# Data Quality Assurance

Nathan Bunker
-
Michigan Care
Improvement Registry

# Agenda

- Lessons Learned
- What We Did
- Demonstration of DQA Report

# Lessons Learned

1. **IIS is Responsible:** The IIS is ultimately responsible for data quality and is uniquely qualified and positioned to identify and resolve data quality issues.
2. **More than Validation:** Data quality assurance is much more than just message format and required field checking.
3. **Errors and Warnings:** Data quality assurance process can identify both errors and warnings.
4. **Pattern Detection:** Data quality assurance process can identify problematic patterns that only appear when reviewing a larger set of messages.
5. **Integrated and Ongoing:** Data quality assurance process must be a integrated process that is used for all incoming data, before and after going to production.

# DQA #1 IIS is Responsible

- Spot the problems:
  1. On July 22, 2012 Sally was administered on OPV at a local Kalamazoo, Michigan clinic in her left arm.
  2. Family clinic in Flint, Michigan regularly reports administering DTaP to teenagers.
  3. Pediatric clinic in Grand Rapids, Michigan regular reports large amounts of data but never reports administering MMR.

# DQA #1 IIS Is Responsible

- IIS is ultimately responsible for data quality.
- Data quality assurance tasks can not be shifted to outside entities.
- Data quality is a process not a software.

# DQA #2 More than Validation

- Data Quality includes Message Validation.
- Message Validation includes:
  - Structure
  - Required Values
  - Correct Values
- Message Validation is a critical first step.
- Data Quality is more than validation.

# Example: Hiring for IIS Help Desk

**Sam Latham**
516-555-0000
sidthekid@yahoo.com

**Education**

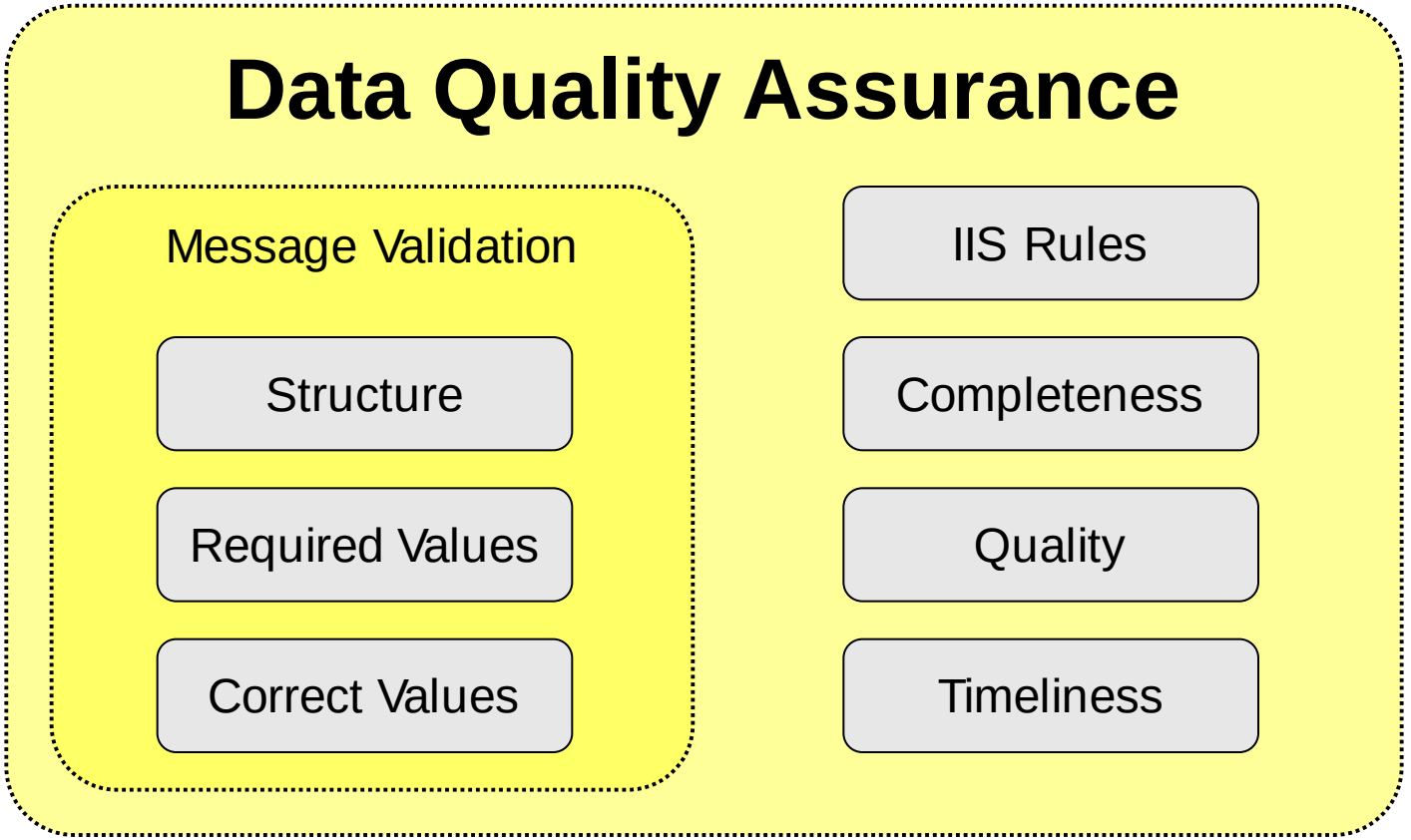
| | |
|---|---|
| 2010 - present | Pine View Middle School |
| 2004 - 2011 | Woodside Elementary |

**Experience**

| | |
|---|---|
| 2012 | Lawn care specialist: Responsible for the care and maintenance of yard and landscaping. |
| 2011 - 2012 | Pet and plant care: Responsible for tending plants and animals while owners are out-of-town. |
| 2009 | Part owner of refreshment stand: Responsible for development of advertisement materials (sign board) and mixing the lemonade. |

- Analysis:
  - No spelling mistakes
  - No grammar problems
  - Well organized
  - Clear and concise
  - Looking great!
- Would you call Sam in for an interview?

# DQA #2 More than Validation



# DQA #3 Errors and Warnings

- Error
  - An issue that invalidates an entire message.
  - Indicates a problem that must be fixed.
  - In HL7, creates a negative acknowledgment.
- Warning
  - An issue that is noticed, may be a data quality problem, but the message can still be accepted with the issue in place.
  - Ideally the sending system should review and resolve warning issues.

# Error and Warning Examples

- Errors
  - Patient date-of-birth is missing
  - Vaccination code is not recognized
  - Vaccination given before date-of-birth
- Warnings
  - Patient phone number is missing
  - Vaccination, not used in the US (for example OPV), is listed in historical record

# DQA #4 Finding Patterns

- Data quality assurance process can identify problematic patterns that only appear when reviewing a larger set of messages.
- Examples:
  - Submitter never sends phone number.
  - No last names are longer than eight characters.
  - Submitter never sends MMR vaccinations.

# DQA #5 Integrated and Ongoing

- Data quality assurance process must be a integrated process that is used for all incoming data, before and after going to production.
  - New interfaces need to measured with a consistent quality assurance process.
  - Once standards are meet and interface is approved data from the new interface can flow to production.
  - Interface should should continue to be measured by the same quality assurance process.
  - This process needs to be automated as much as possible in order to reduce staff time and fatigue.

# What We Did

# Open Source DQA

- In 2011 an open source Data Quality Assurance engine was created by Texas IIS, ImmTrac and MCIR.
- Accepts HL7 vaccination update (VXU) messages and returns acknowledgment (ACK).
- Focused on data quality and IIS business rules and not message validation.
- Creates DQA report and DQA score for batch submitted and weekly reports.

# DQA Issues

- DQA has a list of issues that it looks to identify in each message.
- IIS defines a profile that indicates the status of each issue:
  - Error
  - Warning
  - Accept
  - Skip

# DQA Report and DQA Score

- Standardized report with standardized score
- Three areas of measurement:
  - Completeness
  - Quality
  - Timeliness
- Contains no patient identifiable information.
- Clean HTML format, can easily be emailed or displayed by IIS.

# DQA Score

- **90-100 Excellent** meets and exceeds expectations
- **80-89 Good** meets expectations
- **70-79 Okay** meets expectations, improvements may need to be made
- **60-69 Poor** does not meet expectations, improvements likely need to be made
- **0-59 Problem** does not meet expectations, improvements must be made

# Demonstration of DQA Report

# Scoring Summary

## Scoring Summary

| DQA Score | Description |
|---|---|
| 75 | Okay |

| Measurement | Score | Description | Weight |
|---|---|---|---|
| Completeness | 93 | Excellent | 50% |
| - Patient | 99 | Excellent | 22% |
| - Vaccination | 87 | Good | 22% |
| - Vaccine Group | 92 | Excellent | 5% |
| Quality | 70 | Okay | 40% |
| - No Errors | 100 | Excellent | 28% |
| - No Warnings | 0 | Problem | 12% |
| Timeliness | 0 | Problem | 10% |

- Score and description represents score of the sub section
- Weight indicates how much this score affects the overall score

## Data Received

| Received | Count | Percent |
|---|---|---|
| Patients | 100 | |
| Next-of-Kins | 100 | |
| Vaccinations | 200 | |
| - Administered | 100 | 50% |
| - Historical | 100 | 50% |
| - Deleted | 0 | - |
| - Not Administered | 0 | - |

- Data can duplicated
- Administered: given by submitter
- Historical: reported by submitter
- Deleted: previously sent but should be deleted
- Not Administered: Not actually given for some reason

# Completeness

## Completeness

Completeness measures how many required, expected and recommended fields have been received and also indicates if expected vaccinations have been reported.

### Score

| Completeness Score | Description |
|---|---|
| 93 | Excellent |

| Measurement | Score | Description | Weight |
|---|---|---|---|
| Patient | 99 | Excellent | 45% |
| Vaccination | 87 | Good | 45% |
| Vaccine Group | 92 | Excellent | 10% |

# Patient Completeness

**Patient**

| Patient Fields | Score | Description | Weight |
|---|---|---|---|
| Overall | 99 | Excellent | |
| Required | 100 | Excellent | 16% |
| Expected | 96 | Excellent | 4% |
| Recommended | 100 | Excellent | 2% |

# Completeness Categories

- **Required** must be sent in every message, every time
- **Expected** should be sent normally but there are some occasional scenarios where a value is not expected
- **Recommended** should be sent if possible
- **Optional** could be sent but it is either not expected or not necessary, optional fields are noted but neither count nor detract from scoring

# Completeness Scoring Example

| Required | HL7 | Count | Percent | Description | Weight |
|---|---|---|---|---|---|
| Patient Id | PID-3 | 100 | 100% | Excellent | 3.5% |
| First Name | PID-5.2 | 100 | 100% | Excellent | 1.7% |
| Last Name | PID-5.1 | 100 | 100% | Excellent | 1.7% |
| Birth Date | PID-7 | 100 | 100% | Excellent | 3.5% |
| Sex | PID-8 | 100 | 100% | Excellent | 1.7% |
| Address | PID-11 | 100 | 100% | Excellent | 0.7% |
| - Street | PID-11 | 100 | 100% | Excellent | 1.7% |
| - City | PID-11 | 100 | 100% | Excellent | 0.4% |
| - State | PID-11 | 100 | 100% | Excellent | 0.4% |
| - Zip | PID-11 | 100 | 100% | Excellent | 0.4% |

# Completeness Scoring Example

- **Required** short description of the field
- **HL7** short reference to where in the HL7 message this data should be sent
- **Count** the number of times this field had a non-empty value
- **Percent** the percent of the total opportunities for this field to have a value
- **Description** indicates score status
- **Weight** how this field score affects the final DQA score

# Vaccine Group

- **Expected** vaccination is normally administered and is expected to be sent
- **Recommended** vaccination is often administered and is normally expected to be sent
- **Optional** may be given, but not routinely or not by all providers
- **Unexpected** not routinely given as part of general patient care in the US

# Vaccine Group Example

**Vaccine Group**

| Expected | CVX | Label | Count | Percent |
|---|---|---|---|---|
| DTaP | 20 | DTaP | 10 | 10% |
| Hep B | Problem: no vaccines received for this group | | | |
| Polio | 10 | IPV | 12 | 12% |
| Hib | 49 | Hib (PRP-OMP) | 5 | 5% |
| | 48 | Hib (PRP-T) | 14 | 14% |
| Influenza | 141 | Influenza, seasonal, injectable | 11 | 11% |
| | 140 | Influenza, seasonal, injectable, preservative free | 10 | 10% |
| MMR | 94 | MMRV | 2 | 2% |
| Varicella | 94 | MMRV | 2 | 2% |
| | 21 | varicella | 2 | 2% |
| Pneumococcal | 133 | Pneumococcal conjugate PCV 13 | 9 | 9% |

# Quality

- A good score indicates that there were not too many error or warnings
  - **No errors** a score of 100 indicates that there were no errors
  - **No warnings** a score of 0 indicates that there were more warnings than is expected by MCIR
- Expectations:
  - No more than a 1% message error rate
  - No more than 10% message warning rate

# Timeliness

- Measures the number of days between
  - The date of the latest administered vaccination in a single message
  - The date the message was received by MCIR
- Messages that only contained historical immunizations are not considered
- Older vaccinations reported in the same message with newer vaccinations are not considered

# Timeliness Categories

- **Early** wow that was fast!
- **On Time** received just on time
- **Late** received later than required
- **Very Late** received much later than required or expected
- **Old Data** ancient history

# Codes Received

- List of values received
- Current status of value:
  - **Valid** value is good to send
  - **Invalid** value should never be sent
  - **Unrecognized** value is unknown to MCIR
  - **Deprecated** value is recognized but a newer, better value should be sent (rare)
  - **Ignored** value represents a concept that is ignored and normally skipped by MCIR (rare)

# Vaccine Group

**Vaccine Group**

| Expected | CVX | Label | Count | Percent |
|---|---|---|---|---|
| DTaP | 20 | DTaP | 10 | 10% |
| Hep B | Problem: no vaccines received for this group | | | |
| Polio | 10 | IPV | 12 | 12% |
| Hib | 49 | Hib (PRP-OMP) | 5 | 5% |
| | 48 | Hib (PRP-T) | 14 | 14% |
| Influenza | 141 | Influenza, seasonal, injectable | 11 | 11% |
| | 140 | Influenza, seasonal, injectable, preservative free | 10 | 10% |
| MMR | 94 | MMRV | 2 | 2% |
| Varicella | 94 | MMRV | 2 | 2% |
| | 21 | varicella | 2 | 2% |
| Pneumococcal | 133 | Pneumococcal conjugate PCV 13 | 9 | 9% |

# Contact Information


- Nathan Bunker - Nathan.Bunker@gmail.com

- More information about DQA:
  - http://sourceforge.net/projects/ois-dqa/
  - http://www.openimmunizationsoftware.org/