Employing the TDQM Methodology: An Assessment of the SC SOR

Valerie Sessions
Charleston Southern University
vsessions@csuniv.edu

November 11, 2007
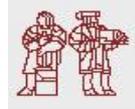
Outline

- Background/ Objectives
- TDQM Methodology
- Creation of the LE Toolkit
- Example Results
- Conclusions
- Discussion and Future Research
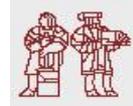
Background

- TDQM [1] is overarching methodology for data quality assessments

- Must be modified and customized for each field

- National Institute of Justice (NIJ) sponsored a grant to customize TDQM for the Law Enforcement (LE) field

- We present here the beginnings of a toolkit and example assessment
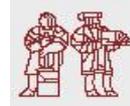
TDQM Methodology

- Developed by DoD/ DISA and MIT
- Based on Deming's TQM constructs
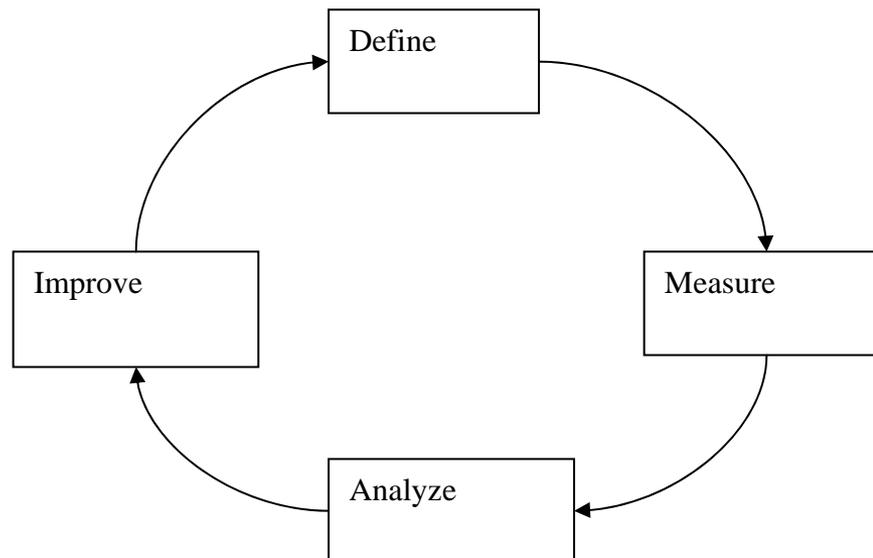
TDQM Methodology

- Great methodology but few tools for LE to put into practice

- Phase I of this study created a toolkit for LE DQ Studies

- Provide an example usage of this toolkit to assess SC Sex Offender data sets
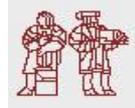
Toolkit

1) Creation of Data Dictionary

2) Query Database

      A. Query for Completeness

      B. Query for Format

      C. Query for Believability

3) Field Interviews

4) Survey
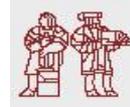
Toolkit

1) Creation of Data Dictionary

Element:	**ADDRESS: COUNTY**
Definition:	Subject's county of actual residence.
Data Type:	AN
Field Size:	2
Tables/Codes:	SC County codes table
Public:	Yes
Drop Down:	Yes
Reported to NCIC:	Yes
Comments:	None

Toolkit

1) Creation of Data Dictionary

    A) Can't analyze your data until you know it

    B) Determine the state of an agency's systems – how was database designed and how does data enter the system (speaks to data flow)

    C) Look for text entry vs. boolean fields

    D) Look for format checking at POE
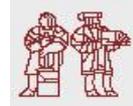
Toolkit

2) Query the Database

   A) Query for Completeness (% missing fields – analyze column and row completeness)

   B) Query for Format

      Ex. Is SSN 9 digits?

   C) Query for Believability

      Cross-Checks with other agencies

      Ex. Does DMV match SOR?

Toolkit

3) Field Interviews – TALK TO YOUR PEOPLE!!

In LE it is especially important to talk to each jurisdiction as they are tied to different laws and requirements.

Conduct:

I. State Interviews

II. County Interviews

III. City Interviews

Sample Questions:

- *What is your position with the state/county/city? What does your job entail (especially related to the Sex Offender registry)?*

- *What laws mandate the fields that must be inputted to the Sex Offender System?*

- *It is our understanding that the County is obligated to report the known Sex Offenders to schools within the county. How often is this down and in what format?*
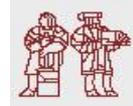
Toolkit

4) Questionnaire/Survey

- Equalizer – fair, quick and easy
- Not well received by LE – no time, small result sets
- Example survey – online
- Used variety of well tested questions from AIMQ [2] paper
- Example survey indicated location information most important to LE
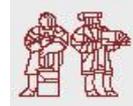
Example Results

- Process is easier to understand with an example

- SC SOR was kind enough to undergo this process on its public SO data

- Present here our results and future avenues for research and improvement

Why Sex Offender data sets?

- Example of a National, State, Local data set

- Public and LE-only data sets

- Results applicable on a more National level

- 1 in 4 girls, 1 in 6 boys sexually abused by the age of 18

- Most offenders repeat abuse
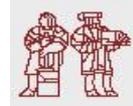
Example Results

- Performed all "toolkit" tasks for the SC SOR

- Created a Data Dictionary using NY State as example [4]

- Queried Database for completeness and format errors

- Extensive interviews with State, County, City officials

- Online Questionnaire/ Survey
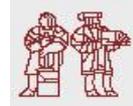
SC Results

1) Timeliness has increased due to Megan's Law.

2) Reporting format to cities problematic.

3) More complete reporting to schools in order.

4) Data Dictionary results are promising.

5) Inaccuracies in address field.

6) Completeness in reporting to NCIC/FBI.

7) Overload at county level.

Future Research

Address Validation

- Mapping systems fail due to inaccuracies in address fields

- Not a problem until data sets are rolled up to state/national level

- Propose to link mapping technology to Address Point of Entry

- Leverage Google Maps to keep costs low for Law Enforcement
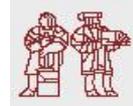
Future Research

School Reporting System

- 179 Registered "Schools" in Dorchester County alone
- With no mapping capability reporting to all schools nearly impossible
- Propose School Reporting System built onto Address Validation System
- Helps both schools and Sheriff's Departments
- Partnering with *Darkness to Light* – Charleston based non-profit

References

[1] Cykana, P. Alta Paul, Miranda Stern. DoD Guidelines on Data Quality Management. *Proceedings of the International Conference on Information Quality.* 1996.154-171.

[2] Lee, Yang W., Diane M. Strong, Beverly K. Kahn, and Richard Y. Wang. AIMQ: a methodology of for information quality assessment. *Information and Management.* 2002. 133-146.

[3] 42 U.S.C.A. § 14071 (2000). James Wetterling Act.

[4] NY Statewide Criminal Justice Data Dictionary <http://criminaljustice.state.ny.us/dict/dict.htm>
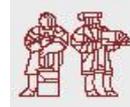
Legalese

Valerie K. Sessions, Ph.D. is a Research Scientist for the College of Charleston and an Assistant Professor of Computer and Information Science at Charleston Southern University. The research conducted by the College of Charleston, Charleston, South Carolina was supported by contract number N65236-06-P-2513 awarded by the Space and Naval Warfare (SPAWAR), Systems Center, Atlantic. The findings, conclusions and opinions expressed in this document are those of the author and do not necessarily reflect the official position or policies of the National Institute of Justice, SPAWAR, Atlantic, the College of Charleston, or Charleston Southern University.