Slide 1

8th International Conference on Information Quality (IQ-2003)

Data Quality Based Applications Testing

Mark Ofori-Kyei
Mediserve Information Systems
kwame50@cox.net

Troy Lamoreaux
McKesson SPS
troy.lamoreaux@mckesson.com

Padma Kulkarni
McKesson SPS
padmaja.kulkarni@mckesson.com

Rick Thompson
McKesson SPS
rick.thompson@mckesson.com

Sai Kalapala
McKesson SPS
sai.kalapala@mckesson.com

Executive Summary/Abstract: In an effort to improve the data quality levels of applications developed by McKesson Specialty Patient Services, the Quality Assurance team has implemented a data quality approach to system testing. The QA team uses system testing activities to identify and resolve data quality defects. Data quality findings reports are developed to form the basis for long term solutions to data quality issues. The details of the current process as well as plans for process improvement are provided in this presentation.

1

Slide 2

8th International Conference on Information Quality (IQ-2003)

Agenda

- Problem description
- The solution
- Implementation examples
- Benefits of the Process
- Process Improvement
- Conclusion

2

Slide 3

8th International Conference on Information Quality (IQ-2003)

The Problem

- Like every other large corporation, McKesson Specialty Patient Services (SPS) has encountered data quality issues
  - **incorrect address information**
  - **invalid patient and physician information**
  - **data elements with unreasonable values**

3

Slide 4

8th International Conference on Information Quality (IQ-2003)

SPS Quality Assurance Team

- Currently has 5 team members
- Established in Dec., 2002 with the following mission statement:
  - **Ensure the delivery of defect-free software products by providing quality assurance services and facilitating the implementation of quality assurance best practices.**
- Systems testing was the original focus

4

Slide 5

8th International Conference on Information Quality (IQ-2003)

The Challenge

- How can typical system testing activities be modified to uncover existing and potential data quality problems?
- Can system testing activities form the basis for data quality process improvements?

5

Slide 6

8th International Conference on Information Quality (IQ-2003)

System Testing Activities

- The following steps are performed in sequence by Quality Assurance Analysts during projects:
  - **Test planning**
  - **Test case development and execution**
  - **Defect tracking and resolution**
  - **Data quality findings report**
  - **Post implementation review**

6

8th International Conference on Information Quality (IQ-2003)

Test Planning

- Based on IEEE Standard for Software Test Documentation (829-1998)
  - **The software test plan prescribes the scope, approach, resources, deliverables, and schedule of all testing activities associated with development.**
- The Data Quality Findings Report is listed as a testing deliverable in our software test plans.

7

8th International Conference on Information Quality (IQ-2003)

Test Case Development

- Based primarily on the following methods:
  - **Boundary value analysis**
    - *A test data selection technique in which values are chosen to lie along data extremes. [3]*
  - **Equivalency partitioning**
    - *A process for identifying a set of classes for input conditions with the aim of minimizing the number of test cases required to test these input conditions. [3]*
  - **Functionality (Boolean conditions)**
  - **Data reasonableness - test for believability**
  - **Requirements coverage**

8

8th International Conference on Information Quality (IQ-2003)

Defect Tracking and Resolution

- Test case failures are documented and defect information is entered in the defect tracking database
  - **Defect number**
  - **Defect date**
  - **Tester**
  - **Description**
  - **Severity**
  - **Priority**
  - **Proposed action**
  - **Status**
  - **Defect classification**
- Defects are classified to form the basis for metrics collection. The categories are based partially on the TDQM information quality dimensions.

9

8th International Conference on Information Quality (IQ-2003)

Defect Classification Definitions

- Accuracy – a measure of data correctness [2]
- Reasonableness – a measure of data believability [2]
- Relevancy - the degree to which system data matches system data definitions [2]
- Timeliness – the degree to which data is available when needed [2]
- Completeness – the degree to which all required data is populated [1] [2]

10

8th International Conference on Information Quality (IQ-2003)

Defect Classification Definitions Continued

- Accessibility – characteristic of being able to access data when it is required [1] [2]
- Security – the prevention of unauthorized access to a system [2]
- Representation – the degree to which data is represented in the proper format [2]
- Functionality – the degree to which the system satisfies the requirements

11

8th International Conference on Information Quality (IQ-2003)

Data Quality Findings Report

- Provides more detailed description and analysis of data quality defects
  - **Uses defect classifications to form the basis for data quality metrics collection.**
    - *For example: There are 5 occurrences of inaccurate data on a report that is generated from data warehouse data. An ETL process extracts data from the OLTP to generated the report. The data quality findings report provides metrics on the amount of inaccurate data in the OLTP system.*
- Aids in data quality process improvement activities
- Components
  - **Description and sizing of each data quality defect**
  - **Source of the defect**
  - **Recommendations for resolution**

12

8th International Conference on Information Quality (IQ-2003)

Post Implementation Review

- The purpose of the post implementation review is to share the results of systems testing activities with the application development team and present opportunities for process improvement.
  - ◆ Review the data quality findings
  - ◆ Review the testing metrics
  - ◆ Review the results of a process assessment based on SEI CMM level 2 Key Process Areas
  - ◆ Present a process improvement plan

13

8th International Conference on Information Quality (IQ-2003)

Process Implementation

- Process has been used on 3 projects:
  - ◆ Test plans and test cases developed with data quality in mind
  - ◆ Data quality based defect classifications formed the basis for testing metrics
  - ◆ Data quality findings reports were generated
  - ◆ Post implementation reviews were held
  - ◆ The 3 projects were:
    - ★ *A reporting application that uses an ETL process to extract data from an online system.*
    - ★ *An online application for a patient assistance program (PAP).*
    - ★ *An online Customer Relationship Management (CRM) application.*

14

8th International Conference on Information Quality (IQ-2003)

Report Application Findings

- There were defects attributed to inaccurate data in the production OLTP that was extracted for reporting by the ETL process:
  - ◆ Unreasonable patient and physician names
  - ◆ Invalid addresses
    - *Incorrect state values*
    - *Incorrect state/zip combinations*
    - *Invalid cities*
  - ◆ Unreasonable ages
  - ◆ Root cause is inadequate controls at the points of data entry.
  - ◆ Logic was added to ETL process to cleanse data for reporting.
  - ◆ Process improvement in place to add controls for data entry.

15

8th International Conference on Information Quality (IQ-2003)

Report Application Metrics

- Classifications
  1. Accuracy (49%)
  2. Reasonableness (21%)
  3. Relevancy (0)
  4. Timeliness (0)
  5. Completeness (23%)
  6. Access (0)
  7. Security (0)
  8. Representation (5%)
  9. Functionality (2%)

16

8th International Conference on Information Quality (IQ-2003)

PAP Application Findings

- There were incidents where the data that was displayed on the screens was not in the format that was specified in the design document:
  - ◆ The code was modified to ensure that the data was displayed in the correct format
- The online screens permitted the entry of inaccurate and unreasonable data:
  - ◆ Additional controls were placed in the code to prevent the entry of inaccurate and unreasonable data.
- There were incidents where the application did not perform as designed.
  - ◆ The code was modified to match the system requirements.
  - ◆ Peer reviews have been recommended as a process improvement for defect prevention.

17

8th International Conference on Information Quality (IQ-2003)

PAP Application Metrics

- Classifications
  1. Accuracy (20%)
  2. Reasonableness (4%)
  3. Relevancy (1%)
  4. Timeliness (1%)
  5. Completeness (1%)
  6. Access (0)
  7. Security (1%)
  8. Representation (17%)
  9. Functionality (55%)

18

8th International Conference on Information Quality (IQ-2003)

CRM Application Findings

- There were incidents where system data did not match the data definition:
  - ◆ **System was modified to ensure that system data matches data definition**
- There were incidents where the data that was displayed on the screens was not in the format that was specified in the design document:
  - ◆ **The code was modified to ensure that the data was displayed in the correct format**
- There were incidents where the application did not perform as designed.
  - ◆ **The code was modified to match the system requirements.**
  - ◆ **Peer reviews have been recommended as a process improvement for defect prevention.**

19

8th International Conference on Information Quality (IQ-2003)

CRM Application Metrics

- Classifications
  1. Accuracy (2%)
  2. Reasonableness (7%)
  3. Relevancy (10%)
  4. Timeliness (1%)
  5. Completeness (1%)
  6. Access (0)
  7. Security (0%)
  8. Representation (13%)
  9. Functionality (66%)

20

8th International Conference on Information Quality (IQ-2003)

Benefits of the Process

- Focuses attention on data quality issues during system testing.
- Provides metrics which can be used for process improvement.
- Defect classifications identify categories that should be well documented in requirement and design documents.
- Data quality activities are integrated with system testing activities so DQ is not seen as additional overhead.[4]

21

8th International Conference on Information Quality (IQ-2003)

Process Improvement Plan

- Focus on data quality defect prevention
- Defect
  - ◆ **A problem introduced in one development phase but found in a subsequent phase.**
- Error
  - ◆ **A problem that is introduced and found in the same development phase.**
- Plan to prevent data quality defects by containing data quality faults within the phase that they are introduced.

22

8th International Conference on Information Quality (IQ-2003)

Data Quality Defect Prevention

- Plan to add a data quality focus to the inspection of requirement and design documents.
- Inspections are evaluation techniques whereby the software requirements, software design, or code are examined by a person or group other than the author to detect faults, violations of development standards, and other problems. [5]
- The checklists that will be used to aid design and requirement inspections will include a section on the categories identified by the defect classifications.
  - ◆ **Develop a design for data quality mindset.**

23

8th International Conference on Information Quality (IQ-2003)

Conclusion

- System testing provides a great opportunity for introducing data quality concepts.
- It is important to keep the process simple.
- Data quality findings and testing metrics must be documented and communicated to the application developers and management so that process improvement activities can be effectively implemented.
- Consider using inspections to facilitate data quality defect prevention.

24

8th International Conference on Information Quality (IQ-2003)

References

[1] English, L.P.,*Improving Data Warehouse and Business Information Quality,* John Wiley and Sons, Inc., New York, NY, 1999. pp. 461-465

[2] Huang, K., Lee, Y.W., Wang, R.Y., *Quality Information and Knowledge,* Prentice Hall PTR, Upper Saddle River, NJ, 1999. pp. 43-47

[3] Kit, E., *Software Testing in the Real World: Improving the Process,* Addison-Wesley Publishing Company, Harlow, England, 1995. pp. 83-87

[4]Lister, J., Mobisson, G., Ofori-Kyei, M., "Evolution of a Data Quality Strategy", *Proceedings of the 7th International Conference on Information Quality*, Cambridge, MA, 2002

[5] ANSI/IEEE Standard 1012-1986, *"Software Verification and Validation Plans"*

25