

A large, thick black L-shaped graphic is positioned on the left and bottom edges of the slide, framing the central text.

DATA QUALITY 2

MSBO Certification course

Rob Dickinson, MPAAA

Data Quality 2

- Session Agenda
 - *Where we Stand – National Data Quality campaign*
 - *MiSchooldata website*
 - *Data Governance at the Policy Level*
 - *Quality Assurance*
 - Data Definitions & Types
 - QA methods
- Questions

Before we begin

Data Quality challenges

- What are some of your data quality challenges?
- What do they cost you?
- How can you fix them / keep them from happening?

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*

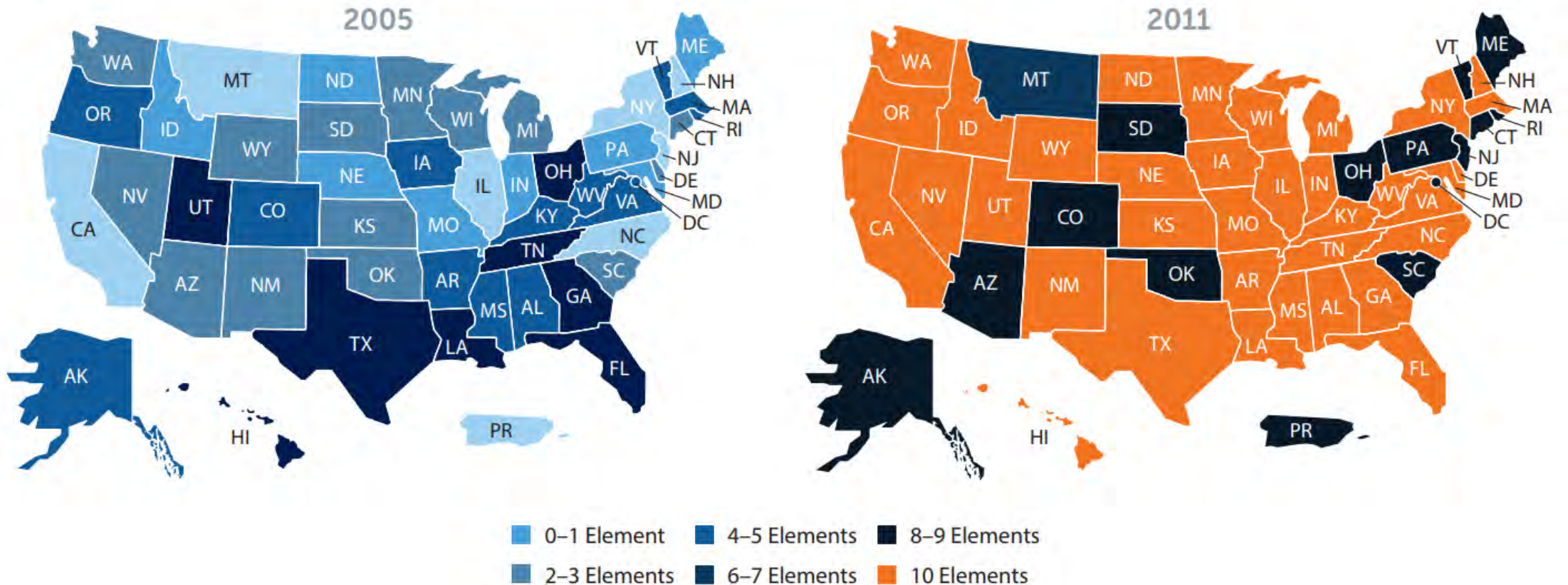
Where we stand

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*

Where we stand

10 Essential Elements of Statewide Longitudinal Data Systems



Where we stand

Data Quality Campaign (DQC)

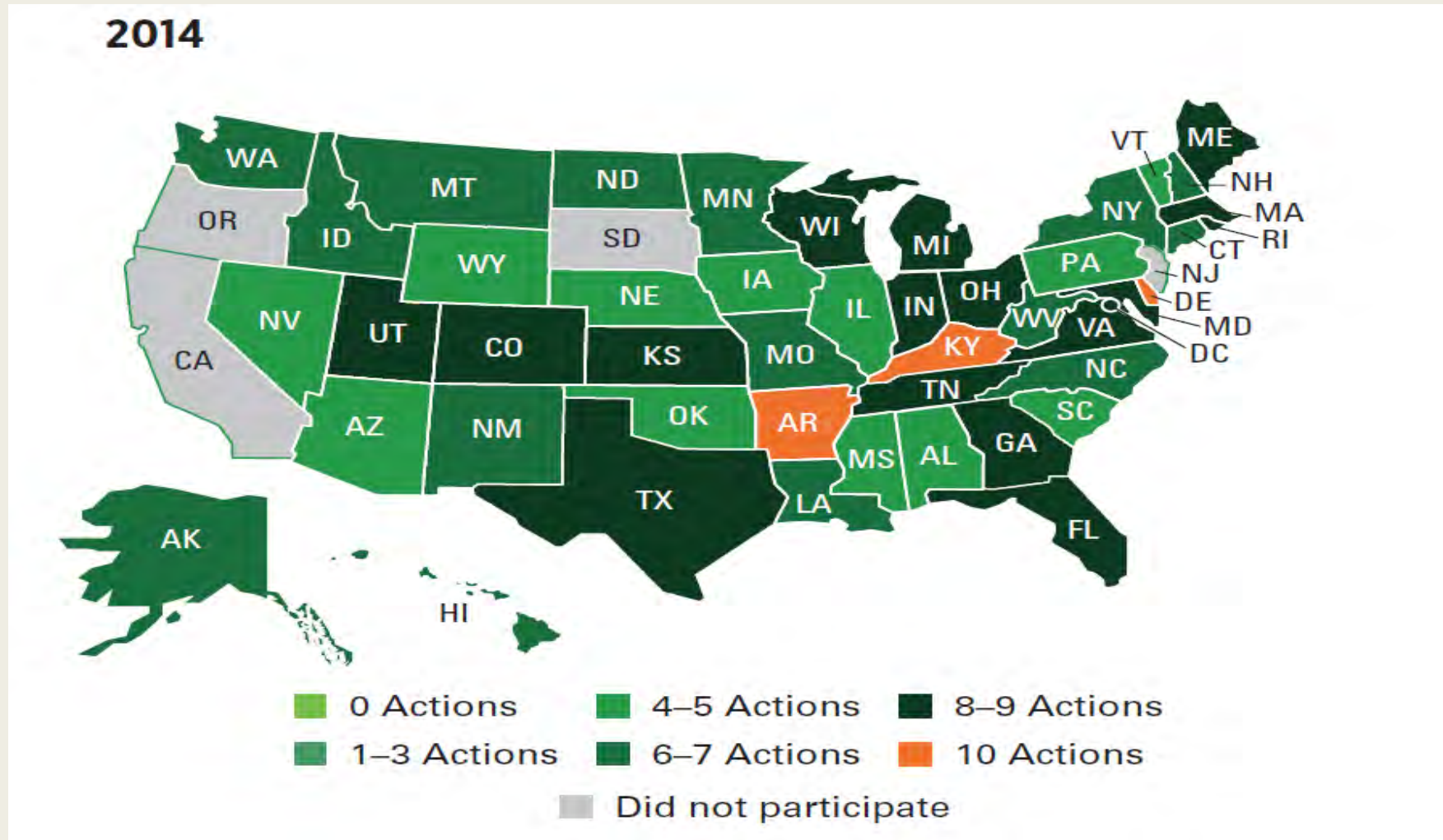
- 10 State Actions to Ensure Effective Data Use
 1. Link K-12 data to Early learning, postsecondary, and workforce data
 2. Create stable support for SLDS system
 3. Create governance structures
 4. Build state data repositories
 5. Data collection for more than compliance purposes

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

Where we stand



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*

Where we stand

Data Quality Campaign (DQC)

- *More recently*
- *Emphasizing USE of data, now that it is in good shape*
- *Beyond Compliance reasons*
- *How data can affect education of students*
- *Published Handbook – Educational Data 101, A Briefing Book for Policymakers*
- <https://dataqualitycam.wpenginepowered.com/wp-content/uploads/2021/03/DQC-EducationData101-031821.pdf>

Where we stand

Educational Data 101

10 things to emphasize about educational data

1. *Student Data helps improve student achievement*
2. *State longitudinal data systems help answer questions and drive improvement*
3. *Student growth data provides a more equitable picture of student and school performance*
4. *Data linkages provide the fullest picture of student and school outcomes*

Where we stand

Educational Data 101

10 things to emphasize about educational data

5. *Data generates the evidence that state and local leaders need to make decisions*
6. *Student data must be kept private and secure*
7. *State report cards provide the public information about school performance*
8. *Data empowers teachers and parents with information to better support learning*

Where we stand

Educational Data 101

10 things to emphasize about educational data

9. *Educator Preparation programs need data to improve teacher training and quality*
10. *Teachers must be equipped with the skills to understand and use data effectively*

Where we stand

Educational Data 101

- <https://dataqualitycam.wpenenginepowered.com/wp-content/uploads/2021/03/DQC-EducationData101-031821.pdf>
- *Excellent resource:*
- *Lays out in greater detail*
- *Links to examples, articles, and surveys*

Where we stand

MiSchoolData website

<https://www.mischooldata.org>

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

Quantitative measures

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

Qualitative measures

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

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:*

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*

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*

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

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 an 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

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*

Quality Assurance

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

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

Field Types

Number

- Used for values, amounts
- Sometimes used for codes
- Significant 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
 - *List of options (listbox) gives greater control*
- 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*
 - *Can be generated by anyone who knows the rules*
- Bad
 - *No privacy – allows guessing (identity theft)*
 - *Needs strong rules*
 - *Limits possible values*
 - *Need 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*

Code Fields

Intrinsic or Non-intrinsic?

UIC – *Non-intrinsic*

SSN – *Was intrinsic, changed late 90's*

MSDS Exit codes '19' – *Non-intrinsic*

MSDS Ethnicity codes '010000' - *Intrinsic*

EEM District codes '41010' - *Intrinsic*

EEM Building Codes '03921' - *Non-Intrinsic*

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

The screenshot shows a Windows form titled "Form1" with the following controls:

- Status
- Race
 - Hispanic
 - African American
 - Caucasian
 - Native American
- Last name:
- Grade:
 - 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 11
 - 12
- Breeds:
 - Airdale
 - Boxer
 - Bulldog
 - Bassetthour
 - Collie
 - Greyhound
 - Bloodhound
 - Mastiff
 - Bullmastiff
 - Shepard
 - St Bernard
 - Pointer

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

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 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*

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

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*

Back to the beginnning

Data Quality challenges

- The Data Quality challenges you discussed at the start of the session:
- Do you have any additional insight now?
 - *Causes*
 - *Fixes?*

Getting Help

- CEPI Helpdesk

- (517) 335-0505, Option 3

- cepi@michigan.gov

- MPAAA

- Rob@mpaaa.org

- (517) 853-1413