CHCF
DATA EXCHANGE EXPLAINER SERIES
MAY 2022

by Brian Dillon, MBA, and Alex Horowitz

Digital Identity Management for California's Health Data Exchange

The Opportunity

Meaningful statewide health information exchange requires that patient data from different organizations and information technology systems can be associated with a unique person. This capability requires an effective approach to digital patient identity. California Assembly Bill 133, Omnibus Health Trailer, sets a vision for data exchange and calls for a strategy by July 2022 for "unique secure digital identities capable of supporting master patient indices, to be implemented by both private and public organizations in California."[1] A successful approach will set up more uniform data standards and an efficient statewide structure that uses organizational capabilities at the local level.

Considerations for Digital Identity Management in California

California has unique characteristics to consider in developing the state's approach to digital identity management, including the state's existing data exchange and identity matching ecosystem, its large and diverse population, the complex and diverse provider landscape, and strong interest in protecting patient privacy.

- **Existing identity matching ecosystem.** California already has regional health information organizations (HIOs) and health information exchange (HIE) networks operating in the state,[2] including significant health information sharing that takes place within provider electronic health record systems. Functionally, these entities already manage patient digital identity at a regional, network, and substate

Core Digital Identity Definitions

Unique digital identity. A finite data set made up of enough unique attributes to identify a person and differentiate that person from other people.

Patient matching. The ability to associate patient information from multiple records and unaffiliated systems with the correct person.

Master patient index (MPI). A discrete set of patient demographic information stored in a database and used to associate disparate records with a unique individual. A MPI system often also includes data matching and data cleansing capabilities. When used across organizations, it is called an enterprise master patient index. A key policy choice is whether there is one or many MPIs in an exchange ecosystem — in other words, its level of centralization.

Data field. An individual data element considered part of an overall digital identity. Often, these data elements are demographic, but the principle can be extended to other data elements unique to an individual. Examples include first name, date of birth, and zip code.

Data standards. Guidelines that define standardized formats of data fields such as name or date of birth, which reduce errors and make data matching procedures more effective. Data standards can be applied to data fields to make data matching procedures more dependable and efficient.

level. While these capabilities are robust, each network has a unique approach to digital identity management. Accuracy within and across these networks is highly variable.

- **State population and demographics.** California is significantly larger than any other state. As the number of people in a data set increases, the uniqueness of any person decreases, and it becomes more difficult to make distinctive matches.[3] In addition, California is an exceptionally diverse state with subgroups of people who share racial and ethnic identities, and similar names and other data attributes, and often live in concentrated geographic areas. These regional variations can create what are known as "fragile populations,"[4] which often match incorrectly and may require special considerations. California's scale and diversity may create a need for more data fields per patient to find unique people accurately.
- **Implementation complexity.** The number and types of organizations that make up California's health ecosystem is vast. It includes payers, providers, and clinical and social service organizations, ranging from technologically well-resourced to bootstrapped. Any state-level requirements related to patient identity management will require changes to these organizations' IT systems and operating procedures, and even small and simple changes will have a significant impact across the state.
- **Patient privacy / data security.** Protecting patient privacy is a priority for California policymakers and consumer advocates. Approaches to digital identity, including whether and how to aggregate, store, and share patient data, need to be carefully considered in that context. Publicly available technology systems designed to bolster data security should also be considered.

Strategies and Tools Available to Policymakers

Policymakers have various tools available to craft the state's approach to identity management and to address the challenges presented by California's unique characteristics.

- **Network participation options.** Larger data volumes improve matching rates. Policymakers influence the robustness of identity management by making participation in data exchange optional or mandatory. Optional participation models with incentives for organizations to join can take many years to generate sizable data sets. On the other hand, they lessen the administrative burden on organizations in the delivery system. Mandated participation enables more robust identity management in a shorter time but may alienate resource-constrained provider organizations that cannot meet the onboarding requirements. A mandated approach with a phased implementation timeline is one approach to balancing these trade-offs.
- **Data standards.** Defining a common set of data elements that make up a digital identity and data formats that health organizations use to exchange information — also known as creating data standards — can help identity management. The more data fields required, the more complete the initial identity description and the more information available for matching. Implementation of data standards by network participants may require multiple years and detailed technical planning, including an approach to converting historical data. Heavy implementation requirements could worsen disparities among provider organizations with and without sufficient resources to meet any new mandates. Policymakers could define levels of data set requirements to be achieved by participants, depending on their organizational capabilities and resources. For example, a minimum data set, such as the United States Core Data for Interoperability (USCDI) v1,[5] could be established as a statewide

standard for demographic data, while state health and social identifiers could be required beyond the USCDI baseline for larger regional and substate organizations.

- **MPI architecture.** The overall design of a state's MPI system, or its MPI architecture, defines the logical and physical relationships between identity matching capabilities (e.g., patient identity databases, matching algorithms, established data standards, policies on information sharing) that exist within and between organizations at the local and regional levels and any capabilities managed by the state. One important policy choice for MPI architecture is the level of centralization in the system (Figure 1).
  - California's current MPI architecture is an example of a decentralized approach. Health systems and HIOs maintain their own MPIs and establish policies and procedures to query patient data from one another. Retaining data and managing exchange locally negates the need for centralized technical resources at the state level and may offer greater data security. However, the quality of patient identity management and efficiency of data sharing is limited by the capabilities of each organization.
  - Fully centralized MPI architecture positions the state as a hub where patient identity management capabilities reside and establishes the use of a unique patient identifier across network participants. A centralized repository of patient identity information, matching algorithms, and data cleansing capabilities are managed and supported by a large technical team with significant security controls.
  - Most state MPI architectures fall somewhere between a fully decentralized and fully centralized

Figure 1. Options for Master Patient Index Architecture Centralization

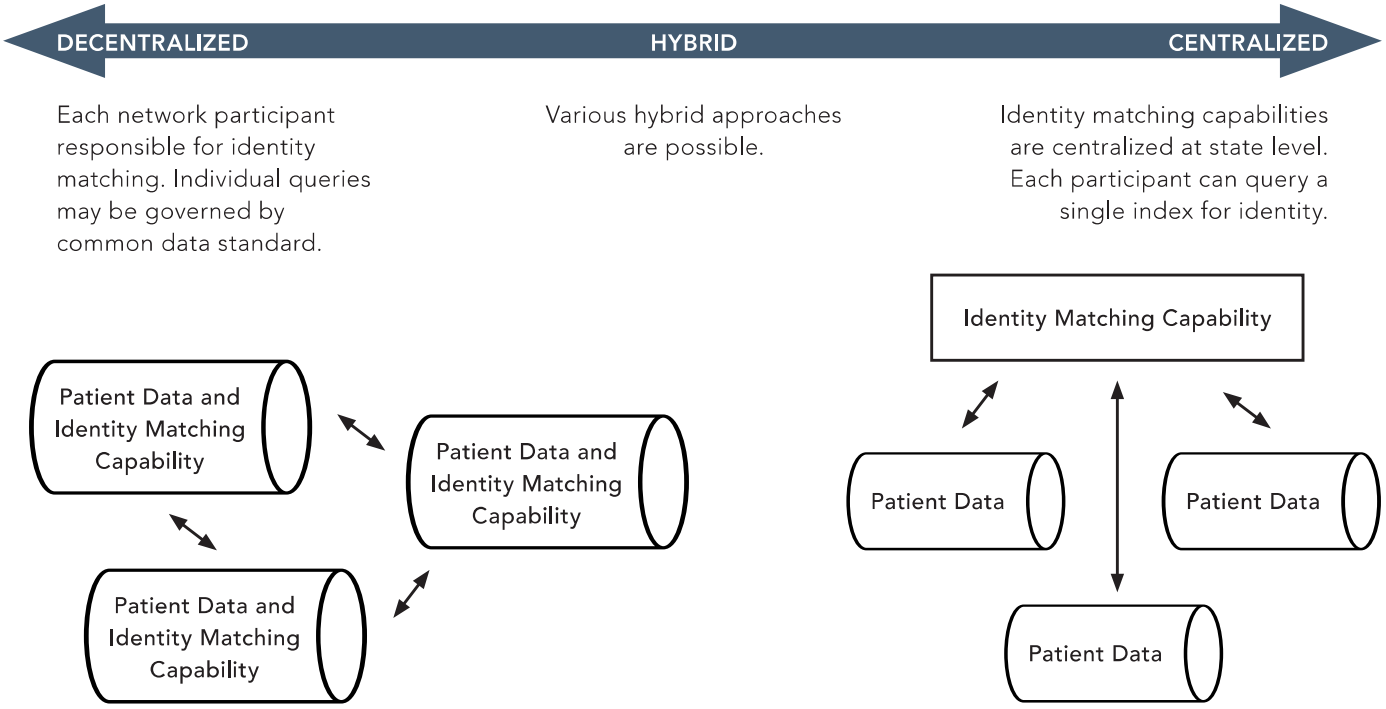

Illustration credit: Direct access storage cylinder by Arthur Shlain from NounProject.com.

approach. In a hybrid model, the state may, for example, supply data matching technology or support linking data but not enforce a unique identifier across network participants.

Matching Technologies

MPIs utilize a variety of technologies to match patient identities. Generally, patient matching compares the patient identity data of disparate records to link patient records and/or assigns a patient a unique identifier that must be used everywhere in an organization or ecosystem. Common technologies include these:

Algorithmic matching compares patients' demographic data (e.g., name, address, date of birth, phone number, social security number [SSN], gender) and other associated identifying information (e.g., federal identifiers like Medicare ID, health-related state program IDs, local identifiers like health plan member or medical record numbers) to match records. The MPI uses data to which it already has access to make the best potential matches.

Referential matching goes beyond the data that exist in the enterprise MPI or any participant databases, and supplements those data with public and semipublic demographic data (e.g., from credit reporting databases), including data that are not health related, to make matches.

A **unique identifier** unambiguously identifies a person; all records for that person are associated with one identifier. The unique identifier is most often a unique number but could also be achieved through a smart card with an encoded number, or through biometrics. **Tokenization** is a technology that replaces sensitive data like personally identifiable information (e.g., SSN) with a token. That token can then serve as a type of unique identifier, linking records with that common data element without exposing sensitive information.

State-Level Approaches to Digital Identity Management

Three state health information exchange organizations offer useful information about how patient matching can be conducted at scale: the Colorado Regional Health Information Organization (CORHIO), the Michigan Health Information Network (MiHIN), and New York eHealth Collaborative (NYeC) and the statewide HIE network it runs, the State Health Information Network-New York (SHIN-NY). (See Table 1 on page 5.) All three states centralize important aspects of data exchange governance, allowing them to direct data sharing policies, to establish data standards, and to ensure that a proper privacy and security framework is in place. However, they may not have a fully centralized approach to MPI architecture.

- **CORHIO** centralizes the storage of patient identities and matches them at the state level through a third-party vendor that uses an approach called "referential matching" and keeps a 350-million-person database to compare and enhance records. Matches are assigned an organizational identifier and considered a unique identity.[6]
- **MiHIN** uses a process called the "Common Key System" to link disparate patient data across participating organizations. When a query for patient data is sent to MiHIN, the organization checks for existing information among its networked members. If information exists, the patient data are associated with an existing organizational ID or "key" that serves as a unique identifier. If no data exist, a new key is created and assigned to be used in the future.[7]
- **NYeC's SHIN-NY** is organized into six connected regional health information organizations (RHIOs), also known as Qualified Entities (QEs), which collectively maintain 90 million records. Each NYeC QE has its own MPI system and staff who manage its population in a process like MiHIN's. Patients are assigned a unique ID within the QE and the centralized entity, SHIN-NY, links patient data across the QEs.[8]

Overall, Colorado centralizes the storage of patient identities and matches them at the state level while Michigan and New York merely connect entities searching for additional patient records across the network and do not permanently store information themselves. CORHIO and NYeC use referential matching in partnership with a third-party vendor, while Michigan relies entirely on data of its participating members. All three use minimum data requirements and data standards to support identity matching. See the table for more details about the three states' approaches.

Table 1. Characteristics of State Approaches to Digital Identity Management

	COLORADO (CORHIO)	MICHIGAN (MIHIN)	NEW YORK (NYEC)
State Population	5.8 million	9.9 million	20.2 million
Year MPI Implemented	2021	~2017	~2014
HIE Governance	Centralized at CORHIO	Centralized at MiHIN	Shared between RHIOs and NYeC
Data Exchange Participation	Optional	Optional	Required
MPI Approach	Centralized	Hybrid	Hybrid
Required Minimum Data Sets and Standards	Yes	Yes	Yes, varies by level
Highest Level of Patient Matching	State	State	State, but typically QE
Matching Rate	With referential matching, 79% match rate	When key assigned, 100% match rate	Not measured at state level

Interviewees

Colorado Regional Health Information Network
Ako Quammie, Vice President, Data Management and Quality

Michigan Health Information Network
Tim Pletcher, Executive Director

New York eHealth Collaborative
Michael Berger, Acting Chief Information Officer

About the Author

Brian Dillon, MBA, a consultant with Intrepid Ascent's Technology Strategy Group developed this fact sheet with Alex Horowitz, vice president of technology strategy at Intrepid Ascent. **Intrepid Ascent** supports communities on their paths to adopting scalable technologies with a human-centered approach focused on both individual experience and collective impact.

About the Foundation

The **California Health Care Foundation** is dedicated to advancing meaningful, measurable improvements in the way the health care delivery system provides care to the people of California, particularly those with low incomes and those whose needs are not well served by the status quo. We work to ensure that people have access to the care they need, when they need it, at a price they can afford.

CHCF informs policymakers and industry leaders, invests in ideas and innovations, and connects with changemakers to create a more responsive, patient-centered health care system.

Endnotes

1. **California Health & Safety Code § 109.7.130290(h)**.
2. Mark Elson et al., ***Health Information Exchange in California: Assessment of Regional Market Activity***, California Health Care Foundation (CHCF), August 2021.
3. Eric Heflin et al., ***A Framework for Cross-Organizational Patient Identity Management***, Sequoia Project, May 30, 2018; and Genevieve Morris et al., ***Patient Identification and Matching Final Report*** (PDF), Office of the National Coordinator for Health Information Technology (ONC), February 7, 2014, 41.
4. Heflin et al., Framework; and Morris et al., *Patient Identification*, 41.
5. ***United States Core Data for Interoperability, Version 1 (July 2020 Errata)*** (PDF), ONC, July 2020.
6. Ako Quammie Interview conducted virtually on March 4, 2022; Sandeep Kapoor, JoAnne Hawkins, and Dawn R. Gallagher, ***Designing a Statewide Health Data Network: What California Can Learn from Other States***, CHCF, March 2021; and Morris et al., *Patient Identification*, 51, 54.
7. Tim Pletcher Interview conducted virtually on March 2, 2022; Morris et al., *Patient Identification*, 51, 54; and Kapoor, Hawkins, and Gallagher, Designing.
8. Michael Berger interview conducted April 7, 2022; Morris et al., Patient Identification, 51, 54; Kapoor, Hawkins, and Gallagher, Designing; and **N.Y. Comp. Codes R. & Regs. tit. 10 § 300**, accessed March 15, 2022.