CRISP

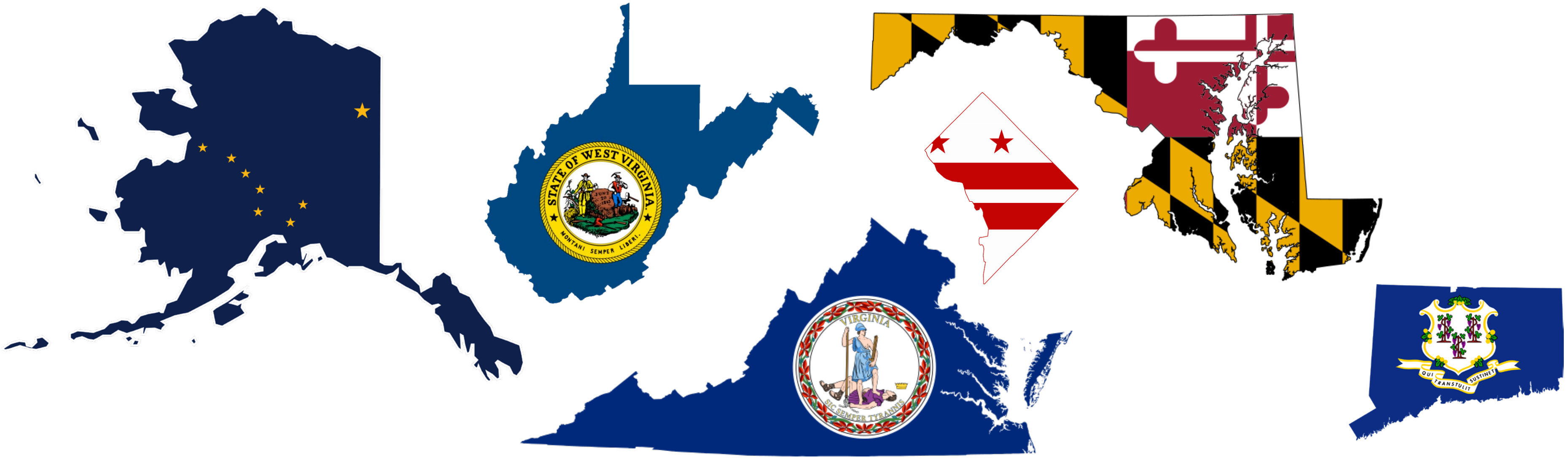
SHARED SERVICES

Data-Quality at Scale:

Rapidly Evolving and Expanding our Master Patient Index with Large Datasets

About CRISP Shared Services

The mission of CRISP Shared Services (CSS) is to: **assist member organizations in achieving economies of scale, pooling innovation efforts, and implementing best practices.**



CRISP Shared Services is a non-profit support organization:

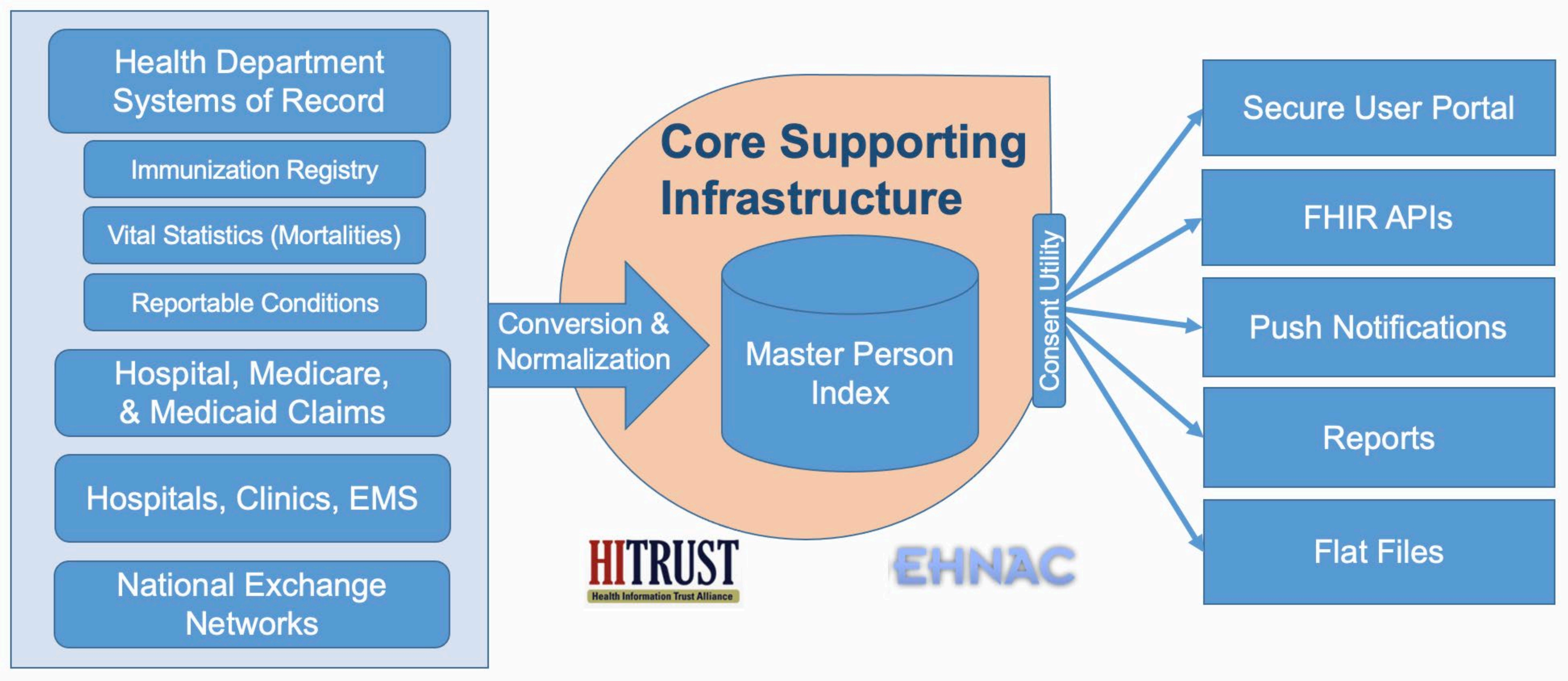
- Data, prioritization, funding, and relationships remain locally-controlled
- Each affiliate organization participates in the governance
- Technology is designed for reuse and operational efficiency

Supporting Health Data Utilities

- An HDU is a statewide entity with the advanced capabilities "to combine, enhance, and exchange electronic health data across care and services settings for treatment, care coordination, quality improvement, and public health and community health purposes."[1]
- An HDU works directly as a member of the community to balance stakeholder needs and meet or exceed local privacy regulations.
- An HDU receives data from many source systems in any available format, conducts normalization and linkages, provides secure cloud storage, and sends outputs through standards-based integrations, analytic files, and applications.
- An HDU provides all data to every authorized receiver.

[1] https://www.civitasforhealth.org/wp-content/uploads/2022/12/Civitas-MHCC-HDU-Brief_FINAL_2022-15-12.pdf

Simplified Technology Overview


Health Department Systems of Record
Immunization Registry
Vital Statistics (Mortalities)
Reportable Conditions
Hospital, Medicare, & Medicaid Claims
Hospitals, Clinics, EMS
National Exchange Networks
Conversion & Normalization
Core Supporting Infrastructure
Master Person Index
Consent Utility
Secure User Portal
FHIR APIs
Push Notifications
Reports
Flat Files
HITRUST
Health Information Trust Alliance
EHNAC

Core technological components include a master person index; flexible and secure exchange layers; standards-based APIs; data lake for converting, normalizing, and mastering; performant datasets

Evolution of the Data Lake

Services	February 2020	February 2023
States	3	6
Entities (EID)	21m	43m
Source and MRN	150m	240m
Microservices architecture	Yes, serving patient access	Yes, serving patient access
Bulk Analytics capability	Limited for clinical data	Comprehensive Data Lake
COVID	Not yet	Testing and Immunization Data for a significant percentage of the population of several States
Ability to detect and improve MPI	Limited	Significant
Ability to bulk load MPI	Available, but single-threaded	Multi-threaded and tuned to MPI capacity
Copy of MPI	Twice a week for analytics	Daily copy and real-time bulk access

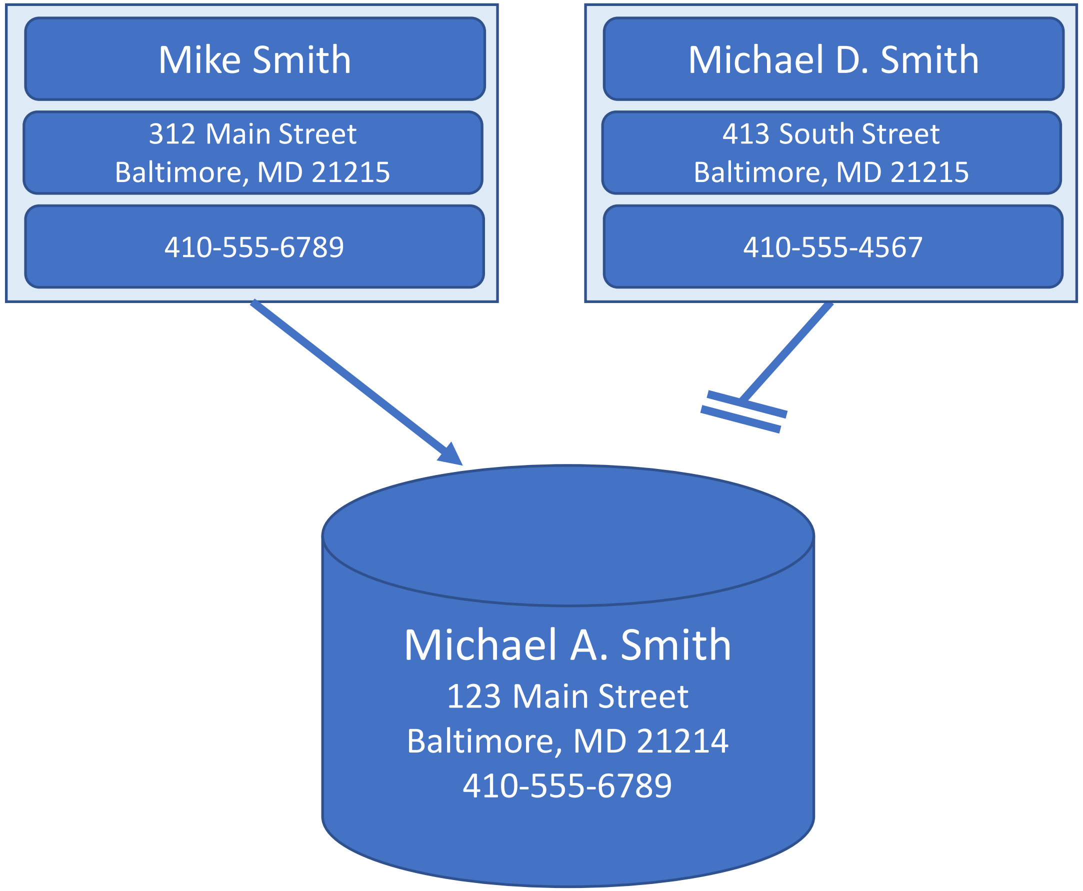
First Steps

- Had funding pre-approved for a data warehouse in Jan 2020
- Interviewed vendors to get us started in Feb 2020
- Launched production Azure/Sparc/Databricks Data Lake in April 2020
- Compelling COVID use case drove adoption and incremental growth.
- Ingestion pipelines initially created to support contact tracing, race/ethnicity enhancement, latest phone numbers, hospitalization status
- Launched hourly contact tracing in May 2020
- Very early on, we realized that the interface with the Master Patient Index (MPI) was critical to success

How the Master Person Index Works

To link patient identities across hundreds of source systems, the Master Person Index conducts probabilistic matching:

- Each feed contains first name, last name, date of birth, address, email, phone, etc.
- Points are assigned as discrete elements match; once the points reach a threshold then the identities link
- The system accounts for typos, common name iterations, reversed first/last names, and phonetic spelling



Storing Address History and Prior Names

Entity (EID)

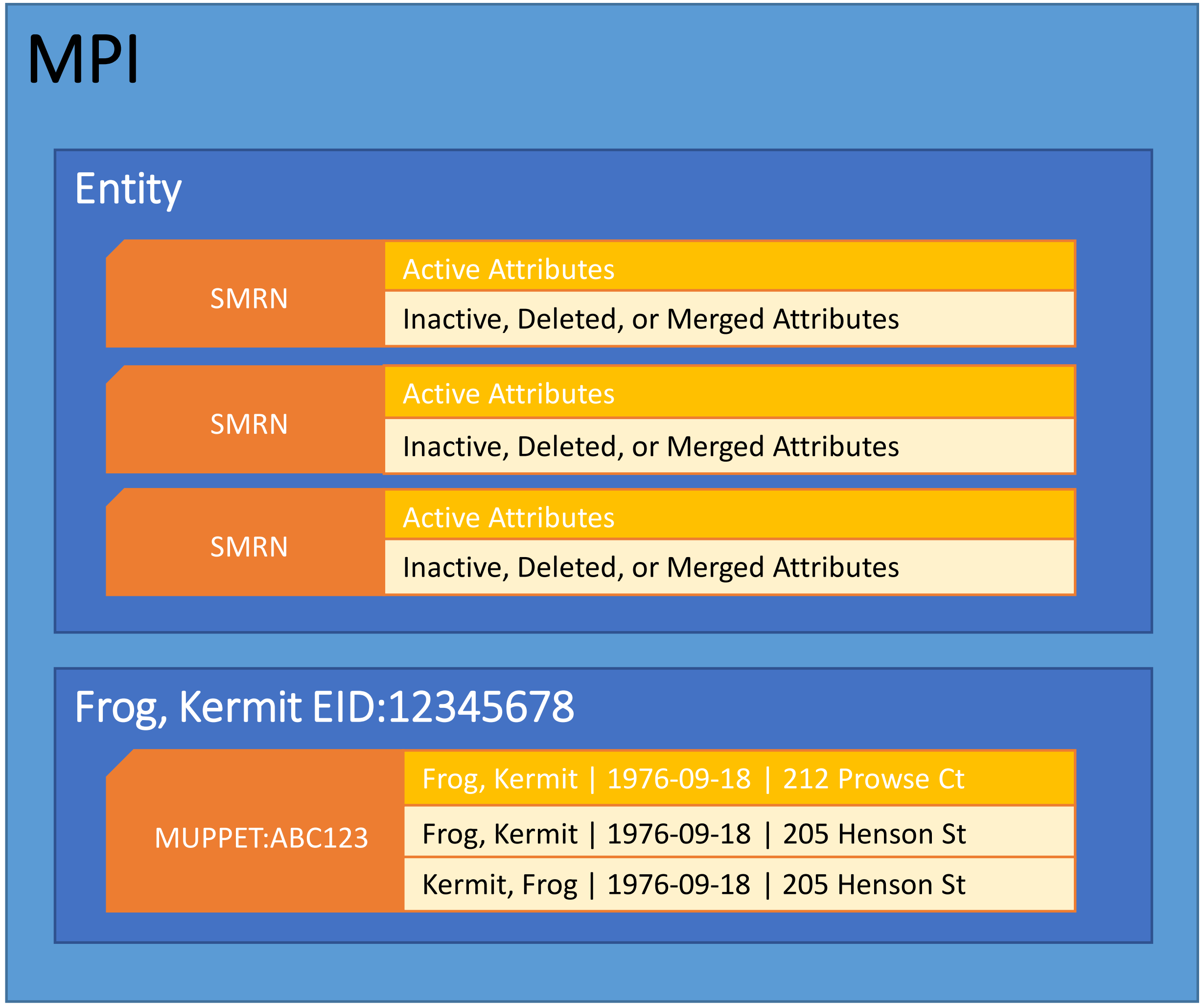
= collection of SMRNs matched together by the MPI based on active and inactive demographic data

Member / SMRN / Record

= demographic information for a single patient from a single source, basically a row in the “entity table”

Attribute

= the MPI categorizes demographics as distinct attributes to facilitate matching



MPI

Entity

SMRN
- Active Attributes
- Inactive, Deleted, or Merged Attributes

SMRN
- Active Attributes
- Inactive, Deleted, or Merged Attributes

SMRN
- Active Attributes
- Inactive, Deleted, or Merged Attributes

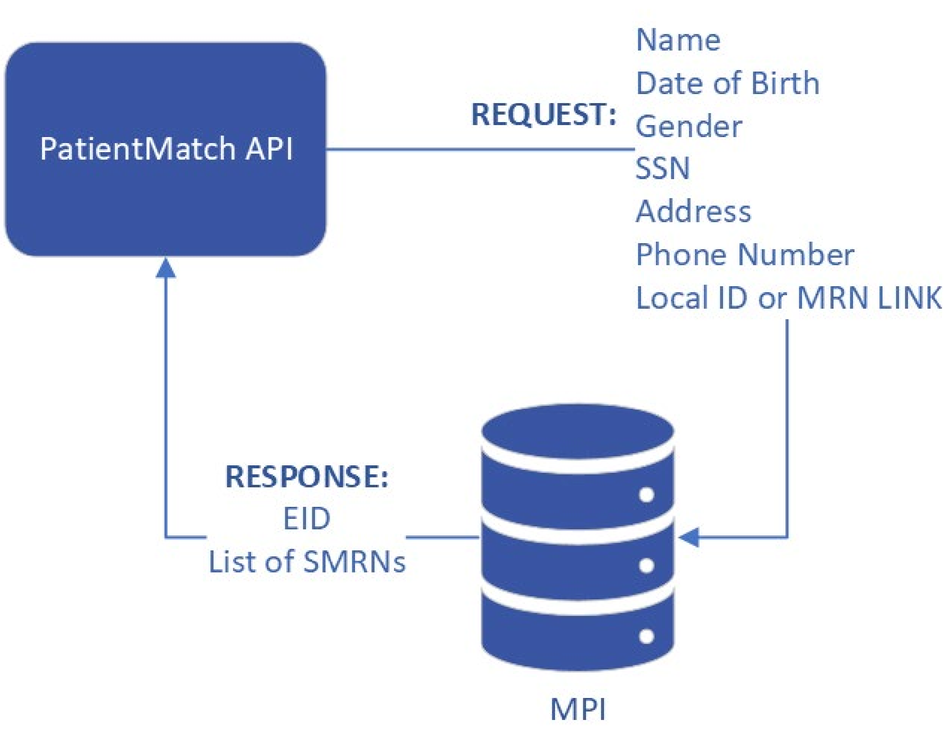
Frog, Kermit EID:12345678

MUPPET:ABC123
- Frog, Kermit | 1976-09-18 | 212 Prowse Ct
- Frog, Kermit | 1976-09-18 | 205 Henson St
- Kermit, Frog | 1976-09-18 | 205 Henson St

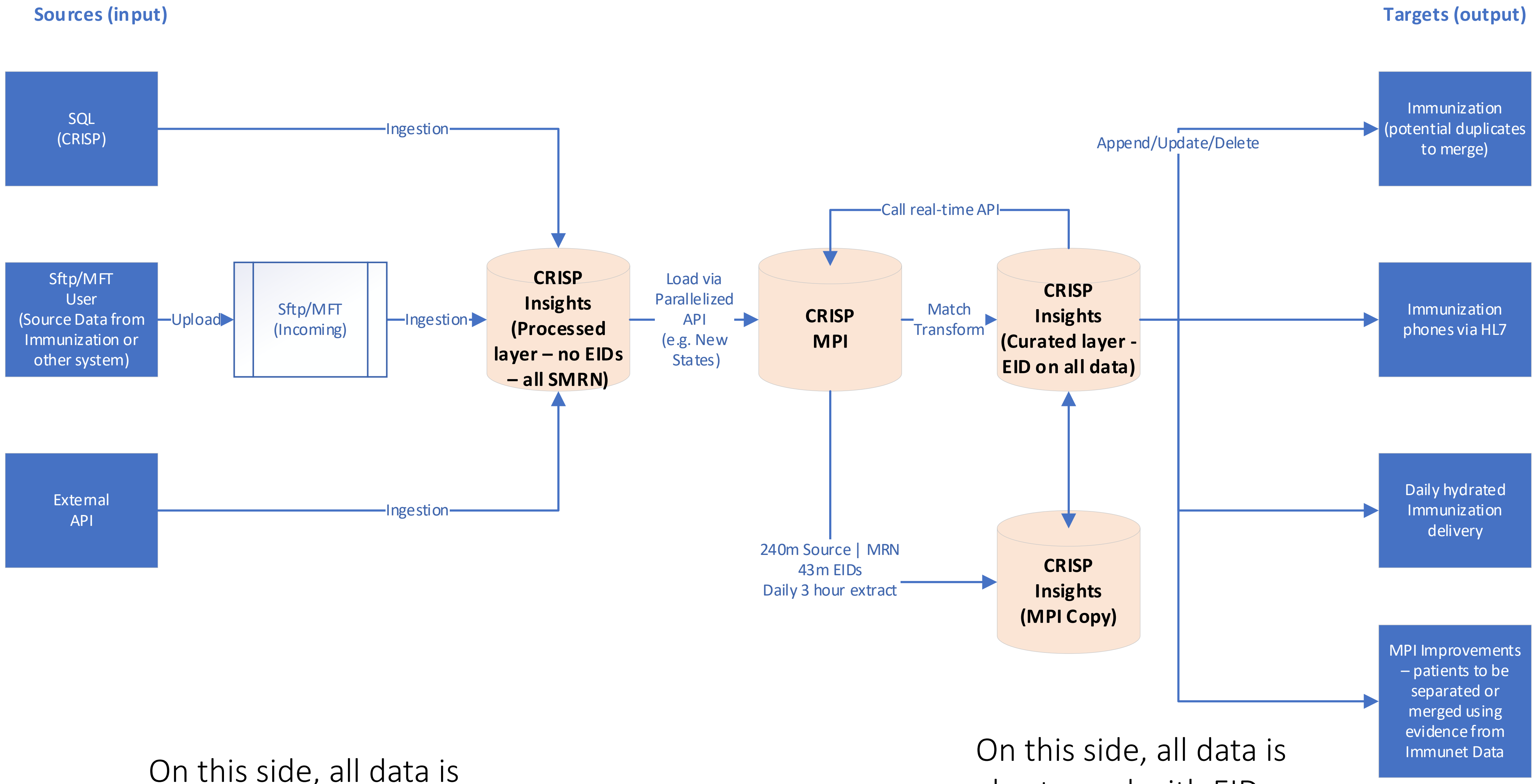
Interacting with a Master Person Index

- About 1bn times a month, we read or update the MPI with new data. From the data lake, a typical call is run – pass in demographics and the MPI returns the entity and all matching sources with demographics that match

PatientMatch: Lookup by Demographics



Using Data Lake to pre-build matched daily datasets



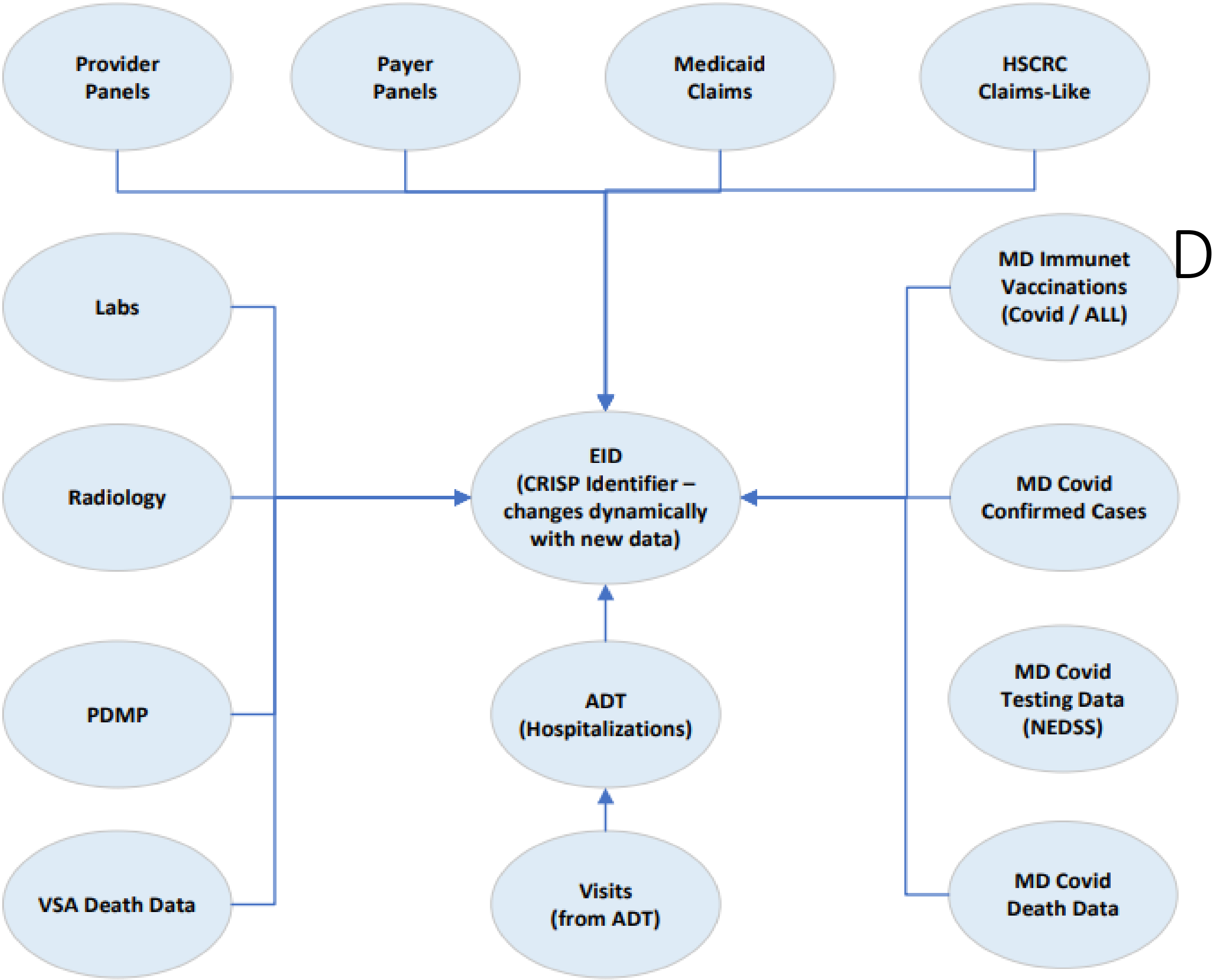
Sources (input)
Targets (output)
SQL (CRISP)
Ingestion
Sftp/MFT User (Source Data from Immunization or other system)
Upload
Sftp/MFT (Incoming)
Ingestion
External API
Ingestion
CRISP Insights (Processed layer – no EIDs – all SMRN)
Load via Parallelized API (e.g. New States)
CRISP MPI
Call real-time API
Match Transform
CRISP Insights (Curated layer - EID on all data)
240m Source | MRN
43m EIDs
Daily 3 hour extract
CRISP Insights (MPI Copy)
Append/Update/Delete
Immunization (potential duplicates to merge)
Immunization phones via HL7
Daily hydrated Immunization delivery
MPI Improvements – patients to be separated or merged using evidence from Immunet Data

MPI allows detection of patients to be merged in source systems

On this side, all data is stored by Source + MRN

On this side, all data is also tagged with EID as of midnight prior day

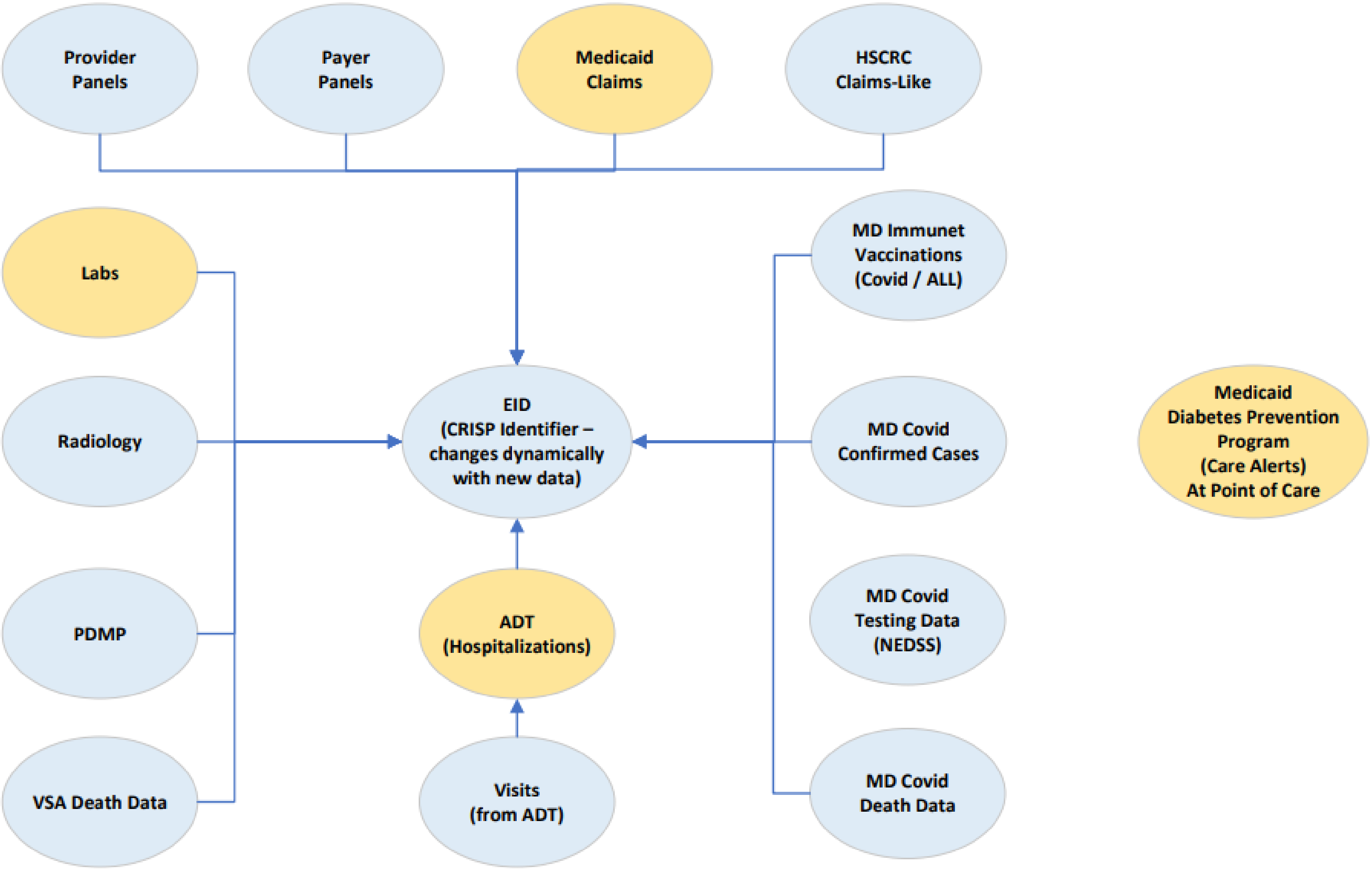
Aligning Data in the Lake by EID



Data Types

- Clinical/Participant
- Claims
- Public Health

Use Case: Bulk Matching for Diabetes Prevention


Provider Panels
Payer Panels
Medicaid Claims
HSCRC Claims-Like
Labs
Radiology
PDMP
VSA Death Data
EID (CRISP Identifier – changes dynamically with new data)
ADT (Hospitalizations)
Visits (from ADT)
MD Immunet Vaccinations (Covid / ALL)
MD Covid Confirmed Cases
MD Covid Testing Data (NEDSS)
MD Covid Death Data
Medicaid Diabetes Prevention Program (Care Alerts) At Point of Care

Current Initiatives

- Use Data Lake multi-threaded processing to bulk import very large new sets of patient data (e.g. New States)
- Use knowledge gained from operations to “pre-process” incoming data to protect the MPI from changes that currently need to be rolled back
- Use Data Lake to build operational reporting and troubleshooting to improve the quality of the Master Patient Index itself
- Rapid response analytics to questions where the data is already in the Data Lake, though follow clear data governance process and rigor
- Serving hundreds of automated outgoing pipelines with large datasets

Technical steps to achieve this

- Incrementally ingest datasets both local and remote (as of midnight)
- Have scaled and tuned bulk MPI database access
- Have scaled and tuned bulk MPI reads and writes by rearchitecting calls to go to a dedicated MPI application server for the Data Lake
- By offloading a lot of calls from the MPI by building the MPI copy for the Data Lake, this has enabled the MPI to continue to support operational calls at average 70ms – despite growth at 500k entities per month, and a massive increase in bulk analysis and reporting
- Build and invest in a talented team every day

Additional Production Capabilities Currently Live

Public Health Disease Investigation

- Immediately view records from across local and national providers via QHIN participation

Public Health Data for Providers and Medicaid

- Immediately notify a hospital if a patient presents with an antibiotic resistant infection
- Share weekly pediatric vaccination gaps with Medicaid Managed Care organizations
- Notify FQHCs when patients are subject to redetermination within 90 days

Prescription Drug Monitoring Reporting

- Enhanced overdose reporting by combining EMS, hospital, and fatality information

Federal and State Reporting

- Reuse data feeds to share information with state and federal disease registries

Public School Support

- Notify Medicaid of absenteeism potentially due to health care events
- Provide immunization records to a school system in bulk

Non-Medical Drivers of Health

- Integrate housing information systems with health care data to support care management
- Capture HIPAA Authorization/consent for sharing with appropriate community supports

CRISP
SHARED SERVICES

Thank you!

Craig Behm – craig.behm@crisphealth.org

Andy Hanks – andy.hanks@crisphealth.org