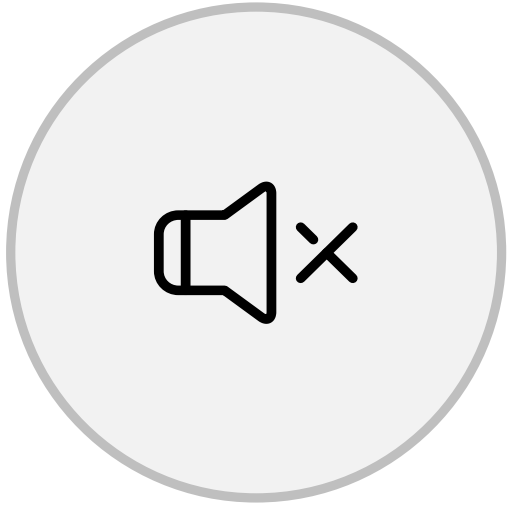
Monitor Data Quality by Provider with the Data at Rest (DAR) Tool - A Pilot Update

Discovery Session
June 22, 2020
4pm EST

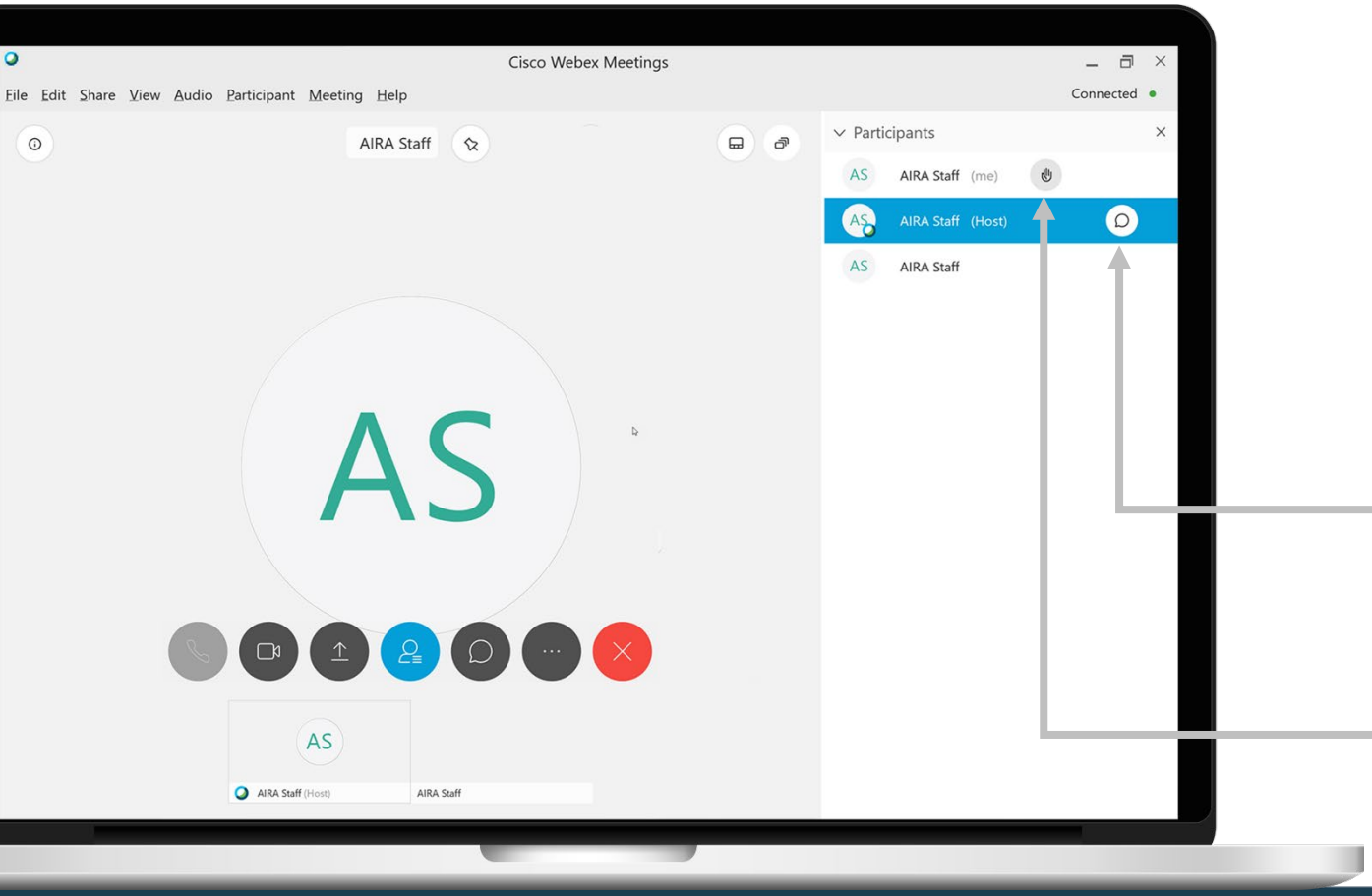
AIRA Discovery Session



All phone lines
are muted

This meeting is being recorded
and will be posted on the
AIRA repository

AIRA Discovery Session



- **How do I ask a question?**
  - There will be time allotted for Q&A following each of the updates, to unmute your line press *6
  - Via WebEx:

Select the chat icon next to the host and type question into the chat box.

Select the hand icon next to your name and you will be called on.

Today’s Topics

- Data at Rest Overview
- Initial pilot with Michigan’s
- Utah’s experience

Today’s Speakers

- Eric Larson, AIRA
- Josh Hull, Michigan
- Jon Reid, Utah
- Tom Romney, Utah

Project Overview

Data at Rest is the measurement of data residing in the IIS database regardless of how it arrived there.

This content area puts into practice data quality indicators found in the May 2018 IIS Data Quality Practices document.

For the pilots, AIRA is measuring:

- 24 Completeness Indicators
- 18 Validity Indicators
- 2 Timeliness Indicators

Example Data Quality Indicators

Completeness Indicators

- The percent of records containing a first name
- The percent of records containing a gender

Validity Indicators

- Lot numbers that violate lot number patterns
- More vaccines than expected for patient age

Timeliness Indicators

- Percent of patient records created in less than X days
- Percent of vaccination records created in less than X days

The 30,000 Foot View

Data at Rest in Four Steps

Extract

- IIS staff will extract a pre-defined cohort of patients from their production database into a flat file.

Transform

- IIS staff will run an AIRA-provided tool on their extracted data. This tool will assess for data quality indicators. The output of this tooling will be a set of aggregate detections void of identifiable patient data.

Load

- The aggregate file generated during the Transform step will be uploaded to AIRA for analysis.

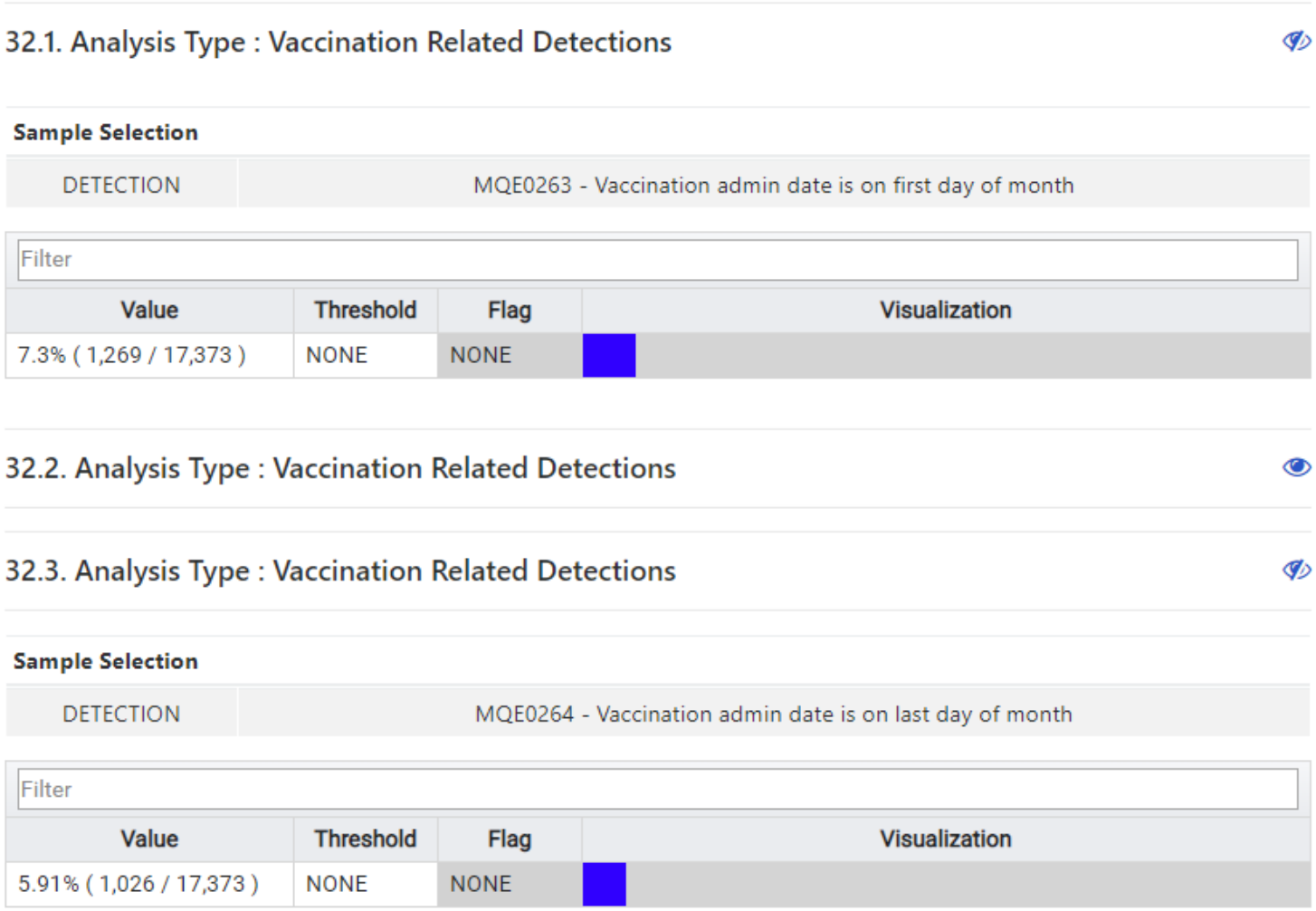
Analyze

- AIRA will analyze and produce individualized reports for the IIS to review and use to improve their data quality.

IIS-Wide Report

The example to the right is highlighting a potential data entry problem for this IIS where data is heavily skewed to the 1st and last day of the month.

This may or may not indicate an actual data quality problem, but it is statistically unlikely that 13.2% of all records (7.3% on 1st and 5.9% on last) were administered on two days of the month.



32.1. Analysis Type : Vaccination Related Detections

Sample Selection

DETECTION	MQE0263 - Vaccination admin date is on first day of month

Filter

Value	Threshold	Flag	Visualization
7.3% (1,269 / 17,373)	NONE	NONE	

32.2. Analysis Type : Vaccination Related Detections

32.3. Analysis Type : Vaccination Related Detections

Sample Selection

DETECTION	MQE0264 - Vaccination admin date is on last day of month

Filter

Value	Threshold	Flag	Visualization
5.91% (1,026 / 17,373)	NONE	NONE	

Provider Breakdown Report

Taking the previous example a bit further, we can look at the same data quality detection by provider.

As can be seen on the right, three providers are contributing to the high rate of 1st day of the month vaccination dates.

This report can now be used to reach out to specific providers to improve the quality of their data.

DETECTION	MQE0263 - Vaccination admin date is on first day of month

Filter

PROVIDER	Value	Threshold	Flag	Visualization
b62b604c7d914efe924b762fc32d440a	0% (0 / 3)	0%	BELOW	
12d2698a56a4db9bdf06033f2e83a395	0% (0 / 3)	0%	BELOW	
075f83ca228c47ac4a4c39826373c9ac	0% (0 / 8)	0%	BELOW	
5d497f9c085350f8451bc6a366a6b05d	0% (0 / 14)	0%	BELOW	
a7c7b157c51f232c832f6311cc3c30c8	7.24% (1,247 / 17,217)	0%	ABOVE	
1ee6993539e172b82649d570f621ca72	0% (0 / 40)	0%	BELOW	
ea57e9600aae601a66f8c74a45ee0298	80% (8 / 10)	0%	ABOVE	
ed5ed25b477d3da20ac147c598b63279	100% (14 / 14)	0%	ABOVE	
b657683c019fcf185c5cd4d5a4babad3	0% (0 / 7)	0%	BELOW	
54413b34bb901a0adaa516f6a6358fb7	0% (0 / 21)	0%	BELOW	
28826a8e762c2f0fe6fbe275c3b2745b	0% (0 / 3)	0%	BELOW	
c62270dab45496e8da533f48ce6c41a3	0% (0 / 21)	0%	BELOW	
62344d4467afad32d1231e366e6a62dd	0% (0 / 12)	0%	BELOW	

DAR Pilots

- Pilot #1
  - Michigan (two times)
- Pilot #2
  - Alaska
  - California
  - Colorado
  - Indiana
  - Oregon
  - Vermont
  - Utah
  - Washington

Forward Progress!

DAR Pilots: 9 IIS

Extract

- IIS staff will pull a pre-defined cohort of patients from their production database into a flat file.

Transform

- IIS staff will run an AIRA-provided tool on their extracted data. This tool will assess for data quality indicators. The output of this tooling will be a set of aggregate detections void of identifiable patient data.

Load

- The aggregate file generated during the Transform step will be uploaded to AIRA for analysis.

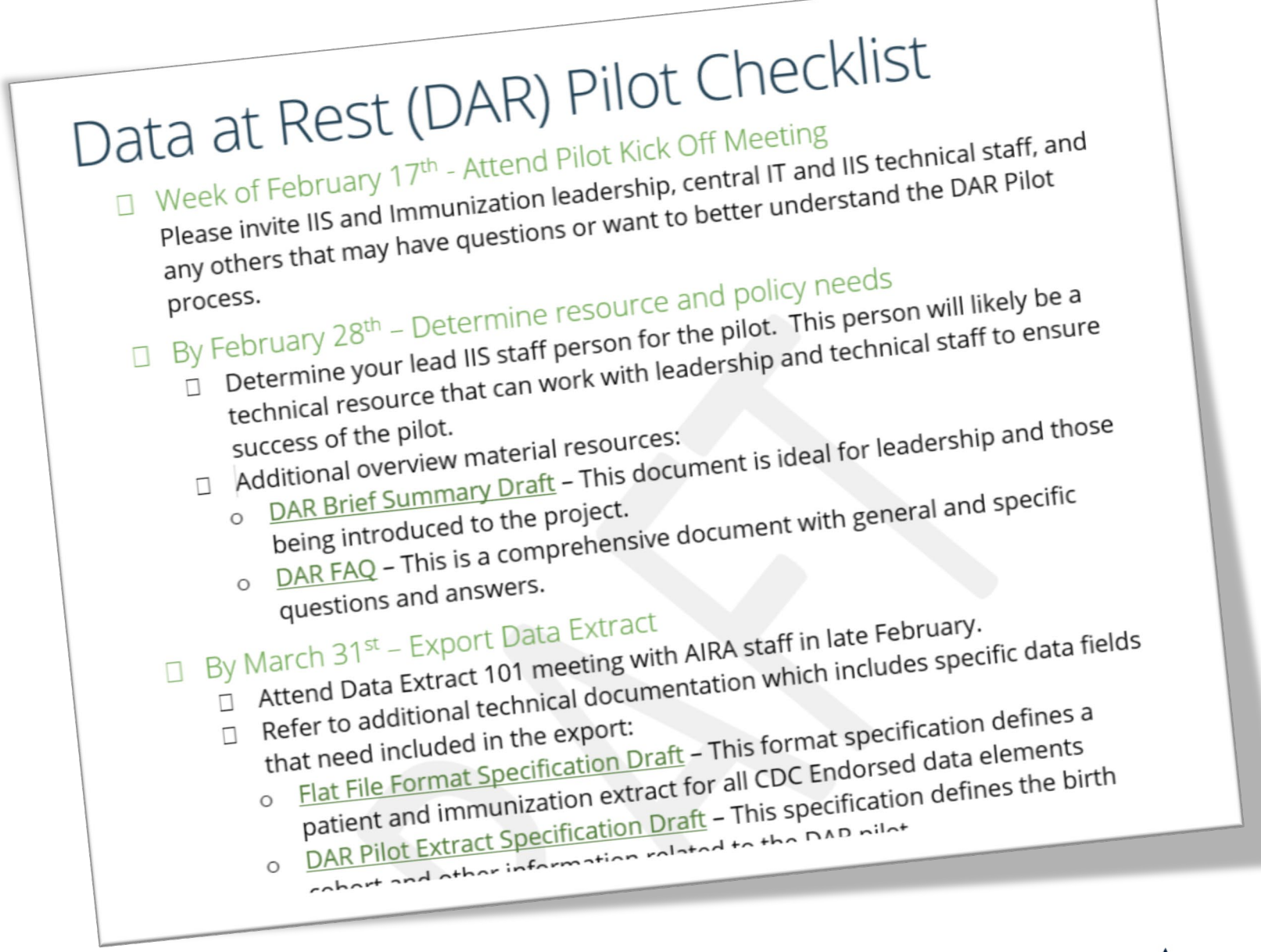
Analyze

- AIRA will analyze and produce individualized reports for the IIS to review and use to improve their data quality.

Pilot Checklist

One-stop reference guide for

- Key activities within each step
- Links to key artifacts



Data at Rest (DAR) Pilot Checklist

☐ Week of February 17th - Attend Pilot Kick Off Meeting

Please invite IIS and Immunization leadership, central IT and IIS technical staff, and any others that may have questions or want to better understand the DAR Pilot process.

☐ By February 28th – Determine resource and policy needs

- ☐ Determine your lead IIS staff person for the pilot. This person will likely be a technical resource that can work with leadership and technical staff to ensure success of the pilot.
- ☐ Additional overview material resources:
  - DAR Brief Summary Draft – This document is ideal for leadership and those being introduced to the project.
  - DAR FAQ – This is a comprehensive document with general and specific questions and answers.

☐ By March 31st – Export Data Extract

- ☐ Attend Data Extract 101 meeting with AIRA staff in late February.
- ☐ Refer to additional technical documentation which includes specific data fields that need included in the export:
  - Flat File Format Specification Draft – This format specification defines a patient and immunization extract for all CDC Endorsed data elements
  - DAR Pilot Extract Specification Draft – This specification defines the birth cohort and other information related to the DAR pilot

DRAFT

Future Directions

2020	2021	
		 
Testing and Discovery	Define Measures	Begin Assessment

MICHIGAN + DAR

Experiences being an early pilot

Everyone keeps asking!

- **1 Hour** Initial read through documentation
- **1 Day** Brainstorming and experimenting with query possibilities
  - ** This includes coming up with a repeatable strategy*
- **2 Days** Running Queries, validating output, running the tool, tweaking, and repeating
- **1 - 2 hours** Every few weeks running the query and DAR tool to debug, work out issues in the DAR side of things, and do pilot-y things.

Estimated total: **~30 Hours**

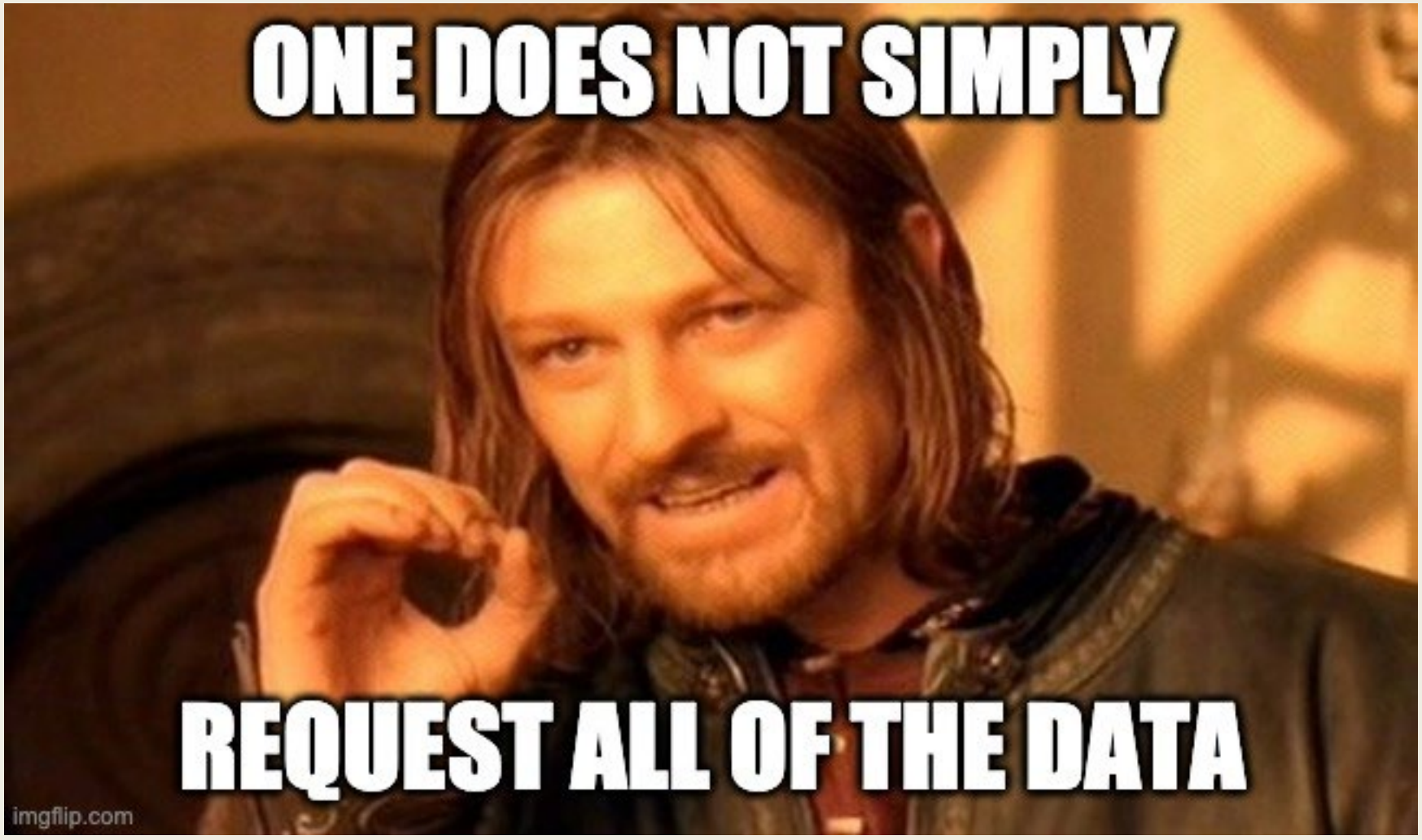
Pilots should reduce the effort for future users

- Discovery
- Debugging
- Tooling
- Documentation

What is it that we’re being asked to do?

- 95% of effort = “Extract” masked data from your database into a text file
- 5% of effort = Run the DAR tool

Extract the Data?



What is an extract, and how do you get one?

What is it?

- A text file
- Contains data from your data source
- Data separated and masked in a way that can be understood by another program, or person.

How do you get it?

- Run a SQL statement, save the results into a file.
- A myriad of more difficult and complicated options.

All of the data? No, just a Cohort

A subset of the whole, chosen for a specific characteristic relevant to the study

```
child.BIRTH_DT ⩾ '20160101' and child.BIRTH_DT ⩽ '20171231'
```

What are Masks?

Extract definition says things like:

- *Patient Name – First*
  - Populate with either [[VALUE_PRESENT]] or [[VALUE_NOT_PRESENT]]
- *Mother's Name – Middle*
  - Default this field to [[EXCLUDED]] for all records.
- *Record Creation Date*
  - Populate with actual value (no time component) from DB.

Patient Name – First : John

Output to Extract : VALUE_PRESENT

Extract Files

Patient Extract:

12345678900 [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_NOT_PRESENT]] [[VALUE_PRESENT]] 2016-01-01 [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_PRESENT]] [[VALUE_NOT_PRESENT]] [[VALUE_NOT_PRESENT]] [[VALUE_PRESENT]] [[VALUE_NOT_PRESENT]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] 2016-01-06 [[EXCLUDED]]

Extract Files

Vaccination Extract:

12345678900 186081818 50012491919 [[EXCLUDED]] [[EXCLUDED]] [[VALUE_NOT_PRESENT]] [[VALUE_NOT_PRESENT]] 45 [[NOT_COLLECTED]] 2016-01-03 UNK 01 [[VALUE_NOT_PRESENT]] [[EXCLUDED]] V00 [[EXCLUDED]] [[EXCLUDED]] [[EXCLUDED]] 2016-01-06 [[EXCLUDED]]

IIS Patient ID - 1234567890
IIS Vaccination Event ID - 2234567890
Reporting Group - 3234567890
Sending Organization - [[EXCLUDED]]
Responsible Organization - [[EXCLUDED]]
Administered at Location - [[VALUE_NOT_PRESENT]]
Vaccine Administering Provider [[VALUE_NOT_PRESENT]]
Vaccine Type (CVX) - 21
Vaccine Type (NDC) - [[NOT_COLLECTED]]
Vaccine Administration Date - 2017-01-25
Vaccine Manufacturer - MSD
Vaccine Lot Number - M012345
Vaccination Event Record Type - 01
Vaccine Route of Administration - SC
Vaccine Site of Administration - RT
Vaccine Expiration Date -
Vaccine Dose Volume - [[VALUE_PRESENT]]
Completion Status - [[EXCLUDED]]
Dose Level Eligibility - V01
VIS - Identifier - [[EXCLUDED]]
VIS - Publication Date - [[EXCLUDED]]
VIS – Given Date - [[EXCLUDED]]
Record Creation Date - 2017-08-11
Record Update Date -[[EXCLUDED]]

The DAR tool

Process the Extract

- You tell the tool where the extract files are
- The tool starts counting things in the extract
- It creates a file of what it counted

What then?

- You share the file of what was counted with AIRA.
- Do what you want to with the extract file (Delete, archive, etc)

The tool counts things

Aggregate Detections File Content


```
{
  "analysisDate" : 1573712155079,
  "extraction" : {
    "Patient – Mother's Name – Last Name" : {
      "valued" : 0,
      "excluded" : 0,
      "notCollected" : 0,
      "notExtracted" : 0,
      "valuePresent" : 200648,
      "valueNotPresent" : 27783,
      "valueLength" : 0,
      "empty" : 0,
      "total" : 228431
    },
    "Patient – Address – State" : {
      "valued" : 0,
      "excluded" : 0,
      "notCollected" : 0,
```

The tool counts things

Age Groups

Age Group	Count
0d -> 1 year	111784
1 year -> 2 years	116647
2 years -> 3 years	0
3 years -> 4 years	0
4 years -> 5 years	0
5 years -> 6 years	0
6 years -> 7 years	0
+	0

The tool counts things

Extraction Completeness

Element	Valued %	Excluded %	Not Collected %	Not Extracted %	Value Present %	Value Not Present %
Patient - Mother's Name - Last Name	0.00 %	0.00 %	0.00 %	0.00 %	**87.84 %**	**12.16 %**
Patient - Address - State	0.00 %	0.00 %	0.00 %	0.00 %	**100.00 %**	0.00 %
Patient - Responsible Party - Name - Middle Name	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %
Vaccination - Responsible Organization	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %
Patient - Ethnicity Code	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	**100.00 %**
Patient - Email Address	0.00 %	0.00 %	0.00 %	0.00 %	**0.15 %**	**99.85 %**
Vaccination - Vaccine VIS - Presented Date	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %
Patient - Name - Middle Name	0.00 %	0.00 %	0.00 %	0.00 %	**88.60 %**	**11.40 %**
Vaccination - Administered at Location	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	**100.00 %**
Patient - Provider Facility Level	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %
Patient - Birth Multiple Indicator	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %
Vaccination - Record Update Date	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %
Vaccination - Sending Organization	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %
Patient - Mother's Name - Middle Name	0.00 %	0.00 %	0.00 %	0.00 %	0.00 %	**100.00 %**
Vaccination - Complete Status	0.00 %	**100.00 %**	0.00 %	0.00 %	0.00 %	0.00 %

The DAR tool

THE SCARY PART – What are we sharing?

- The counts, encrypted

Aggregate Detections File Content


```
{
  "analysisDate" : 1573712155079,
  "extraction" : {
    "Patient – Mother's Name – Last Name" : {
      "valued" : 0,
      "excluded" : 0,
      "notCollected" : 0,
      "notExtracted" : 0,
      "valuePresent" : 200648,
      "valueNotPresent" : 27783,
      "valueLength" : 0,
      "empty" : 0,
```


Everyone Also Asks

How Big is the Extract?

- Cohort: Patients in MCIR birthdates between 1/1/2016 and 12/31/2017
  - *(2 years of birth dates)*
- Vaccination File:
  - *1.02 GB*
  - *4,338,846 Vaccination Records*
- Patient File:
  - *121 MB*
  - *228,431*

Everyone Also Asks

How long does it take to Run DAR tooling on that HUGE file?

4 h 33 m 48 s

*we talked about how to speed it up after this run last NOV

Everyone Also Asks

Any Lessons Learned?

“Yesterday I processed our 250MB patient file, and 1GB vaccine file but… it took 4+ hours.

But I had errors in the extract”

Lesson #1 - Always verify that your format is correct by running a small file against the tool BEFORE you extract the whole thing

(aka – extract a small subset first and run through the process)

Everyone Also Asks

Any Lessons Learned?

Verify that you've mapped, and masked properly.

I noticed some errors even reviewing it for this presentation.

How did we feel about the process, and the output?

- It was a fun challenge
- It didn’t suck up a ton of time
- It was fun to collaborate with AIRA
- We feel prepared for using the tool when its in its final form
- We are really excited about the reporting, and potential for helping us identify areas for improvement

Did we accomplish anything for future DAR participants?

- We debugged a lot of things
- We identified some areas that needed to be improved
  - *Reporting*
  - *Speed*
  - *Clarity of Documentation*
- We made files available of my final versions of queries, and functions:
  - *Dar Filter Function (SQL)*
  - *Extract Query – Patient, and Vaccination (SQL)*

**These were made available for the second pilot. I hope we hear if it was useful!*
**I'd love to share more thoughts on the technical details of the extract making*

UTAH DEPARTMENT OF HEALTH

Data at Rest (DAR)

June 2020

Tom Romney
Business Analyst
Jon Reid
USIIS Program Manager

immunize.utah.gov/usiis

Overall Data Stats

Utah Statewide Immunization Information System (USIIS)

- Authorized by R386-800 (Immunization Coordination Rule)
  - Allows system to coordinate immunizations among healthcare providers to assure adequate immunizations and to avoid unnecessary immunizations.
- Op-out system
- Contains all ages
- Voluntary participations by providers
  - VFC immunizations must be reported
  - Pharmacies required by DOPL Vaccine Administration Protocol

- Total Identities: 5,176,289
- Total Vaccine records (2000-2020): 49,942,692
- Providers: 3163

Utah DAR Timeline

- 2-18-2020 – DAR Kick-off call
- 2-21-2020 – First draft of export SQL
- 2-24-2020 – Updated instructions from DAR Pilot Data Extract 101 call
  - Ran through 4 versions of the export script before final version
- 3-23-2020 – Test full export ran
  - Installed NIST tool to production Linux server
- 3-26-2020 – Send final export to AIRA

- Total time spent on creating data export script and installing the tool was 16-18 hours with 4 hours spent in meetings/calls

Takeaways

- Business staff have full access to our production database
  - We can run script and data extracts as needed
- We do not require partner agreements for non-identifiable aggregate data
- Limited access to production servers
  - Able to extract and run applications
- We have imbedded IT staff in our IIS team
  - When needed our IT staff can review, install, develop custom solutions

- Tom is very familiar with data in database, experience with exporting data and files
- Ran several small test exports and updated script as needed
- Contacted AIRA when we had questions

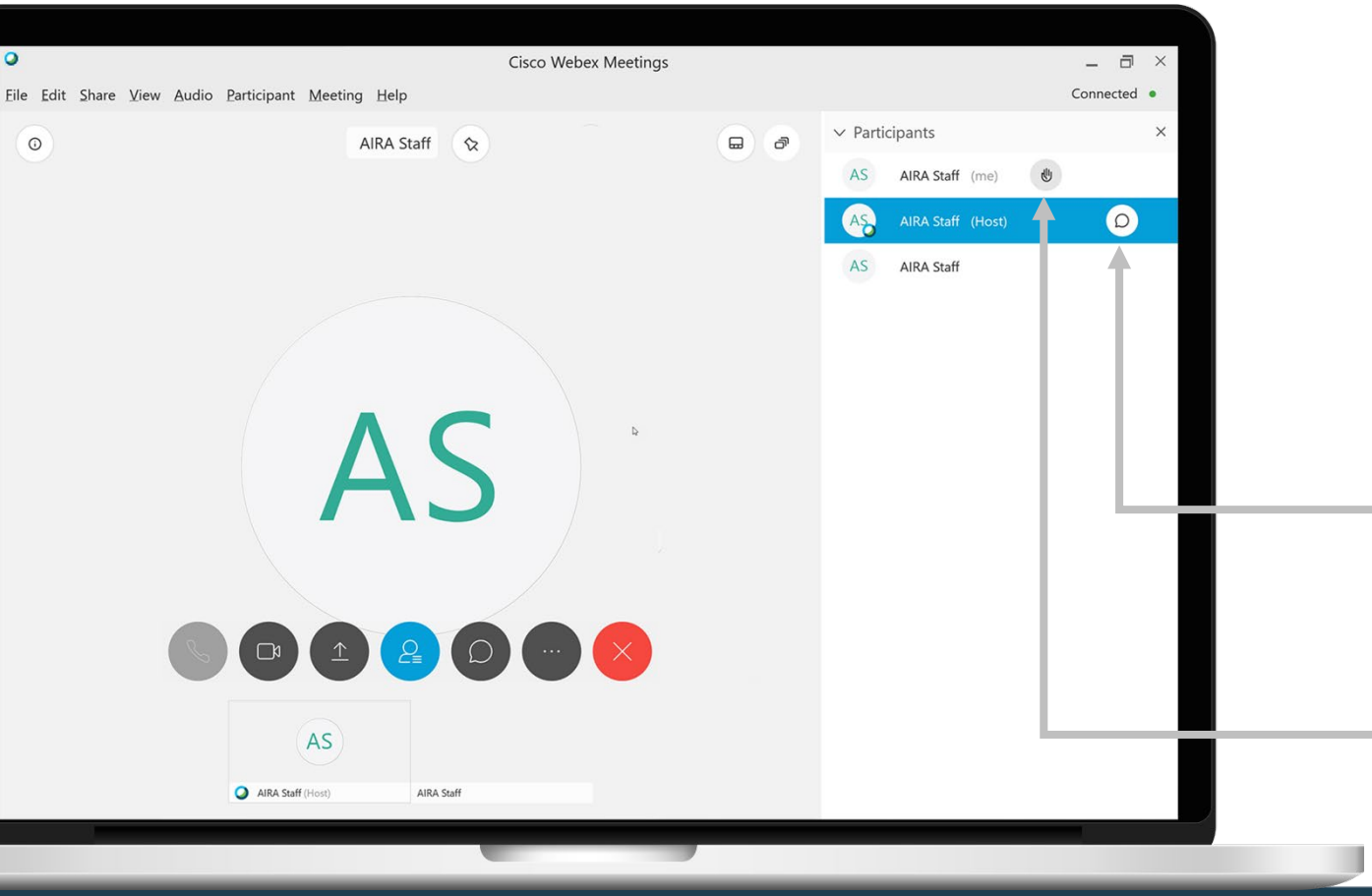
Questions

Tom Romney
Business Analyst
Jon Reid
USIIS Program Manager

immunize.utah.gov/usiis

Questions, Comments, Discussion?

Questions, Comments, Discussion?



- **How do I ask a question?**
  - To unmute your line press *6
  - Via WebEx:

Select the chat icon next to the host and type question into the chat box.

Select the hand icon next to your name and you will be called on.

Thank you to our presenters, and thanks to all of you for joining us!

A brief evaluation survey will be sent out following this webinar

The next Discovery Session will be July 27th at 4pm ET