N3C Data Quality Gates Focus Group Kickoff

Davera Gabriel

5May2020

Invitees

Clair Blacketer (Janssen, OHDSI)	Raju Hemadri (NCATS)	Christian Reich (IQVIA, OHDSI)
Christopher Chute (JHU)	Stephanie Hong (JHU)	Usman Sheikh (NCATS)
Mariam Deacy (NCATS)	Dazhi Jiao (JHU)	Sai Simhadri (NCATS)
Shahim Essaid (OHSU)	Jeff Klann (Harvard, ACT)	Harold Solbrig (JHU)
Tricia Francis (JHU)	Kristin Kostka (IQVIA, OHDSI)	Shyam Visweswaran (UPitt, ACT)
Ed Hammond (Duke, HL7)	Harold Lehmann (JHU)	Anita Walden (CD2H, SageBio)
Davera Gabriel (JHU)	Sandeep Naredla (NCATS, Adeptia)	Jarrett Waller (JHU)
Ken Gersing (NCATS)	Matvey Palchuk (TriNetX)	Griffin Weber (Harvard)
Melissa Haendel (CD2H, OSU)	Emily Pfaff (UNC, PCORNet)	Tanner Zhang (JHU)
		Richard Zhu (JHU)

How We Got Here

- Research Protocol
  - Chris Chute, March-Apr2020
- Initial Workflow Drafts
  - Davera Gabriel, 20APR20
  - Stephanie Hong, 4May20
- Data Ingestion & Harmonization WGMs
  - ‘Large’ & ‘Small’ Teams Meetings
  - March & April 2020
- Adeptia Demos / Technical Reviews
  - 13 April 11am EST - Demo
  - 22 April 3pm EST - Demo w/ workflow
  - 24 April 4:30p EST - JHU Site Configuration
  - 1 May 4pm EST – JHU “Deep Dive”
- NCATS Cloud Architecture / Configuration
  - 20 April 11am EST - Demonstration / Review
  - 28 April 11 am EST - JHU Site Configuration
- CDM Data Quality Process Reviews
  - PCORNet (on SAS) 14 April 1pm EST
    - Emily Pfaff & Adam Lee, UNC
  - ACT i2b2 16 April 4pm EST
    - Jeff Klann, Harvard & Shyam Visweswaran, U Pittsburgh
  - OHDSI (1) 20 April 3p EST
    - Claire Blacketter, Janssen, J&J; Kristen Kostka & Christian Reich, IQVIA
  - OHDSI (2) 28 April 10a EST
    - Claire Blacketter, Janssen , J&J
  - OHDSI (3) 30 April 11a EST *(Technical implementation)*
    - Kristen Kostka & Christian Reich, IQVIA
  - TriNetX – 4 May 4pm EST
    - Matvey Paulchuk, TriNetX
  - PCORNet (on R) - TBD

Proposal for Comprehensive N3C Data Quality Process

Goal: answer 3 questions for N3C

1. **What** DQ tests are desired / required?
   a) Utilize Kahn DQ Taxonomy / Framework
   b) Review one class at a time
2. **Which** available / combination of tools will be used?

   Review attributes of tools or processes that address each test

   a) Platform
   b) Portability
   c) Output / Handoff
   d) …
4. **When**, at what point(s) to conduct each test in the DI&H workflow?

   Assumes… workflow may change

Data Quality Taxonomy / Framework

- Conformance
  - Value
  - Relational
  - Computational
- Completeness
- Plausibility
  - Uniqueness
  - Atemporal
  - Temporal

EDM Forum

EDM Forum Community

eGEMs (Generating Evidence & Methods to improve patient outcomes)

Publish

9-11-2016

A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data

Michael G. Kahn
University of Colorado Anschutz Medical Campus, michael.kahn@ucdenver.edu

Tiffany J. Callahan
University of Colorado Anschutz Medical Campus, tiffany.callahan@ucdenver.edu

Juliana Barnard
University of Colorado Anschutz Medical Campus, juliana.barnard@ucdenver.edu

Alan E. Bauck
Kaiser Permanente Northwest, alan.bauck@kpchr.org

See next pages for additional authors

https://drive.google.com/drive/u/0/folders/1A9uWGjFRcLLEw-JLQ_z07eAwxQjzk-7e

N3C Additional DQ Requirements

- Multiple Source CDMs
  - *Plus* ETL for each
- Mapping & variance between sources
  - Concept mapping
  - Value Set alignment
- Transformation consistency
  - Did we map appropriately?
  - Did we apply maps consistently?

Available Tools / Methods

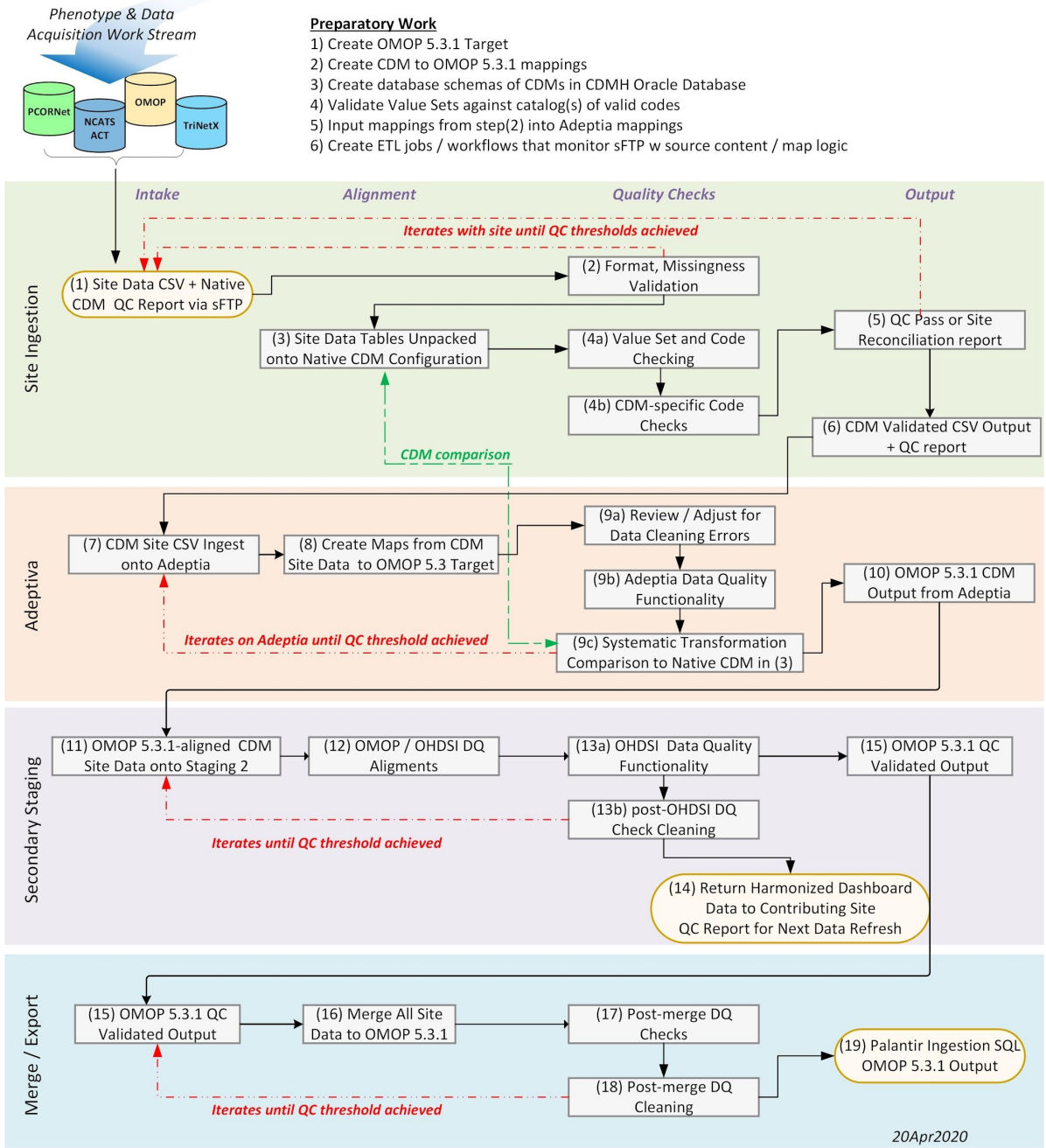
- Native CDM DQ processes
  - PCORNet
  - ACT
  - TriNetX
- SQL Code
- Adeptia
  - Process automation
  - Data & Map Validation
- Proteus
- OHDSI
  - Data Quality Dashboard
    - Kahn-based DQ tests of OMOP dB
  - ATLAS
    - Design / execute analytics on OMOP dB
  - ACHILLES
    - Data Characterization of OMOP dB
  - White Rabbit
    - ETL Prep / Support

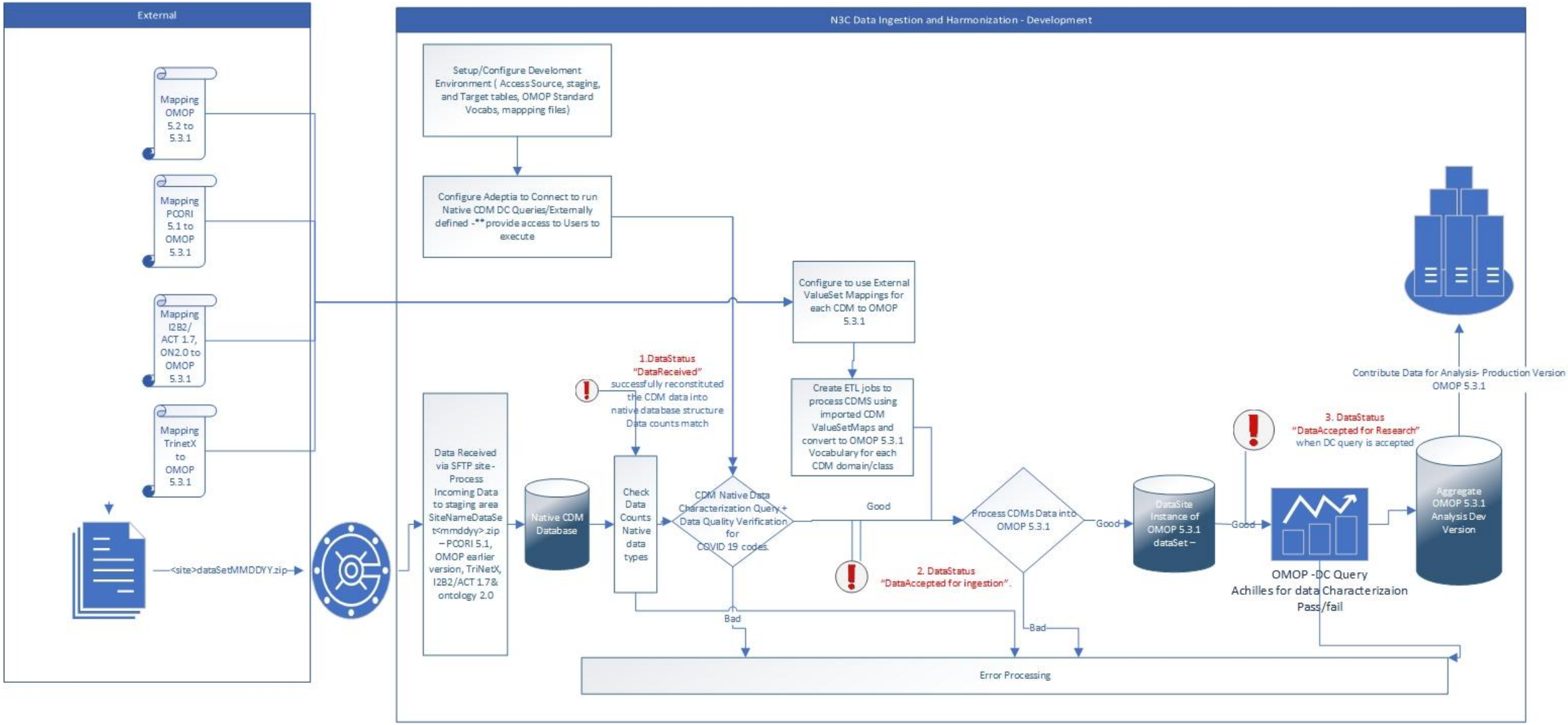
DQ Detailed Evaluation Process

1. Compile & agree to list of required DQ tests, aka “gates”
2. Review one gate at a time
   - Consider goal(s) for & granularity of N3C data set
   - Desired pass / fail threshold(s), implications
   - CDM SMEs envision / model DQ step for their respective CDM
     - *is it included in native CDM process?*
3. Review & contrast available tools, methods which can perform DQ gate
4. Agree on the optimum order of gate(s) in context of entire workflow
   - *may move as additional tests are considered*
   - *may occur in more than one position in the workflow*
   - *may require this step is repeated*
5. Aggregate & promote potential CDM extraction “pruning” upstream

Discussion

N3C Data Ingestion & Harmonization Workflow




External
Mapping OMOP 5.2 to 5.3.1
Mapping PCORI 5.1 to OMOP 5.3.1
Mapping I2B2/ ACT 1.7, ON2.0 to OMOP 5.3.1
Mapping TrinetX to OMOP 5.3.1
<site>dataSetMMDDYY.zip
N3C Data Ingestion and Harmonization - Development
Setup/Configure Develoment Environment (Access Source, staging, and Target tables, OMOP Standard Vocabs, mappping files)
Configure Adeptia to Connect to run Native CDM DC Queries/Externally defined -**provide access to Users to execute
Configure to use External ValueSet Mappings for each CDM to OMOP 5.3.1
Create ETL jobs to process CDMS using imported CDM ValueSetMaps and convert to OMOP 5.3.1 Vocabulary for each CDM domain/class
1.DataStatus "DataReceived" successfully reconstituted the CDM data into native database structure Data counts match
Data Received via SFTP site - Process Incoming Data to staging area SiteNameDataSet<mmddyy>.zip – PCORI 5.1, OMOP earlier version, TriNetX, I2B2/ACT 1.7& ontology 2.0
Native CDM Database
Check Data Counts Native data types
CDM Native Data Characterization Query + Data Quality Verification for COVID 19 codes.
Good
Bad
2. DataStatus "DataAccepted for ingestion".
Process CDMs Data into OMOP 5.3.1
Good
Bad
DataSite Instance of OMOP 5.3.1 dataSet –
Good
3. DataStatus "DataAccepted for Research" when DC query is accepted
OMOP -DC Query Achilles for data Characterizaion Pass/fail
Aggregate OMOP 5.3.1 Analysis Dev Version
Contribute Data for Analysis- Production Version OMOP 5.3.1
Error Processing