

## THE ISSUE

Clinical and community organizations capture specific information on the individuals and families they serve. Linking individual information across organizations builds capacity for program evaluation and provides a more complete picture of health at an individual and population level, including the social determinants of health. However, linking data across organizations can create concerns due to the risks around sharing an individual's personally identifiable information (PII), unclear guidelines regarding ownership and use of linked data, and unique information security requirements. Linkage can be especially difficult in new partnerships, when a track record of trusted data exchange has not been established.

## THE AIM

To build trusted data sharing partnerships between partner organizations, the U.S. Centers for Disease Control and Prevention (CDC) is leading the Clinical and Community Data Initiative (CODI) to leverage information technology tools to facilitate the linkage of individual data across clinical and community sectors. CODI harmonizes data to create a **linked longitudinal record** and affiliate individuals living together using a household address. CODI protects individual privacy and secures PII using privacy-preserving record linkage (PPRL), a linkage method where PII is not shared.

## DATA SHARING

CODI creates a technology and partnership network that allows partners to maintain control over their own data. Instead of submitting data to a repository, queries are used to extract on the minimum required data, without PII. Before data is shared, CODI implementing partners will establish governance to build trust and reconcile regulations, policies, and norms needed for sustained network data exchange. For CODI, governance includes policies, processes, and agreements to assure Data Partners that **data sharing is ethical, transparent, and compliant with appropriate laws and regulations.**

### CODI Roles

- **Data Owner** – contributes data for queries and may require a Technical Partner and/or Data Partner for data exchange
- **Data Partner** – facilitates CODI data exchange by hosting data (its own or other Data Owners'), responds to data requests from other CODI partners, and implements CODI data models and PPRL tools
- **Technical Partner** – supports Data Owners that want to contribute data but lack technical capabilities to support CODI infrastructure (e.g., installs software, troubleshoots technology)
- **Data Coordinating Center** – assembles longitudinal, de-identified records from data provided by Data Partners
- **Linkage Agent** – performs individual and household linkage and generates unique identifiers for Data Partners
- **Key Escrow** – generates a unique key used in the deidentification process

## PRIVACY-PRESERVING RECORD LINKAGE

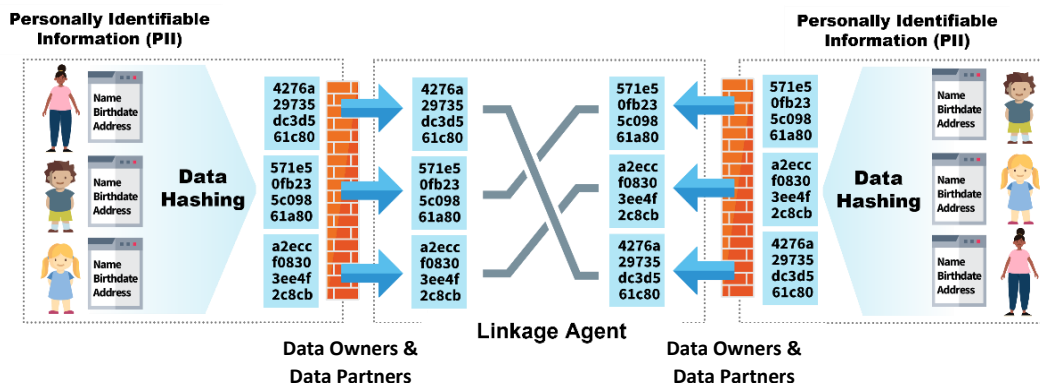
PPRL is the method where links are **established between records without disclosing PII** outside the originating organization. **PII remains secure within the originating organization's system.** PPRL relies on encoding of the individual's data, called "hashing." **Hashing prevents access to the individual's PII**, but allows records belonging to the same individual to be matched across organizations and within households. An individual's PII is passed through a hashing function that:

- Garbles the data.
- Is "one-way." It is very difficult, if not impossible, to extract the PII from the garbled data.
- Is deterministic. The same input will always generate the same output.

## PPRL DEFINITIONS

- **Linkage:** Identifying multiple records that represent a single person.
- **Linkage Agent:** Performs data linkages on behalf of Data Partners.
- **Household Linkage:** Multiple individual records linked to the same shared dwelling.
- **PII:** Personally identifiable information such as name, address, or social security number.
- **Hashing:** A function that scrambles data.
- **Plain Text:** Unencrypted data that is intended to be encrypted and decrypted.

The hashing function transforms the plain text PII to a data file in which the PII is garbled. This file can then be shared directly with the linkage agent to perform matching. As a simplified example, two organizations that want to link records will use a hashing function to encode PII in a data file and then share the encoded file with a linkage agent. **The linkage agent does not have access to the PII**, but they know that records exist for the same individuals—or individuals within the same household—at both organizations because the hashing function values are the same from both organizations.



## FREQUENTLY ASKED QUESTIONS

**Q. Do personal identifiers (e.g., participants' name, address) ever leave my firewall in the CODI project?**

**A.** No. The PPRL process is designed so organizations PII never has to leave their firewall.

**Q. Does participant-level data leave my firewall?**

**A.** Yes. However, each encoded data set is bound by terms of agreement, defined in collaboration with all Data Owners, to which a data requester must agree prior to receiving any individual level data. This data use agreement (DUA) includes explicit expectations around approved uses of the longitudinal records. Data Owners may specify terms such as restricting access to data to anyone outside of the DUA, maintaining ownership of the data, authority to terminate the DUA at any time, or destruction of the data upon termination of the DUA.

**Q. How does the Health Insurance Portability and Accountability Act (HIPAA) apply to the CODI project?**

**A.** HIPAA applies to patient health data which does not include all community partner data. A partner's data may be bound by laws, organizational policies, or agreements other than HIPAA. CODI's governance process will uncover these and execute an agreement that will meet all the regulations and policies of the Data Owners.

**Q. We do not have technical data experts at my organization. Is there support for this?**

**A.** The CODI project will work with Data Owners to identify a Data Partner and Technical Partner that can provide technical support. For this relationship, a DUA or other appropriate agreement may be developed to ensure that the Data Owner would make all decisions related to the use of the data. Data Owners will decide what technical tasks they would like the Data Partner to conduct on their behalf. For example, a Data Owner may prefer to hash their own data but have a Data Partner respond to queries for them.

For more information, contact [CODI@cdc.gov](mailto:CODI@cdc.gov).

Note: The following legend should be used for all software and technical data developed using contract funds (including MOIE funding). This legend should be affixed to the software/technical data prior to delivery to the customer.

NOTICE

This (software/technical data) was produced for the U. S. Government under Contract Number 75FCMC18D0047, and is subject to Federal Acquisition Regulation Clause 52.227-14, Rights in Data-General.

No other use other than that granted to the U. S. Government, or to those acting on behalf of the U. S. Government under that Clause is authorized without the express written permission of The MITRE Corporation.

For further information, please contact The MITRE Corporation, Contracts Management Office, 7515 Colshire Drive, McLean, VA 22102-7539, (703) 983-6000.

©2022 The MITRE Corporation.