DATA QUALITY 2

MSBO Certification course

Janell Craig, Oakland Schools

1

Data Quality 2

- Session Agenda
  - *Where we Stand*
  - *Defining Data Quality & Data Governance*
  - *Quality Assurance*
    - Data Definitions & Types
    - QA methods
  - *Data Governance in practice*
- Questions

2

2

Where we stand

Data Quality Campaign (DQC)

- 10 Essential Elements for a Statewide Longitudinal Data System
  1. *Unique Student Identifier*
  2. *Student-level enrollment, demographic, program participation information*
  3. *Ability to match student level testing year-to-year*
  4. *Information on Untested students*
  5. *Teacher identification system*

3

3

Where we stand

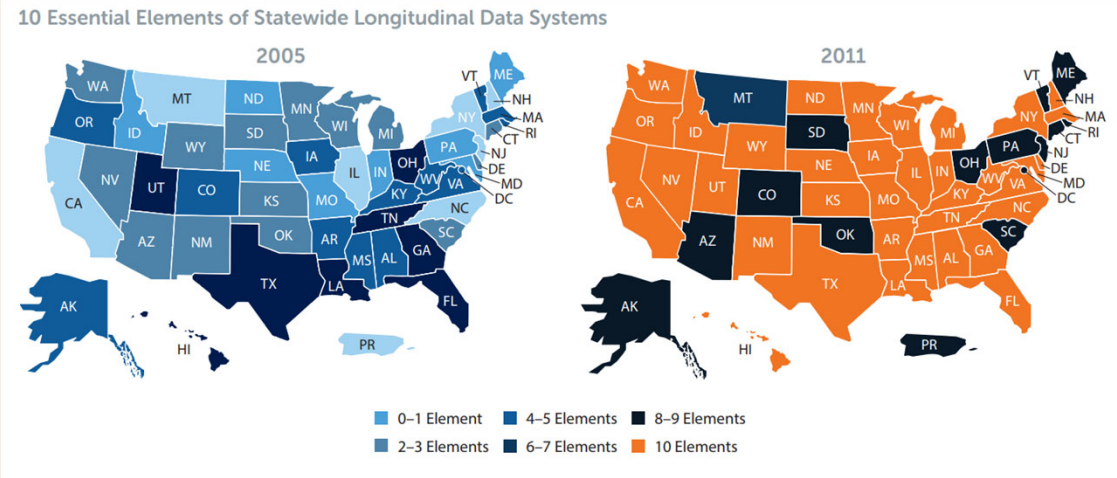
Data Quality Campaign (DQC)

- 10 Essential Elements for a Statewide Longitudinal Data System
  6. *Student-level transcript data*
  7. *Student-level college readiness test scores*
  8. *Student-level Graduation and Dropout data*
  9. *Match students from K-12 to postsecondary systems*
  10. *Statewide data audit system*

4

4

Where we stand



Source: dataqualitycampaign.org

5

5

Where we stand

Data Quality Campaign (DQC)

- 10 State Actions to Ensure Effective Data Use

1. Link K-12 data to Early learning, postsecondary, and workforce data
2. Create stable support for SLDS system
3. Create governance structures
4. Build state data repositories
5. Data collection for more than compliance purposes

6

6

Where we stand

Data Quality Campaign (DQC)

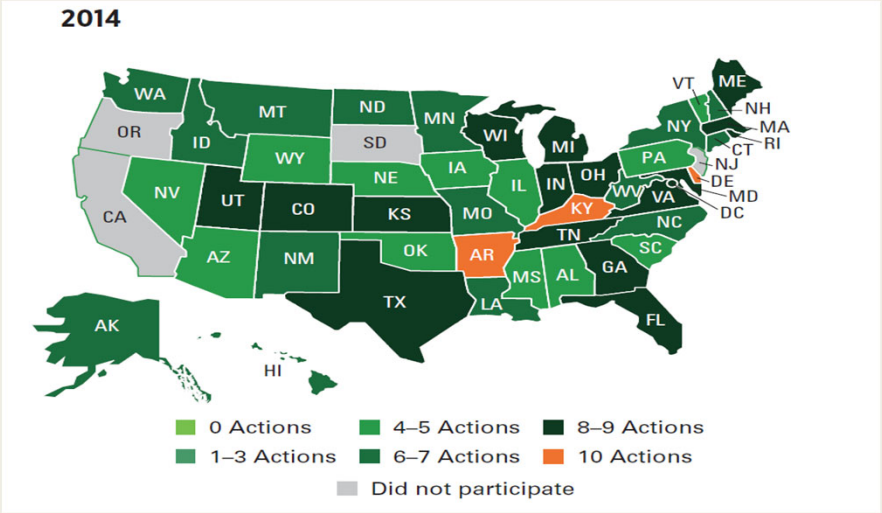
- 10 State Actions to Ensure Effective Data Use

6. Create reports for educators, students, parents
7. Have reports provide longitudinal statistics
8. Develop a purposeful research agenda
9. Implement policies to build educators capacity to use data
10. Raise awareness of available data

7

7

Where we stand



Source: dataqualitycampaign.org

8

8

Where we stand

Data Quality Campaign (DQC)

- *By 2014, Michigan met all elements and all Actions*
- *By 2017, all states with SLDS systems met all Elements*
  - Several still working on Actions
- *Michigan continues to lead in national data quality discussions*
- *CEPI staff involved in DQ policy and practice discussions at the national level*

9

9

Defining Data Quality

Wikipedia - ***Data quality*** *refers to the condition of a broad set of qualitative and quantitative variables. There are many* ***definitions of data quality*** *but* ***data*** *is generally considered high quality if it is "fit for [its] intended uses in operations, decision making and planning".*

- Not 'perfect', or 'error free'
- Involves both tangible (Quantitative) and intangible (Qualitative) measures

10

10

Quantitative measures

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

11

11

Qualitative measures

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

12

12

Defining Data Quality

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

13

13

Data Governance

- 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*

14

14

Data Governance

- 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*

15

15

Data Governance

- 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 these things add up to Data Governance

16

16

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*

17

17

Quality Assurance

- Controlling data as it enters your systems
- Important part of system design/installation & maintenance
- 3 Major areas
  - *Data field design*
  - *Input control functions*
  - *System modification/customization*

18

18

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

19

19

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, or blank is considered one of the values
- Won't allow for any future re-definition

20

20

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*

21

21

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*

22

22

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)

23

23

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*

24

24

Code Fields

Intrinsic or Non-intrinsic?

UIC

SSN

MSDS Exit codes '19'

MSDS Ethnicity codes '010000'

EEM District codes '41010'

EEM Building Codes '03921'

25

25

Code Fields

Intrinsic codes

- SSN, Gender, Special ed program codes
- Good
  - *Easy to understand*
  - *Built in error checking*
- Bad
  - *No privacy – allows guessing (identity theft)*
  - *Needs strong rules*
  - *Limits possible values*
  - *Need to know all possible values*

26

26

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*

27

27

Field Formats

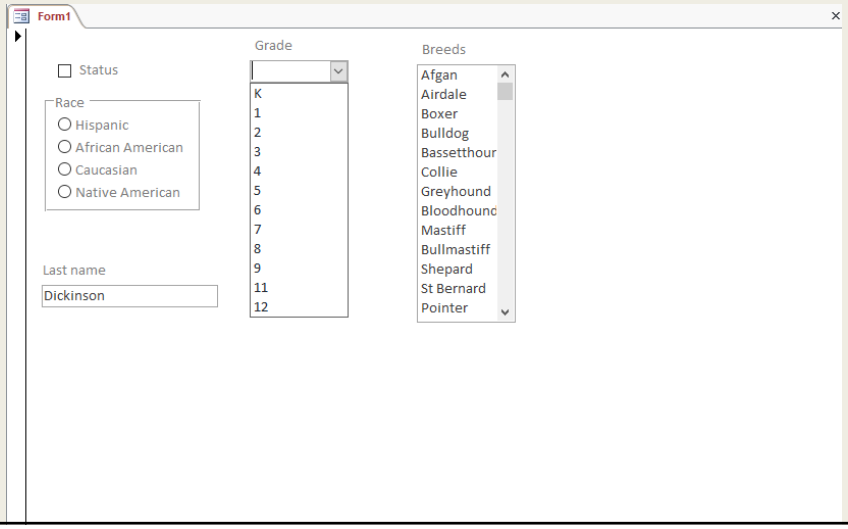
The interface that controls how the data is entered

- 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*

28

28

Field Formats



29

29

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
- New screens, reports, queries follow same rules

30

30

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?'*

31

31

Training for Input Staff

Make sure staff entering data is aware of its 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*

32

32

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

33

33

Error checking routines

New screens or reports or queries MUST follow same QA rules to prevent 'trapdoor' errors

- New screens – 'All in one' or audit screens
- Reports – Do report generators follow security rules?
- Queries – double edged sword
  - *Easy to run, change data quickly*
  - *Usually avoids all the user input rules*
  - *Can fix - or break - large amounts of data very fast*

34

34

Data Governance

- Data Horror stories
  - *Social Security Administration, 1936*
  - *Bear Sterns, 2002*
  - *Japan Stock Market, 2005*
  - *SID data – West Michigan, 2011*
- Impact of poor data governance

35

35

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?*

36

36

Data Governance

Technology & Architecture

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

37

37

Data Governance

Governance Organization

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

38

38

Data Governance

D.G. Processes

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

39

39

Data Governance

D.G. Policies

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

40

40

Data Governance

Data Monitoring/Investigation

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

41

41

Getting Help

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

42

42