# DATA QUALITY 2

MSBO Certification course
Rob Dickinson, MPAAA Executive Director

## Data Quality 2

- Session Agenda
  - *Putting DQ in context*
  - *Quality Assurance*
    - Data Definitions & Types
    - QA methods
  - *Data Governance*
- Questions

## Defining Data Quality

Involves both tangible (Quantitative) and intangible (Qualitative) measures

## Quantitative measures

- Accuracy
- Integrity across systems
- Consistency
- Completeness
- Uniqueness
- Accessibility
- Precision
- Timeliness

## Qualitative measures

- Relevance
- Usability
- Usefulness
- Believability
- Unambiguous
- Objectivity

## Quality Assurance

- Controlling data as it enters your systems
- Usually part of system design/installation
- 2 Major areas
  - *Data field design*
  - *Input control functions*

# Putting DQ in Context

- Data Quality is one part of larger model – Data Governance
- Data Governance:
  - *Policies, processes, and practices that control our data and ensure it's quality*
  - *Hard to see directly, easier by example:*

# Putting DQ in Context

- Where most Organizations are:
  - *Data is defined inconsistently across systems*
  - *Student data is duplicated*
  - *Staff time wasted massaging data*
  - *Fragmented view of students exists*
  - *Accuracy issues in key data elements*
  - *Inefficient, leads to 11th hour scramble*

# Putting DQ in Context

- The goal is:
  - *Key data elements sync across systems*
  - *Student information is not duplicated*
  - *Staff spends time analyzing, not verifying*
  - *Systems show a COMPLETE picture of student*
  - *Systems report efficiently for all compliance needs*
  - *Certification deadline is just another day*

## Putting DQ in Context

- Not just data
  - *How well is staff trained on data definitions?*
  - *Are field 'owners' known to all?*
  - *How are staff informed of inevitable changes in these things?*
  - *Are staff encouraged to analyze data?*
  - *Does EVERY staff know data privacy rules, and live them?*
- All there things add up to Data Governance

## Putting DQ in Context

- Data Quality
  - *2 primary focuses*
- Quality Assurance
  - *Methods and ways to keep bad data from getting into systems*
- Quality control
  - *Ways to find and correct bad data once it's in our systems*

## Data Field Design

- Selecting the most appropriate type of field for the data it will hold, and assigning properties to that field to limit bad inputting.
- Field Types: Boolean, number, text, date
- Coded fields: Intrinsic, non-intrinsic
- Field Formats: Check boxes, buttons, selection lists, input fields

# Field Types

Boolean

- ONLY 2 values - Yes/No, True/False
- Status (Participant status, Enrolled, Was Absent on Count day)
- Can NEVER hold a 3rd option
- Usually cannot be left blank
- Won't allow for any future re-definition

# Field Types

Number

- Used for values, amounts
- Sometimes used for codes
- Significate digits are important
- Subtypes
  - *Integer – 1, 2, 3 (no decimal)*
  - *Currency – Always 2 digits of decimal*
  - *Floating Point – No functional limits*

# Field Types

Text

- Used for list of values, string input
- WEAK choice for number only input
- Direct input – Almost impossible to analyze
- Using text for numbers
  - *Allows leading '0', fixed width*
  - *Only for list of codes*

# Field Types

Dates

- Used for inputting dates, sometimes times
- Sometimes stored as number
- Usually built-in error checking for valid dates
- Allows date math
- Formatting for century (3/1/2016 vs 3/1/16)

# Code Fields

- Stores limited list of values
- List determines field type (number, text, etc)
- Good error checking
- Adding & deleting values is a problem
- When creating – Intrinsic vs non-intrinsic
  - *Intrinsic – the stored data conveys information*
  - *Non-intrinsic – stored value has no meaning on its own*

# Code Fields

Intrinsic or Non-intrinsic?

UIC

SSN

MSDS Exit codes '19'

MSDS Ethnicity codes '010000'

EEM District codes '41010'

EEM Building Codes '03921'

# Code Fields

**Intrinsic codes**

- SSN, Gender, Special ed program codes
- Good
  - *Easy to understand*
  - *Built in error checking*
- Bad
  - *Needs strong rules*
  - *Limits possible values*
  - *Needs to know all possible values*

# Code Fields

**Non-intrinsic codes**

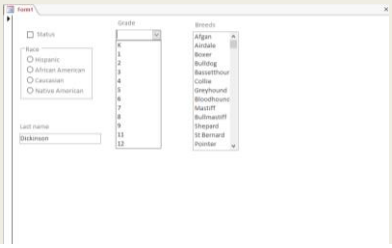
- UIC, EEM Building codes, MSDS Exit codes
- Good
  - *Not limited by rules*
  - *Can accommodate growth/change*
- Bad
  - *Has no value in itself, needs value chart/list*
  - *Can run into limits (field width)*
  - *Can only work if there is only 1 place generating values*

# Field Formats

**The interface that controls how the data is entered**

- Checkboxes, radio buttons
  - *Boolean data, 1 choice among very few*
- Lists, Dropdown lists
  - *List choices available, one or more than 1*
- Input box
  - *Most freeform, hardest to control input*

# Field Formats




# QA Methods

**Ways to ensure data is entered into your systems correctly**

- Error checking at input
- Training for input staff
- Error checking routines run at regular intervals

# Error checking at Input

**Prevent bad data from getting into the system**

- Data Types, field formats
- Error checking rules behind the field
- Make it difficult to allow non-standard data to be input
  - *Can't make it so hard that it is ignored*
  - *'Are You Sure?'*

# Training for Input Staff

**Make sure staff entering data is aware of it's importance**

- Initial training
  - *Bring new staff up to speed*
  - *Familiar with systems*
- Recurring training
  - *Letting everyone know what's new, changed*
  - *Reminders on problem areas*

# Error checking routines

**Frequently run reports/queries designed to find errors soon after input**

- Find and fix before it is used, propagated to other systems
- Nightly, over weekend, end of attendance period
- Can be system report, email, faxed, etc.
- Do you fix, or do they?
- Balance of finding errors vs overwhelming users

# Data Governance

- Data Horror stories
  - *Japan Stock Market, 2005*
  - *Bear Sterns, 2002*
  - *SID data – West Michigan, 2011*
- Impact of poor data governance

# Data Governance

**Data Governance Strategy**

- Overall vision for improvement
- Program Implementation plan
- Linking data Quality back to District policies and objectives
  - *How does good data make education easier?*

# Data Governance

**Technology & Architecture**

- Flexibility to change
- Open and Common Standards
- Data accessibility among systems
- End-to-end data security

# Data Governance

**Governance Organization**

- D.G. recognized at a organizational level
- Data quality as an embedded competency for ALL staff
- Data Stewards recognized and known
- Senior Stakeholders recognized and known

# Data Governance

**D.G. Processes**

- Correction processes
- Root cause analysis
- Best practices and methods
- Focus on Improvement
  - *Starting on Key elements*
- Supply chain approach

# Data Governance

**D.G. Policies**

- Common definitions
- Data Standards
- Review of Policies and Standards
- Defined Controls

# Data Governance

**Data Monitoring/Investigation**

- Qualitative understanding of issues
- Key data pieces identified
- Ongoing monitoring
- Tracking of issues for Improvement

## Getting Help

- CEPI Helpdesk
  - *(517) 335-0505, Option 3*
  - *cepi@michigan.gov*
- MPAAA
  - *Rob@mpaaa.org*
  - *(517) 853-1413*