

Application of Six-Sigma for Data Quality Improvement in an Insurance Company

Mohit Panwar

Abstract: Billions of dollars annually is what poor data quality costs Insurance businesses, according to the Data Warehousing Institute. Poor data is also the leading cause of many IT project failures. Many companies are designing their data management programs and quality initiatives to deal with this problem of erroneous and inconsistent data. This paper discusses these topics as well as those detailing how companies can improve their data quality using the quality improvement technique of applying Six-Sigma. The method of creating a new initiative for identifying, measuring and controlling the data errors is discussed in this paper. Finally, the paper ends with a conclusion and expected results and benefits of the technique. The insurance industry is data-dependent. Today, carriers and intermediaries are engaged in improving data capture to help them to better manage their business, manage their risk and know their customers. Business and regulatory drivers are pushing the industry to manage its data better.

Index Terms: Data quality, Insurance data management, Six sigma for data, Data Augmentation

1. Introduction

a) Insurance companies and data

Insurance business is one of the fastest growing businesses in the world. More than 20% of the Fortune 500 companies are insurance providers. Many of these companies have a customer base of more than 50 million people. Almost all of the companies require essential data from the customers, which include the personal details, the employment details and in some cases, even the private medical information OR SSN. There are many channels to collect this information. It can be seen that most of the data coming as input to the systems of these insurance companies are either from the Customer care OR from the customer himself, by typing in the information on the company's websites, portals or by mailing in the data. As humans, we have a tendency of making mistakes which can even have false consequences. In relation with data, mistakes are more possible to occur if not typed in properly. For example, John Maxwell is a different person than John R Maxwell OR John Maxwell Jr. Even a single alphabet can cause serious consequences. Insurance companies and other financial institution deals with monetary reimbursements and even a word mistake can cause financial losses to the companies. That is why these companies are seeking more reliable ways to cope up with the data problems. The main benefit of using new technologies is their ability to handle any structure data – all your data –without requiring extensive data integration efforts. But at some point, even the input data to these new systems are also created by humans and hence, prone to errors.

- *Mohit Panwar*
Software Engineer II, Metropolitan Life Insurance Company (Metlife), Greenville SC, USA, Email: mohitpanwar@yahoo.in

b) Background.

What is Data Quality?-Data quality is an essential characteristic that determines the reliability of data for making decisions. High-quality data is: Complete: All relevant data —such as accounts, addresses and relationships for a given customer—is linked. Accurate: Common data problems like misspellings, typos, and random abbreviations have been cleaned up. Available: Required data is accessible on demand; users do not need to search manually for the information. Timely: Up-to-date information is readily available to support decisions. [1]

Business leaders recognize the value of data and are eager to analyze it to obtain actionable insights and improve the business outcomes. Unfortunately, the proliferation of data sources and exponential growth in data volumes can make it difficult to maintain high-quality data. To fully realize the benefits of big data, organizations need to lay a strong foundation for managing data quality with best-of-breed data quality tools and practices that can scale and be leveraged across the enterprise [1]

Types of Data in Insurance products- For maintaining a policy for a customer, an insurance company has to keep a record of information related to that customer. The information include both personal and industry specific data. Personal data may include Name, Address, Phone numbers, attributes (Height, weight, age etc.), dependents and beneficiary information. Industry specific data includes Medical history and drug history (For health and life insurance), Accident and ticket history (For auto and renters insurance) and any legal involvement (for legal insurance plans). All these information are provided by the customer to the company in form of text or numbers. In the internal software systems, this data is saved in the data warehouse as alphabetic or numeric data. Other data can be classified as system generated, which is generally error free. This includes the payment info, the policy amount and installments etc.

Importance of Data Quality Management- Maintaining data quality within the system is indeed important to reduce the number of errors, which is the basic target for Six-sigma. It is important to include both – the reactive components address problems that already exist, and the proactive components diminish the potential for new

problems to arise. The partnership between the business and information technology groups is critical for a successful program. Once IT understands the business rules, it is in a position to deploy the technology needed to ensure that data is being managed as a corporate asset. Unfortunately, many companies learn about the importance of data quality management the hard way. Only after several documented problems with the data do they recognize the need to improve its quality. The U. S. Government estimates that billions of dollars are lost annually due to data quality problems. The Data Warehousing Institute concluded that the cost of data quality problems exceeds \$600 Billion annually. Additional estimates have shown that 15-20% of the data in a typical organization is erroneous or otherwise unusable.

Challenges in Data quality Improvement- Deploying a data quality management program is not easy; there are significant challenges that must be overcome. Some of the most significant reasons companies do not pursue a formal data quality management initiative include:

- No agreements between departments for the responsibility and initiative.
- Often various departments are needed to be involved. This includes the IT and Database management.
- It requires the organization to recognize that it has significant problems.
- It requires time and discipline.
- It requires an investment of financial and human resources.
- It is perceived to be extremely manpower-intensive.
- The return on investment is often difficult to quantify.[1]

c) Six-Sigma : An Approach

Definition- Six Sigma is a management philosophy developed by Motorola that emphasizes setting extremely high objectives, collecting data, and analyzing results to a fine degree as a way to reduce defects in products and services. The Greek letter sigma is sometimes used to denote variation from a standard. The philosophy behind Six Sigma is that if you measure how many defects are in a process, you can figure out how to systematically eliminate them and get as close to perfection as possible. In order for a company to achieve Six Sigma, it cannot produce more than 3.4 defects per million opportunities, where an opportunity is defined as a chance for nonconformance. There are two Six Sigma processes: Six Sigma DMAIC and Six Sigma DMADV, each term derived from the major steps in the process. Six Sigma DMAIC is a process that defines, measures, analyzes, improves, and controls existing processes that fall below the Six Sigma specification. Six Sigma DMADV defines, measures, analyzes, designs, and verifies new processes or products that are trying to achieve Six Sigma quality. All Six Sigma processes are executed by Six Sigma Green Belts or Six Sigma Black Belts, which are then overseen by a Six Sigma Master Black Belts, terms created by Motorola.

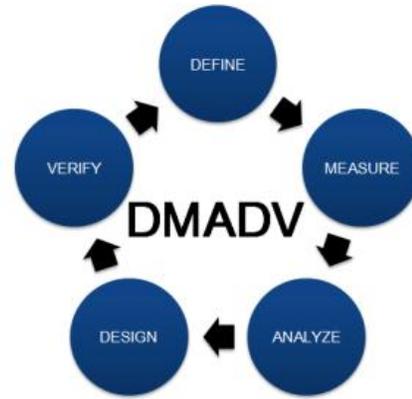


Figure 1 : Six-Sigma DMADV cycle

Benefits of Six Sigma- Six Sigma proponents claim that its benefits include up to 50% process cost reduction, cycle-time improvement, less waste of materials, a better understanding of customer requirements, increased customer satisfaction, and more reliable products and services. It is acknowledged that Six Sigma can be costly to implement and can take several years before a company begins to see bottom-line results. Texas Instruments, Scientific-Atlanta, General Electric, and Allied Signal are a few of the companies that practice Six Sigma.

2. Quality Management Plan

a) Approach

Companies can tailor business process management framework to understand data quality: Start by understanding your important corporate mandates that depend on data quality. For a credit card company, these mandates might include financial reporting, credit risk and management reporting about trends that would signal predatory or discriminatory practices. For each of these mandates, document their information dependencies (i.e., the critical data elements needed to obtain information to fulfill the mandate). Determine where data are provisioned within the company, and whether they are created by upstream business processes or obtained from customers or third parties.[2]. With this knowledge, determine criteria for “good” or “bad” data — criteria for completeness, accuracy, consistency, reasonableness, and other relevant quality dimensions. Embed these criteria in query software to measure data, use your method of choice to calculate data quality, and begin spending time on root cause analysis. A variety of formats is available to present measurement and analysis. Consider modifying statistical process control charts to track data quality. Create a simple scorecard to summarize trends by showing critical data elements on one axis, “mandates” on another axis, and data quality yield rates by mandate in cells. Feedback on other approaches, methods and presentation tools is encouraged. [2]

b) Roles and Responsibilities

One of the more significant challenges is that no single business unit is responsible for all of the data in an enterprise, and the inclusion of a responsibility for data quality in a job description is very unusual. Furthermore,

once the data is in the computer, business units often wash their hands of the problem and blame it on IT. IT cannot create the business rules nor should it be held responsible to make business decisions concerning the data. IT can only ensure that electronic rules, based on business rules, operate correctly. Effective data quality management requires organizations to adopt a data stewardship approach. Stewardship is different than ownership. A steward is a person who is expected to exercise responsible care over an asset that he or she does not own. The data is actually owned by the enterprise. The steward is responsible for caring for that asset.[1]. Another requirement is to get business people to focus on the data issues. The business leaders are concerned with their functional responsibilities. A Marketing Vice President understands customer segmentation and campaign management; data quality management is not his or her forte. The Marketing Vice President must recognize that unless his or her quality expectations are established for the data, it is unlikely that the condition of the data will be sufficient to support his or her needs. Similarly, the Manufacturing Vice President is focused on producing the products. He or she must recognize that data about the products (e.g., specifications, substitutable raw materials, inventory levels at warehouses and customer sites) impacts the department's ability to profitably produce the products.[1]

c) Return on Investment

Data quality management efforts are difficult to fund because the cost of "unquality" is not documented. The documentation of these costs requires recognition of the problem (as discussed earlier), and also requires managers to admit that they are wasting money or that they are not effectively utilizing resources at their disposals. Making these admissions, particularly in tough economic times, is risky. It is imperative that top management create an environment in which people are not unduly penalized for admitting to past problems. People within most organizations are aware of data quality problems and have taken steps to work around them. Solicit information from people working with the data and canvass customer and supplier complaints and try to discern the ones that may have been caused by erroneous data. While this information may not provide a return on investment prediction in financial terms, you are likely to find enough examples of problems to justify addressing at least one area. When you do address an area, be sure to document both the costs and the resultant savings, and use that information to justify data quality management initiatives in other areas.

d) Four Pillars of Data Quality

Once the data quality management initiative is sanctioned, the specific data subjects to be addressed and their priorities are determined. (A method for doing this is discussed within the Getting Started section.) A four-phase process for achieving successful data quality management for any particular set of data follows. Product data is used as an example within this paper. As we will see, proper tools can significantly reduce the effort required to perform each of the four steps.

Data Profiling - Data profiling is the process of gaining an understanding of the existing data relative to the quality specifications. It involves studying the existing data for errors.

Data Quality - In this step, we build on the information learned in data profiling to understand the causes of the problems. For example, the data profiling activities could reveal that we have duplicate data.

Data Integration - Data about the same item often exists in multiple databases. This data can take virtually any form (customer name and address data, product data, etc). Data quality management / integration can be applied to virtually any data problem.

Data Augmentation - Data augmentation is the last step for increasing the value of data. Data augmentation entails incorporating additional external data not directly related to the base data.

3. Plan Execution/Quality Control

a) Prerequisite

At a high level, controlling data quality is all about wrapping a process around the tasks of sourcing, transforming and publishing data that enable data quality/certification. Six Sigma provides a framework or structure around the collection, analysis and control of these processes to improve the level of data quality/certification. Two sets of interdependent processes are used to accomplish these data quality (DQ) objectives: off-line and in-line. The off-line DQ process is run outside of the certified data production process, while the in-line DQ process is run in synchronization with the certified data production process. [3]

b) Define

Majority of insurance companies depend on inputs given by their customer during the time of policy registration OR beneficiary addition. All the data provided by the customer is either by the customer himself thru website OR by phone call/email/mail thru customer care. This data then enters the system by a data processing channel and gets saved into the data warehouse for future use. As of now, many insurance companies rely on the data provided by the customer and use the same during the time of reimbursement or policy change. Various systems involved in the above process are the ones utilizing the data. There can be some more like the claims and reimbursement systems, which might encounter problems due to data errors. All such observations can be utilized to get clarity on data errors.

Voice of Customer- It is important for every company using quality improvement process to listen to the voice of customer. In the data quality management scenario, the complaints from the customers in past should be analyzed in order to narrow down the main type of errors. If required, a survey may be needed from the customer to get their point of view for Data changes. Also, the recordings from customer care portals can be utilized to get more clarity on types of errors encountered. Amongst the well-known market research methods available, surveys and focus groups can be useful for validating what an organization already presumes to know about customer needs. Questionnaires can help capture new information an

organization knows to exist and specifically seeks. These methods are scripted by the organization and typically address questions about a product or service.

Project Goal- The goal of Quality should be predefined and approved by all stakeholders. It should be measurable and defined. The target should be the overall data quality improvement and maintenance if required.

c) Measure

Current Data Issues- The quality of data in legacy systems that were developed without a data quality management program in place may be inadequate for meeting new business needs, as shown by the different representations of the same data in Figure below. For example, the accuracy of customer information may be good enough to bill a customer, but not good enough to understand the profitability of the customer; parts information may be good enough for each manufacturing facility's needs, but not good enough for understanding inventory levels and moving to a virtual parts warehouse approach. Another data problem that requires reactive action results from mergers and acquisitions. Data from the two previously separate companies needs to be combined, and this is often a daunting task, particularly if it must be undertaken without adequate tool support.

Mill Names
Courtland Mill
Courtlan
Mill, Courtland

Product Names
Savvy Matte 87 60-100
courtland Savvy Matte 60-100 87
Courtland 60-100, 87, Savvy Matte

Figure 2 – Common data writing formats [1]

Type of Data errors- Types of errors in Data can be classified as four, discussed below Incomplete – This includes data with incomplete information. This may be a result of missing records or missing fields, incomplete data due to past disasters OR due to noise in data transfer links.

Incorrect – This includes data with wrong information. This may be a result of wrong calculations, duplicate records OR duplicate processing/duplicate data entry by customer and incorrect codes used.

Incomprehensibility errors – This includes data with errors due to incorrect initial design OR planning. Example, fields within fields like First name and Last name within the field NAME, weird formatting to conserve disk space etc. Inconsistency errors – These types of data errors are generally result of manual mistakes OR incorrect codes written. This includes inconsistent codes, inconsistent timing and atomic codes, Inconsistent use of nulls, spaces,

empty values, etc. [4]. All these types of errors can be identified and a plan to mitigate can be developed.

Existing Warehouse Conditions- Data Warehouse is the home for a major part of the solution provided as all data resides in it. Data reconciliation is needed for existing data. The complexity of a data warehouse increases geometrically with the number of sources of data fed into it – Having to reconcile inconsistent systems is the reason. For example, if it takes 100 hours to reconcile data from two source systems, one can expect that it will take on the order of 400, not 200, hours to reconcile data from four source systems. The complexity of a data warehouse increases geometrically with the span of time of data to be fed into it. The data warehouse environment is interesting in that it is the source of information used by the business to make strategic decisions, but it does not actually create any data. That means that data quality problems in the data warehouse originate in the source system environment, are created because of faulty data acquisition and delivery processes, or are due to interpretation problems.

d) Analyze

Based on the information gathered in measure phase, the causes of data errors can be narrow down and identified for mitigation.

Finding the cause: We all know data quality problems when we see them. They can undermine your organization's ability to work efficiently, comply with government regulations and make revenue. The specific technical problems include missing data, misfielded attributes, duplicate records and broken data models to name just a few. But rather than merely patching up bad data, most experts agree that the best strategy for fighting data quality issues is to understand the root causes and put new processes in place to prevent them. Below are some common causes.

External Causes

- (1) Data errors caused by data coming from outside the organization.
- (2) Data errors caused by external customer
- (3) Data errors caused by manual mistakes by customers.
- (4) Data error due to inconsistent data received from group customers

Internal Causes

- (1) Data errors caused by human error in customer care
- (2) Data error due to incorrect codes
- (3) Data error due to complex databases OR older data.
- (4) Data error due to change in standards.
- (5) Data error due to merger/acquisitions.

e) Design

In this phase, the new protocols and measures to improve data quality will be discussed. The same can be used to implement process in order to get the overall data reconciliation.

Standards and Protocols.-When data quality problems are encountered importing data into the data warehouse, there are four viable actions that can be taken. We can:

- Reject the error
- Accept the error
- Correct the error
- Apply a default value for the erroneous data

When accuracy is more important than completeness, it may be appropriate to reject the error. When the data is known to be erroneous, but it is within the tolerance level, then it is appropriate to accept the error. When the correct value can be determined, then the error can be corrected. Lastly, when the correct value cannot be determined and completeness is very important, then a default value can be substituted for the erroneous data. Regardless of the course of action taken, it is very important that the data warehouse users understand the data quality implications of the chosen remedy. When data is corrected coming into the data warehouse, we are faced with an interesting situation. The data warehouse ceases to match the source system, and becomes more accurate than the source system. Minimally, the difference needs to be explained. Ideally, the process that was used to correct the data in the data warehouse can trigger a transaction that also corrects data in the source system.

New Implementations for Data improvement- Some of the measures that can be implemented to avoid data errors are discussed below –

- **Training** – Training for customer care staff and other people involved in data collection is required for data inputs to avoid human errors.
- **Coding standards** – Organization should follow one single code standard and data standard in order to avoid data errors caused by codes
- **Reviews** - Design a plan for conducting quality assurance audits. This involves routine checks of data submitted to the processing centers and management. Steps would be to:
 - assign responsibility for the quality review program;
 - formulate a program to check for completeness of coverage, accuracy and consistency of data against pre-determined standards;
 - determine what is to be checked;
 - conduct the audit on all statistical reports;
 - review findings and follow-up problems identified and provide appropriate feedback to all staff involved. This is essential at all times but particularly for staff in remote areas working in isolation;
 - retrain staff if required; and
 - repeat the process at a specified time period.
- **Improved data standards** – This is a very important change required. There are some standards needs to be defined for data and information storage. This includes changes in data fields as per the four pillars of Data quality discussed above. For example, 'NAME' field should be changed to 'FIRST NAME' and 'LAST NAME', the 'ADDRESS' fields should be changed to 'STREET', 'CITY' and other related fields.

- **Data Integration** – This includes changes in data coming from external organization OR a merged company. For example in Company A, the phone number was used as 800-123-4567 ext 2345 and in company B, the number was used as (800)123-4567 ext2345. Now, when Company A acquires Company B, the number format should be changed to Company A format.
- **New Technology** – A new process is required to reject the error nous data from external sources. This will be a firewall between the Data processing facility and Data warehouse. If any data is rejected, a function will demand new copy of the same data. This info will be sent to associated business for required measures to be done in order to get the relevant data. This measure can cope up with problems of incomplete and redundant data.
- **Existing data reconciliation** - Using an Enterprise Application Interface (EAI) bus to develop interfaces to extract data from legacy applications. Once these application interfaces are developed once, they can be used repeatedly. This is particularly important if your legacy system may run in parallel with your new improved system for an extended period of time.
 - Implementing a Master Data Management Model that clearly identifies systems that are primary owners of data elements as well as other systems that use the data.
 - Using Extract, Transform and Load tools (ETL) to extract legacy data, store in a temporary data base where data cleansing is performed.
 - After cleansing, legacy data is reloaded into legacy systems in order to improve the quality of source data.
 - Steps 3 and 4 can be repeated multiple times, with quality metric reports (produced by tools) used to assess and validate that quality gates have been met.
 - The same ETL tools used for cleansing data can be used in the 'final' extract, transform and load into the new target system.

f) Verify

Now it is time to verify and validate the new approach and training to be used within the organization to improve data quality.

Reviewing and verifying design-Reviewing the new approach involves all stakeholders in a meeting to review the new data standards and changes to existing data. This is required to make sure that the design should not cause any loss to data OR any information, which may cause financial loss to company. Also, it is required to review the training approach and feasibility for the company and business, related to data quality management. If in the review, any change is suggested OR any design flaw is found, the whole program needs to be revisited. It is necessary before implementation to get all things in place.

Quality Assurance: Pilot run- Now when we have all things in place as required, we can go ahead and test the new changes in a small system OR business function and record the impacts of it. Even though the design is now complete and a pilot run is completed in the experimental set up, when it is implemented live on the floor there could be some changes in the output. So, before going live, a simulation in the usual environment is done. All minute changes in the output and any hiccups faced during the process should be recorded. A detailed RCA (Root cause Analysis) should be done and the root causes for variation should be arrived at. After pilot run,

- Collect and Evaluate Pilot Results.
 - Analyze the gaps between the predicted performance and the actual performance.
 - Root cause the gaps to determine why and if solution changes are needed.
 - Analyze the pilot plan. What worked? What didn't? What had to be added or changed?
 - Communicate Pilot Results
- Create a summary of the strategy used to pilot the solution and communicate the results achieved.
 - Solicit Stakeholder Feedback:
 - Change management is a key part of project success. Soliciting feedback during stakeholder interviews gives you access to thoughts from those impacted by the project.
 - Review the original stakeholder analysis to determine how/if anything has changed, and what you may want to do to address those results.

4. Quality Closure: Hand Over.

Now it is the time to implement the whole new quality process throughout the organization. Implementation should be done in phases to avoid sudden impacts and also, to mitigate any issue that is encountered during the process. The implementation should be done according to the project plan and goal developed during the design phase. This involves a lot of effort from all of the employees as a new system will impact the existing and ongoing tasks of the company. The implementation date should be priorly communicated to the business with some advance lag time. The hand over not only includes delivering the new system, but also monitoring it for a specific time to get more things clear and revived with ongoing tasks. This improves the quality process to yield even better results. Generally, large insurance companies have separate quality management teams to look upon this process.

5. Conclusion

Deploying a data quality management program is not an easy task, but the rewards are enormous. Deploying a disciplined approach to managing data as an important corporate asset will better position a company to improve the productivity of its information workers and to better serve its customers. This is no longer an option, particularly in today's competitive and regulatory climate. To move forward, the key stakeholders must be educated, a stewardship function implemented, and appropriate technology must be acquired. With these in place, six sigma

program can be effectively pursued. With proper training and correct process in place, small data errors, which can lead to even large financial problems can be avoided. The return on investment cannot be calculated, but improvement in data quality is indeed rewarding for a long time for sure.

6. References

- [1]. Jonathan G. Geiger "Data Quality Management" - Paper 098-29, Intelligent Solutions, Inc., Boulder, CO
- [2]. Charles Mckinney 'Six Sigma and Data Quality' – www.isixsigma.com
- [3]. Puneet Bharal and Hamir Alform -"Making Sense of Big data in Insurance"-, Marklogic Inc. 2012
- [4]. The data Warehousing information center-<http://www.dwinfocenter.org/errors.html>
- [5]. S. Thomas Foster 'Managing Project quality' Fifth Edition -, Brigham Young University 2013
- [6]. Timothy J. Kloppenborg and Joseph A. Petrick - "Managing Project Quality", (Management Concepts, 2002)
- [7]. A Guide to the Project Management Body of Knowledge, Fifth Edition, (Project Management Institute, 2013)
- [8]. International Business Machines Ltd. Official Website <http://www.ibm.com>
- [9]. "Customer Data Quality – How to Measure it, Improve it, and Make it Stick" - EMC corporation.Six-Sigma – Wikipedia.org
- [10]. Data Quality Assessment, written by Arkady Maydanchik.