

Data Quality 1

MSBO Certification course

Rob Dickinson, MPAAA Executive Director



Data Quality 1

- Session Agenda
 - Putting DQ in context
 - Defining our terms
 - Why is quality data so hard to achieve?
 - Finding the problems
 - Fixing the problems
 - Data Privacy (FERPA/FOIA)
- Questions

History of Data

- Data – information recorded in an organized fashion for quick lookup and retrieval
- 150 AD – Ptolemy Star Catalog
- List of 1,100 stars, their constellations, brightness, position, etc.

History of Data

Longitudo et Latitudo ac Magnitudo stellarum fixarum

Forme et Stelle	Longitudo			S	Latitudo		mag. oc. mag.
	o	g	m		g	m	
Que est in medio reclinatozj sedis	0	7	50	S	51	40	3
Que est in extremitate reclinatozj	0	7	50	S	51	40	6
Illaz g̃ tredecē stellaz in magnitudine tertia sunt q̃tuor. in quarta sex. in quinta vna. in sexta due							
Stellatio Leleub: cui nomē i latino ē p̃seus: ⁊ ē deferēs caput Algol. Imago Undecima							
Stella q̃ ē in resolutione nebulosa: q̃ ē sup extremitatē man ^o dextre	0	27	40	S	40	35	nebulosa
Que est super marsic dextrum	1	1	10	S	37	30	4
Que est super spatulam dextram	1	2	40	S	34	30	4 .e.l.
Que est super spatulam finistram	0	27	30	S	32	20	4 .e.l.
Que est super caput	1	0	40	S	34	30	4
Que est inter duas spatulas	1	1	30	S	31	10	4
Lucida que est in latere dextro	1	4	50	S	30	0	2
Antecedens trium que sunt post eam in hoc latere	1	5	20	S	27	30	4
Media trium	1	7	0	S	27	40	4
Sequens earum	1	7	40	S	27	30	3
Que est super marsic finistram	1	0	40	S	27	0	4
Lucida earum que sunt in capite Algol	0	29	40	S	23	0	2
Sequens earum	0	29	10	S	21	0	4
Antecedens lucidam	0	27	40	S	21	0	4
Antecedens hanc etiam: ⁊ est secunda	0	26	50	S	22	15	4
Que est in genu dextro	1	14	50	S	28	15	4
Antecedens hanc: ⁊ est supra genu	1	13	50	S	28	10	4
Antecedens duarum que sunt in ventre coxe	1	12	20	S	25	10	4
Stella postrema earum in vnitute ventris coxe	1	14	0	S	26	35	4
Que est super musculum cruris dextri	1	14	10	S	24	30	5
Que est super calcaneum dextrum	1	16	20	S	28	35	5

Perseus

History of Data

- 1st large database – United States Social Security Number system
- Established in November of 1935 – 25 million numbers initially issued.
 - 1st number went to John Sweeney Jr, of New Rochelle, NY
 - Over 1,000 post offices (later SSA offices) had to coordinate issuance of numbers, recording of income

History of Data

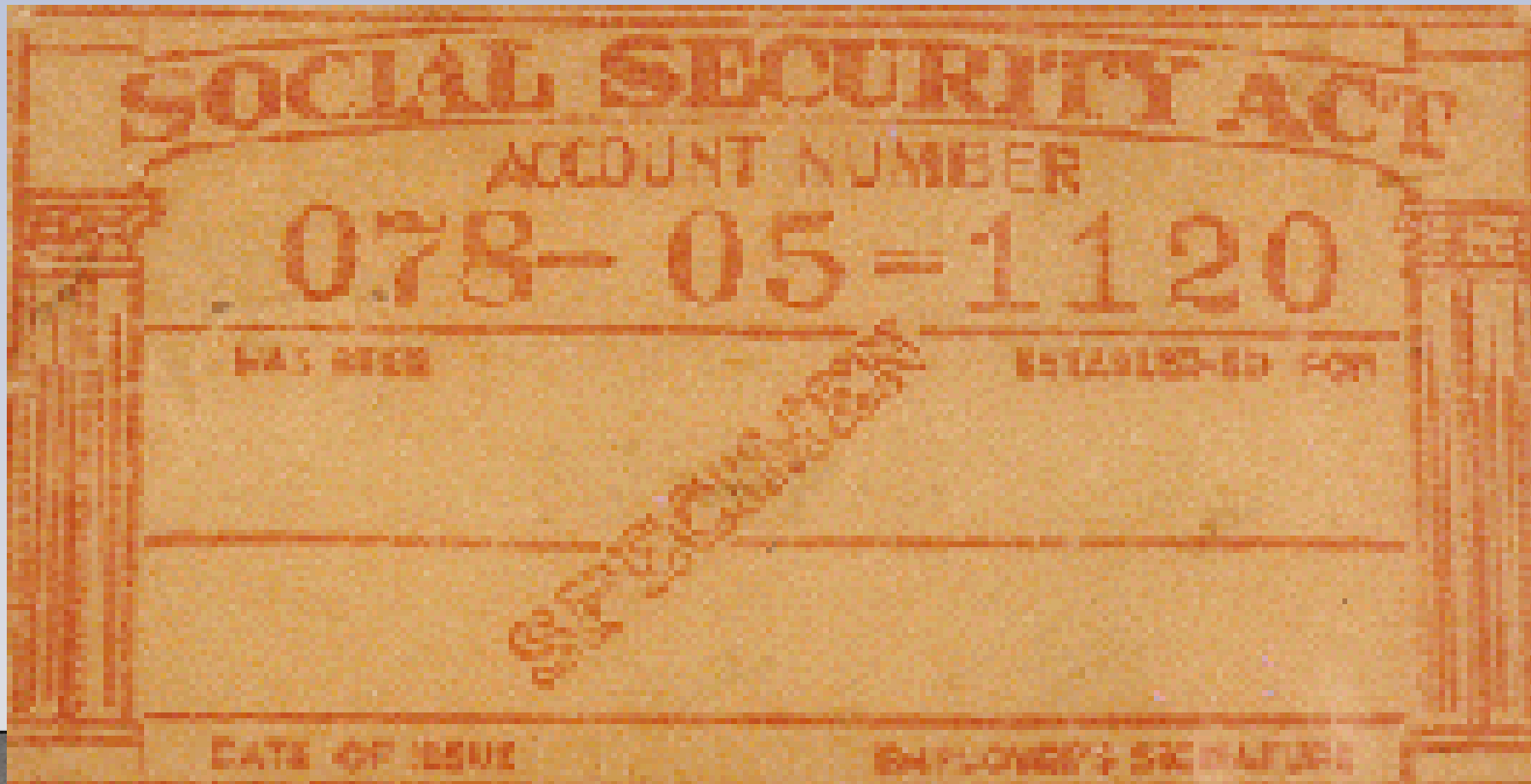
- SSN – 293-47-2031
- Issued by Toledo OH office in 1971
- Issued to someone who's last name began with 'D'
- 31st person that year whose last name began with 'D' to get a SSN at the Toledo office in 1971
- Identity theft, anyone?

History of Data

- 1st data issue – Hilda Witcher
- 1936 – EH Ferree Company – New York
- New national contract for Woolworths Dept Store
- Clear box covers (new acetate plastic)
- Wanted something to put in the wallets to catch attention
- SS cards had been going out for around 3 months, lots of newspaper attention

History of Data

- 1st data issue – Hilda Witcher



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 its 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 these things add up to Data Governance

Defining our terms

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

Defining our terms

- Pupil Accounting terms
 - FTE – Full Time Equivalency
 - CEPI – Center for Educational Performance and Information
 - MSDS – Michigan Student Data System
 - General Collection – 3 time/year snapshot data collection
 - SRM – Student Record Maintenance
 - EEM – Educational Entity Master

Defining our terms

- Pupil Accounting terms
 - UIC – Unique Identification Code
 - PIC – Personnel Identification Code
 - SIS/SMS/SRS – Student Information/Management/Record System

Defining our terms

- Pupil Accounting terms
 - MDE – Michigan Department of Education
 - OAS – Office of Accountability Services
 - Secure Site – District user site maintained by OAS for districts to manage the testing of students

Defining our terms

- Data Quality
 - Data that is fit for its intended use
- Not “Perfect data”

Why is it so hard?

- Complex systems becoming interlinked
 - Student records – Special Ed
 - Student records – Food service
 - Student records – Bus Routing
 - Student records – Personnel / HR
 - Student Records – Financial?
 - Student Records – Public portals/Websites

Why is it so hard?

- Projects are very goal driven, usually compliance driven & punitive (negative feedback only)
 - Fosters attitude “meet THIS requirement, then move on”
- Do you/anyone have time to analyze data, find & fix errors?
- Does your work environment/culture value or invest in data quality?

Why is it so hard?

- Software Development methods don't emphasize data quality
 - Systems are evaluated on functions/features
 - Look of screen, how it functions
 - Information integrity is not valued as a decision-making criteria
 - If customers don't ask for data integrity, vendors won't build it.

Why is it so hard?

- Data Quality exercise:
 - Form being handed out
 - Create data input rules for date of birth
 - Ages where warning or errors SHOULD occur
 - Building or District level?
 - Work with your table

Why is it so hard?

- Data problems are hard to find
 - Data quality evaluated by different systems than capture it
 - Delays in time
 - Personnel, source data no longer available
 - Once problem propagates, much harder to root out.

Why is it so hard?

- Input fitness
 - Data is usually only made clean to the level needed by the person inputting it.
 - Example - building staff, working with parent
 - No incentive to maintain high data quality
 - Errors only show when data is summarized and integrated – which is usually at time of reporting

Why is it so hard?

- Do you know all your data's "Intended use"?
 - Data exists in our systems a LONG time
 - Impossible to know ALL intended uses at time of entry
 - Collection systems can't anticipate every future need
 - Reactive legislature adds to the problem

Who Cares!?!

- Why is quality data important?

Who Cares!?!

*The price of quality data
is far lower than the cost
of the alternative*

Who Cares!?!

- Costs of bad data
 - Financial
 - Embarrassment
 - Reputation

Finding the problem

- Who are the flag wavers?
 - Who will know something is wrong?
 - How can they fix it, or raise flag?
- How can YOU find errors?
 - What data can you analyze?
 - What's be best time for analysis?

Finding the problem

- Flag waivers
 - Parents
 - Parent portal
 - Opportunity to send message on bad data
 - Double edged sword

Finding the problem

- Flag waivers
 - ETL processes & partners
 - Data errors that occur when files transfer to other software
 - Nightly with special ed system
 - Occasionally with transportation/other systems

////////////////////////////////////
If you want it done right . . .

- Do it yourself!
 - Compare summary data
 - Do totals by various categories seem close?
 - Need good feel for your data

Finding the problem

- It can be challenging
 - Student ethnicity – African American
 - CEPI data “010000” or “000100”?
- You shouldn't rely on memory

Finding the problem

- Data analysis
 - Can you run queries?
 - Download data from your SIS?
 - Can reports be dumped to Excel?
- Beware of privacy issues!

Deciding to fix the problem

- Can the data be fixed?
- Should the data be fixed?

Can the data be fixed?

- Is the correction a new value for this field?
 - No – Easy fix
 - Yes – Not so easy

Can the data be fixed?

- New values for a field
- Mechanical issues
 - Will new value fit nature of field?
 - Will it fit in size of field?
- Policy/process Issue
 - What else will new value affect?

Can the data be fixed?

- New data values – Effects:
 - All existing reports, queries
 - Any existing error checking
 - Other users of this field
 - Other systems that link at this data

Should the data be fixed?

- Reasons NOT to fix data
 - Fix has no net benefit
 - Cost greater than benefit
 - Lack of resources
 - Internal need greater than external

Fixing the problem

- At what level should the fix be executed?
- How should the error be fixed?
- Who should fix the error?

What level?

- Sometimes, you don't fix the data
 - Internal needs override external
 - Different uses at different levels
- Options:
 - Fix the PROCESS, not the data
 - Create a reporting field

Create Reporting field

- Internal field used to break out, or consolidate data for reporting
- Place to build in checks
- Usually NOT available to users
- Change data submission routine to use this field

Fixing with Query

- ALWAYS use 2 step process
 1. Run list of errors that shows bad data to be changed as it exists **before** the change
 2. Use EXACT same logic to fix error

Fixing with Query

- Keep list of errors, showing how data *WAS* before the correction
- **ALWAYS** list before changing
- The more you can do, the more you can damage!

Fixing with direct input

- One-off error, single fix
- Run through user interface whenever possible
- Allows any existing error traps to run

Who should fix the data?

- Users vs Finders
 - Users are the inputters of the data in normal day to day usage
 - Finders are the data staff who collect/report/analyze the data

Who should fix the data?

- User fix
 - Closest to that particular piece of data
 - 'Owner' for Data Governance
 - Typically, small part of their job
 - Hard to get their attention

Who should fix the data?

- Finder fix – District level staff
 - Has greatest motivation to make data right
 - Able to concentrate on correction
 - Disconnecting users from their data

Fixing the data: A special case

Fixing errors in CEPI submissions

- A special case
 - Two approaches
 - Reload
 - Online correction

Errors in CEPI Submissions

Reload data

- Fix the error in your system, recreate the upload file, re-upload
 - Eliminates synchronization errors
 - Longest process

Errors in CEPI Submissions

Online correction

- Fix the error in CEPI online system
 - Fast fix
 - Now your source data is different than your reported data

Errors in CEPI Submissions

What is your transition point?

- First load w/many errors, reload
- At some point in process, switch to online
- Keep synchronization in mind

Make your life easier

What can you do to make data quality easier?

- Input forms match input screens
- Controls are LIMITED to avoid bad data
- Keep data originators there while you input

Sharing Records

FOIA and FERPA

- Freedom of Information Act (FOIA)
 - All District records not exempted may be requested
- Federal Educational Rights and Privacy Act (FERPA)
 - Allows student record to be seen by appropriate individuals

Release of Records under FOIA

- Anyone can request records from a public body
- Ask to inspect, copy, or receive a copy of the record
- No age restrictions (cannot refuse the request of a minor BECAUSE they are a minor)
- 5 days to respond with the record, or request for more information, or a 10 day extension to prepare for complex requests
- Not required to CREATE documents, merely share existing ones

Review of Records under FERPA

- Parents have the right to view their children's educational records
- Students assume that right when they turn 18
- Districts have obligation not to share records with outside parties without parental consent
- Districts must provide parents notice of their rights under FERPA

Protecting Records

- Freedom of Information Act (FOIA)
 - Exempt records are not required to be shared
 - Reasonable costs can be charged
- Federal Educational Rights and Privacy Act (FERPA)
 - Any record containing Personally Identifiable Information (PII) is protected

FOIA Exceptions

- Student records covered by FERPA
 - FERPA trumps FOIA
- Law enforcement records
- Records covered by attorney-client privilege
- Pending bids on public contracts
- Test questions and answers or scoring keys
- records relating to civil actions
- Requests by prisoners in County, state or federal correctional facilities

Getting Help

FOIA Assistance

- Michigan Attorney General Website
- Go to www.michigan.gov/ag
- Choose 'FOIA and Open Meetings' from top menu

Getting Help

FOIA Assistance

- Frequently Asked Questions
- Informational pamphlets to distribute
- All current legal citations
- Open Meetings handbook as well

FERPA

- Parents have the right to view their children's educational records
- Students assume that right when they turn 18
- Districts have obligation not to share records with outside parties without parental consent
- Districts must provide parents notice of their rights under FERPA

FERPA

What are educational records?

- Every record that the district holds that contains Personally Identifiable Information (PII)
 - Exception – personal notes for personal use only
- PII – Any information that discloses an individual's identity
 - Even summary data that can reasonably point to an individual

FERPA

Parents:

- FERPA defines a parent as a natural parent, a guardian, or an individual acting in the place of a parent or guardian. In the case of divorce or separation of a student's parents, both parents maintain full rights under FERPA unless some legally binding document or law revokes those rights.
- It is NOT the school's responsibility to hunt down or find those documents. Only to act on them when they are presented.

FERPA

Accessing and viewing records

- District has 45 days after a written request to make records available
- District must comply with reasonable requests for explanation of documents
- Copies do not have to be made, so long as parent can view documents

FERPA

Amending records

- Parent has the right to request that incorrect data be corrected in the record.
- School is obligated to CONSIDER the request.
- If the school decides to not change the data, they must inform the parent of their right to a hearing.
- If a hearing is held, and the parent the decision is made to not change the data, the parent has the right to put a statement into the student's file explaining their position on the contested information.
- School has to maintain that statement and produce it whenever it discloses the contested data on the student.

FERPA

Amending records

- Under FERPA, schools are not required to consider requests to change:
 - A grade for a student because they feel the student should have been given a better grade
 - Opinions
 - Disability placement decisions regarding the student
 - Discipline decisions regarding the student

FERPA

Student rights

- When a student turns 18, their FERPA rights transfer from the parents to the student
 - Also applies when a student enters college at any age
- Parents may no longer view students records, unless:
 - Student is listed as a dependent on parent's current tax records
 - Pursuant to a health emergency
 - Student consents

FERPA

Sharing protected records

- District may not disclose information without parental consent, unless:
 - Directory Information
 - Exempt recipient

FERPA

Directory information

- Information that is not harmful, or intrusive to privacy
 - Name
 - Address & other contact information
 - Date and place of birth
 - May NEVER include Social security number
- District may determine what it constitutes directory information

FERPA

Directory information

- Student numbers
 - Can be considered directory information
 - Can be displayed on student ID
- UNLESS
 - Number is used as a password, without other PIN or counter code, to gain access to student data
 - In that case, it cannot be on the badge, and is not eligible to be considered directory information

FERPA

Directory information

- May be disclosed without consent if:
 - Annually, notice is given to all parents of what records district deems to be directory
 - Gives parents the option to opt out of disclosure
- If parent opts out, Directory info for that student **MUST NOT BE PROVIDED**.
 - Most SIS have FERPA opt-out check boxes
 - Important that custom reports reference that status

FERPA

Exempt recipients (within the district)

- Records may be shared with:
 - Employees of the district with a 'legitimate educational interest'
 - Teachers, administrators
 - Attorneys, counselors, nurses
 - IT staff
 - Contractors, consultants, 3rd parties providing services

FERPA

Exempt recipients (outside the district)

- Records may be shared with:
 - Public health and law enforcement
 - Emergency responders
 - Only for imminent or already existing emergencies.
Not as preparation for future hypothetical events

FERPA

Exempt recipients (outside the district)

- Records may be shared with:
 - School a student transfers to
 - Accrediting organizations
 - Local, state or federal officials with audit or program evaluation needs
 - Organizations the student has applied to for admission for school, or for financial aid
 - To comply with a subpoena or judicial order

FERPA

Annual Notice

- Must include:
 - Parent's rights to inspect records
 - Right to request corrections
 - Right of consent to disclosure of PII
 - Procedure for inspecting records
 - Procedure to requesting corrections
 - Information of reporting complaints to USED Family Policy Compliance Office
- Frequently included in student handbook

FERPA

Training assistance

- Privacy Technical Assistance Center (PTAC)
 - Part of US Department of Education
- <https://studentprivacy.ed.gov/>
- Rules, current case law
- Training videos
- Will track learning, provide documentation of compliance
- Excellent source for training staff

Getting Help

- Pupil Accountant
 - Closest, knows your data best
- ISD Auditor
 - Knows the rules and regulations
 - Does your Pupil Accountant want to be a liaison?

Getting Help

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

