A recipe for shared services: Cloud, MPI and Governance

Facilitator: Bill Brand, MPH, Public Health Informatics Institute | May 21, 2021
Finding your way around Zoom

- The audience is muted.
- Type your questions into the chat or use reactions to communicate with presenters.
Implementation of an Integrated Reporting Portal to Improve Public Health Reporting at the Connecticut Department of Public Health

Gary V. Archambault, MS, Connecticut Tracking Coordinator
Nancy L. Barrett, MS, MPH, Infectious Disease Informatics Program Lead

Data Modernization Virtual Workshop
May 21, 2021
Outline

• Background
• Problem to be addressed
• Solution employed
• Secondary benefit
• Data lake/data factory
• Modernizing analysis, visualization and reporting
• Next steps
Background

- Connecticut Department of Public Health (CTDPH) uses the Conduent Maven application for reportable disease surveillance and case management (called CTEDSS)

- Maven is a hyper-configurable case management system
  - “Easy” to make changes as the need arises
  - Not as easy to query/report on the data contained within the database

- COVID-19 resulted in an approx. 3,000 percent increase in data volume
  - Lab test data (>10.5 million total)
  - Case report data (>2.5 million total)
  - CTEDSS SQL database approaching 0.5 TB in size
Problem to be addressed

• Office of the Governor required daily reports, including calculations of percent positivity of COVID-19 tests
• CTEDSS flexible data model but complicated data tier
  • Too much data to be handled efficiently by built-in Maven report/query tools
  • Required >31 custom reports/queries to extract data
  • Tried various ways of timing these extracts
  • Was taking over ten hours to complete
  • Generated 31+ CSV files for data analytics staff
• Resulted in delays in generating the daily report
  • Data was consumed but difficult to extract in a complete and timely manner
Initial solution

• Direct data pull from CTEDSS SQL database into on-premise (on-prem) SQL database in CTDPH Data Center
• Used SQL server integration service packages
  • Queries backend data
  • Automated start time
• Daily extracts for the COVID-19 data analytics team
  • Data flows from the CTEDSS replica database to the on-prem SQL server database
  • Daily report team rewrites “R Studio” code to work from on-prem SQL server
  • Eliminated the CSV data wrangling
  • First runs identified >300,000 ‘missing’ results
• Extracts ready for data analytic team by 4:00 a.m.
  • Started on 1/15/2021
  • >115 days with zero failures or delays

Problem solved!
Secondary benefits

• Data analytic staff time freed up to do more important work
  • Most days, the daily report is complete by 8:30 a.m.
• Less stress on staff
  • Start their day with data already in the integrated reporting platform
  • Some semblance of work/life balance
• Less stress on the CTEDSS Platform
  • Removes reporting/extract burden from application tier
  • Timed to run to minimize impact on database
• Use daily report extracts to generate CELR files for CDC
  • Improve timeliness and completeness of reporting
Building the integrated reporting platform

• Integrated reporting platform
  • DPH SQL server
  • DPH Azure data lake
  • PowerBI

• Populate with selected COVID-19 data from CTEDSS
  • Initially populate from the on-prem SQL database
  • Move to populating data directly for additional work
  • Other data sets to follow

• Process improvements
  • Full daily pulls to incremental pulls
    • Ready for production
    • Coordinating release
  • Serve data to the CT open data portal
Primary data lake uses

• Platform for analytics
  • COVID-19-related data
    • Surveillance (case/test data)
    • Contact tracing
    • Immunization
    • Vital records

• Platform for interoperability
  • Add COVID-19 vaccine information into the CTEDSS
    • Make the joins within the data lake
    • Send CTEDSS extracts with only the necessary vaccination information
    • CTEDSS used as source for Message Mapping Guide

• Platform for data sharing with CT local health departments
  • 65 Local Health Departments/Districts
  • Most query multiple systems daily for COVID-19 related line lists
  • Data Lake will maximize efficiencies and eliminate burden on primary systems
Modernizing analysis, visualization, and reporting (AVR)

• Data lake use will be expanded beyond COVID-19
• Use platform for integrated reporting
  • Multiple systems feeding data lake
  • Robust highly scalable environment
• Web-based data portals and dashboards
  • Migrating to single CTDPH environment
  • Data stored in data lake
  • Visualizations using PowerBI
  • Public dashboards
  • CT leadership dashboards
• Eliminate ‘data wrangling’ from epidemiologist job description
Next steps

• Governance
• Data security
• User management for non-CT.GOV users
• Develop the team to manage it all
  • Informatics
  • Information technology
Thanks to our teams

• CT Tracking Team: Gary Archambault, Tom Bavone, Patricia Przysiecki
• Epi Data Analysts: Alex Senetcky, Jess Brockmeyer
• Informatics Program: Nancy Barrett, Alycia McNutt, Carl Bondeson, Zack Fraser
• CT DPH IT: Steve McConaughy
• Deputy State Epidemiologist: Lynn Sosa
Questions?

For more information, please contact CTDPH informatics/DMI co-leads:

Gary Archambault  gary.archambault@ct.gov
Nancy Barrett  nancy.l.barrett@ct.gov
What is a Master Person Index (MPI)

- Strategy that links records from multiple sources together for the same person.
- Source data systems are not changed.
- Provides a single source of truth about person, so data can be updated and managed in one place.
Master Person Index (MPI) Project

- In Washington, several manual processes were established to address disease surveillance challenges of the COVID-19 response.

- Master Person Index Project (MPI) developed to reduce manual efforts, notably data linkage and deduplication.

- MPI presented opportunity to address immediate COVID-19 needs as well as lay the foundation for long-term development towards interoperability across several key Washington DOH data systems.

- MPI supports the goals of Washington DOH Data@Health modernization initiative that preceded COVID-19.

- Pre-COVID, MPI established as priority initiative of the Washington Health and Human Services Coalition (HHS Coalition) that includes multiple state agencies.
Example where MPI useful for COVID-19

- Person is identified as a contact of a patient with COVID-19
- Person tests negative for COVID-19 multiple times
- Person tests positive for COVID-19
- Person is interviewed as part of case investigation
Master Person Index Project Goals

- Reduce the burden of deduplication on our Washington disease reporting system (WDRS) and thereby help stabilize WDRS.

- Reduce manual efforts in WDRS and case risk and exposure surveillance tool (CREST), alleviating staff resource strain.

- Reduce manual efforts in tracking down missing/inaccurate contact information for individuals who test positive and their contacts, reducing the time to initiate case investigations.
WA Health and Human Services Coalition Partners

Ex-Officio Advisors

Office of Financial Management

Chief Information Officer
WA Health and Human Services Coalition Drivers

How did the HHS Coalition governance model come to be?

- HHS Coalition Agency Leadership Input
- Federal (CMS) Expectations
- Washington Governor & Legislature Expectations
Cross-agency MPI leadership

- Healthcare authority
- Department of health
- Other coalition partners: DCYF, DSHS, HBE
- Authorizing environment
Project Scope

UMPI is Verato’s Universal Master Person Index product

- Multi-source person record aggregation
- Identity resolution, via deduplication
- End user data stewardship

Enrich is Verato’s data enhancement product

- Contact information (e.g., email address, phone number, address)
- Social determinants of health (e.g., age, occupation, income)
Verato Enrich Helps You Build the Complete Picture

**Contact Data**
- Data like:
  - Name
  - Phone
  - Address
  - Birthdate

**Health Data**
- Data like:
  - Clinical
  - Claims
  - Labs
  - Prescriptions

**SDOH Data**
- Data like:
  - Race
  - Ethnicity
  - Occupation
  - Household

**Solution**
- Verato can enrich your person records with additional contact details like names, addresses, and phones.
- Verato can help link together health records from many data sources, even if records contain different, errored, out-of-date, and incomplete demographics.
- Verato can enrich your person records with SDOH data like race, ethnicity, and occupation.

Verato Proprietary. Do Not Copy, Reproduce, or Use Content Without Explicit Written Permission from Verato
Current State

- Contract, data sharing agreements and security design review complete
- Environments and transport mechanisms established
- Final WDRS/CREST architecture design complete for WDRS
- Development of link database design and development complete
- One time WDRS data load has been sent to Verato; awaiting results
- Auto deduplication process for WDRS matches of repeat lab reports
- HHS Coalition MPI workgroup reviewing conceptual architecture
Challenges

- Managing scope and expectations
- Accelerated timeline needed for DOH MPI solution
- Shared understanding of terms, tools and approaches across programs and agencies
- Leveraging staff expertise during emergency response
- Establishing an MPI link repository at DOH
- Balancing need for immediate solution while planning strategically for the future
- Security, privacy protection and data access
- Aligning agency and coalition needs
Successes

- Collaborative project management expertise at agency and coalition levels
- Partner agencies understanding of DOH emergent needs and barriers
- Focus on minimum viable product for implementation
- Inclusion of HHS Coalition workgroup members in DOH planning
- Strong communication across DOH and HHS workgroups
- Opportunity to move critical statewide needs forward
Questions?

Cathy Wasserman
cathy.wasserman@doh.wa.gov
To request this document in another format, call 1-800-525-0127. Deaf or hard of hearing customers, please call 711 (Washington Relay) or email civil.rights@doh.wa.gov.
The Clinical Community Data Initiative (CODI) Governance Infrastructure: 
Governance as a Shared Service

Emily Kraus, MPH, PhD
Consultant, Public Health Informatics Institute
May 21, 2021
What is the problem?

A large amount of health data is stored outside of clinical entities.
What is the problem?

Data scattered across a fragmented health and community landscape

Researchers and public health stakeholders looking at data from a single institution see only part of the picture
CODI is a partnership between CDC, the MITRE corporation, the Public Health Informatics Institute and local partners in Colorado and North Carolina to integrate clinical and community data for end users.

What is the objective of CODI?

Transform data scattered across a fragmented health and community landscape into individual-level longitudinal records that integrate information across clinical and community organizations.
How can patient records be linked?

Privacy Preserving Record Linkage (PPRL)

Unique Identifier (a.k.a. LINK_ID)
PPRL governance advantages

- Expert determination is used to show that hashes and the unique identifiers generated by PPRL are de-identified
- A third party organization performs the linkage
- Each contributing organization gets data on only their patients
- Privacy is protected because a repository of PII is never created and only de-identified data is shared
- Open source tools are available to support this linkage work
  - CODI Denver used Anonlink
PPRL governance needs

• Identify an organization to perform PPRL (the linkage agent)
• Define a process for PPRL
• Build consensus around identifiability of the unique identifier and the hashes and sensitivity of this information
• Identify the appropriate agreements to facilitate PPRL data exchange
• Document processes, policies, and approved/prohibited uses for the unique identifier
  • Can unique identifiers be reused for other purposes?
  • Can unique identifiers be used to re-identify individuals?
Governance tasks

• Assign roles and responsibilities
• Map out a detailed workflow
• Document approved and prohibited uses of the unique identifier
• Build processes to monitor and prevent prohibited uses
• Create and implement necessary agreements
PPRL workflow example

Key workflow information:
- What data are being exchanged?
- To whom are the data being sent?
- How will the data be sent?
- What will the receiving organization do with that data?
- How will the receiving information store and secure the data?
- What data will be returned?
Lessons learned

• Record linkage and PPRL are complex concepts and methods that require a common terminology
• Governance artifacts from other projects are highly portable and translatable when the functions are similar
• Technical, project, legal, and contract staff had to acquire a detailed understanding of the record linkage process
• Understanding and addressing sensitivities around the unique identifier was crucial
CODI PPRL Governance as a Shared Service
CODI partners could reuse PPRL for another project or expand record linkage to other partners or other data source because:

• Organizations understood the requirements for PPRL and record linkage
• A record linkage service was established (linkage agent) including connectivity with multiple data contributors
• Processes were established to conduct record linkage
• Agreements existed with a record linkage service to share hashed data
• Data quality tools and processes had been built to examine the quality of the link
Acknowledgements

Alyson Goodman
Ray King
Renee Porter
Megan Harrison
Pradeep Podilla
Dawn Heisey-Grove
Andy Gregorowicz
Peter Mork
Ken Scott
Rachel Zucker
Art Davidson
Jim Jellison
Matt Daley
Sara Deakyne Davies
Dan Chudnov

The findings and conclusions in this presentation are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control & Prevention.
Learn more about the CODI project: www.cdc.gov/obesity/initiatives/codi/community-and-clinical-data-initiative.html
Questions?

Emily Kraus, MPH, PhD
Consultant, Public Health Informatics Institute
Next steps

• Continue the discussion on the message board (link provided in the chat)

• Five-minute break: 2:45– 2:50 PM EST

• Next session: 2:50 – 3:20 PM EST
  • Workshop reflections
Thank you.