

Use Case: Linking and Enhancing Multiple External Data Sets Using CRISP IDs and Geocodes for Research

Submitted by Research Subcommittee for approval by CRISP Clinical Advisory Board.

Last updated 09/29/2017

Background

The CRISP Enterprise ID (EID) is a core element for making health information exchange possible as it associates data from multiple data sources into a collated set of data through probabilistic matching of patient demographic information with the same information from other sources. CRISP uses an instance of a commercial tool, IBM's Initiate, aka, our Master Patient Index (MPI), to assign, maintain and link EIDs to patient records for the more than 17 million unique patient identities in the CRISP network.

CRISP's MPI service is primarily used for driving our core clinical services – namely the Encounter Notification Service and the Clinical Query Portal (Mirth Results) – as well as CRISP Reporting Services where patient data are matched across health systems and providers. CRISP also uses the MPI to provide the HSCRC Case Mix data to researchers through a [use case approved in March of 2017](#). In that use case, if the HSCRC approves the use of their data and the enhancement of the data through three basic services: linking the data to the MPI; scrubbing the data of demographic information and replacing it with a “hashed” MPI; and enhancing the data by adding geocodes of the generalized location of the patient.

Recent data requests for adding CRISP identifiers and geocodes to the HSCRC Case Mix data have also requested the same services for other data sources, thus linking patient information from two different data sources while de-identifying the individual patients. This use case serves as a generalized version of the earlier HSCRC Case Mix use case for multiple data sets.

Use Case Description

As with the Case Mix data use case, this use case would support the application of three CRISP services to external data sets that include sufficient demographic information as to enable patient matching.

- Data Linker: CRISP runs the demographic information for each received data set into CRISP's MPI, and assigns it an EID.
- Data Scrubber: CRISP creates a hashed (a masked version of the) EID for each patient that is unique to the research study but shared across data sets so that the data can be linked to other data sets that are part of the study without exposing the actual CRISP ID or making the identifier useful for linking the data to any data outside of the research study. CRISP then removes the demographic information from the data set, essentially replacing it with the hashed EID. Some data set holders will send only the demographic information and request the hashed EID, which they will append to their data and perform their own de-identification process on the data before sending it to the researcher. Other data de-identification services could include randomization of service dates.
- Data Enhancer (if requested): CRISP adds geocodes to the data at a [Census Block](#) (most granular), [Census Block Group](#), or [Census Tract](#) (least granular) level based on the patient's last known address. Researchers can look at the geographic distribution of the data without knowing the precise location of the person's address. The level of data protection through grouping will be evaluated by the Research Subcommittee following the principles of providing the minimum data necessary.

When these services are applied to multiple data sets, researchers can use the hashed EID to associate patient-level information across data sets without having any identifying demographic information on the patient.

Note that this use case currently does not include the use of CRISP-managed data – that is, data coming from CRISP participating organizations that might be found in the Clinical Query Portal or Encounter Notification Service– other than the CRISP EID and geocodes. Data sets must be provided to CRISP for linking.

In studies approved under this use case, CRISP serves as a data manager and aggregator of these data sets by leveraging our core capabilities without doing additional analysis on the data.

Researcher responsibilities

To participate in this service, researchers are required to do the following:

- Submit a data use request to CRISP with the appropriate accompanying documentation for an IRB-approved study using the data sets intended for submission. The request must include the name of the CRISP participant entity requesting the data; name of the IRB board; name and role of principal investigator and any co-investigators; details about the purpose of the study and study protocol; the start and end date of the study; justification for the geocode granularity requested, if applicable; and a description of how the data request meets policy requirements and authorizes CRISP to allow the researcher access to the requested data sources to obtain the individual’s information.
- Secure approval from the data owners or stewards for the use of the data in the study. Data sets such as the HSCRC Case Mix data must be approved through their own governance process if such a process has been established. Evidence of approval must be submitted to CRISP in advance of the data set being delivered to CRISP.
- Establish a data governance agreement among the data contributors wherein each data provider acknowledges the purpose of the study and the conditions under which the data sets may be combined.
- Demonstrate adequate safeguards to ensure confidentiality of source data, prevent re-disclosure that would identify an individual, and prevent re-release of source data through signing a Data Use Agreement in a CRISP-approved form.
- Provide acknowledgement of the use of CRISP data in publications resulting from its use.
- Reimburse CRISP’s chargeable costs and expenses.

CRISP responsibilities

CRISP is responsible for the following:

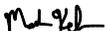
- CRISP manages the EID and geocode services as part of its CRISP Reporting Services (CRS) work and will provide an estimate to the researcher of the level of effort required for performing the outlined CRISP services on the data sets provided.
- Review and approve data access requests in accordance with CRISP’s governance process.
- Establish secure mechanisms for transfer of data sets among data providers, CRISP and researchers.
- Provide boilerplate language that accurately describes the data provided for the study for use in publication.

Permitted Purpose Category

Permitted purpose #4, Research.

Approval:

This Use Case Policy has been approved by the Clinical Advisory Board.


Mark Keleman (Dec 7, 2017)

Dr. Mark Keleman, Chairperson

21 November 2017

Date Approved



Linking and Enhancing Multiple External Data Sets Using CRISP IDs and Geocodes - APPROVED - 2017-11-21

Adobe Sign Document History

12/07/2017

Created:	11/22/2017
By:	Ross Martin (ross.martin@crisphealth.org)
Status:	Signed
Transaction ID:	CBJCHBCAABAaHYPjDdTIPfbW3_hqecZ3iTHDKuKCdPii

"Linking and Enhancing Multiple External Data Sets Using CRISP IDs and Geocodes - APPROVED - 2017-11-21" History

- Document uploaded by Ross Martin (ross.martin@crisphealth.org) from Acrobat
11/22/2017 - 1:17:19 PM PST - IP address: 96.244.108.50
- Document emailed to Mark Kelemen (mkelemen@som.umaryland.edu) for signature
11/22/2017 - 1:17:55 PM PST
- Document viewed by Mark Kelemen (mkelemen@som.umaryland.edu)
11/22/2017 - 1:30:15 PM PST - IP address: 174.229.139.103
- Document viewed by Mark Kelemen (mkelemen@som.umaryland.edu)
11/29/2017 - 3:10:16 PM PST - IP address: 174.229.139.251
- Document e-signed by Mark Kelemen (mkelemen@som.umaryland.edu)
Signature Date: 12/07/2017 - 10:18:08 AM PST - Time Source: server- IP address: 174.205.17.73