senge,

"There seems to be a particular lack of appetite in many American [organizations] for the hard work of articulating our mental models conceptually. Developing explicit models. . . of complex [organizational systems] to test alternative processes and strategies strikes many action-oriented managers as too theoretical. This is especially troubling in light of the widely recognized difficulties . . . in transferring [knowledge, information, data] from one group to another."¹

Another explanation involves the nature and current frequency of change. Macroeconomic forces have made constant change an imperative for organizations that want to survive. However, change in organizations is often met with inwardly focused cultures

¹ Peter Senge. *The Fifth Discipline*. Doubleday, New York. 1994

(organizational and/or departmental), complacency if not outright resistance, and most of all, the genuine and natural human fear of the unknown.

It is not within the scope of this paper to address these two issues—1) modeling the organization and its dynamics, including IQ initiatives and 2) resistance to change within organizations as an IQ challenge—except to bring them to light. In addition these issues strongly suggest directions for future studies.

ACRONYMS

CPI	Continuous Process Improvement
DMAIC	Define, Measure, Analyze, Improve, Control
ETL	Extraction, Transformation, Load
IQ	Information Quality
IQPI	Information Quality Process Improvement
IQS	Information Quality & Strategies
IS	Information Systems
MDR	MetaData Repository
SDW	Shared Data Warehouse
SPC	Statistical Process Control
TQM	Total Quality Management
UHG	UnitedHealth Group

REFERENCES

- Kahn, Beverly K., Katz-Haas, Raïssa, and Diane M. Strong, "How to Get an Information Quality Program Started: The Ingenix Approach", *Proceedings of the 2000 Conference on Information Quality*, October 20-22, 1999, Cambridge, MA, pp. 28-35.
- McCue, Michael T. "The Best of Both Worlds: Executive Profile". *Managed Healthcare Executive*. May 2001. p. 22
- Pande, Peter S., Neuman, Robert P., and Roland R. Cavanaugh. *The Six Sigma Way: How GE, Motorola, and Other Top Companies are Honing Their Performance*. McGraw-Hill, New York, 2000
- Senge, Peter. *The Fifth Discipline*. Doubleday, New York. 1994
- Strong, Diane M., Yang W. Lee, and Richard Y. Wang, "Data Quality in Context", *Communications of the ACM*, Vol. 40, No. 5, May 1997, pp. 103-110.