Date Revised
July 2020

Date Effective
September 2014

Extraction of De-Identified Data from the AHC Information Exchange

Responsible University Officer: Clinical and Translational Science Institute Biomedical Informatics Director

Policy Owner: Constantin Aliferis, califeri@umn.edu

POLICY STATEMENT

To create efficiencies and better serve our constituents, this policy permits the extraction of de-identifed data from the AHC Information Exchange (IE) for those use cases that meet the criteria described herein, without approval from the Executive Leadership & Governance Committee (the "Committee") for each individual request. Use cases that do not meet the criteria described in this policy must be approved by the Committee before extraction can be performed.

Use cases that would require the extraction of de-identified data from the IE are as follows:

1) Extraction required to use computational power or applications not available within the IE secure data environment, or
2) Extraction required to compile with other data (such as with data at another research institution, data already existing in another physical location that cannot reasonably be moved, reporting to a data registry)

General Requirements for Extraction

1. Extraction must be for purposes of:
   - IRB approved research; or
   - Healthcare operations (other than fundraising) as permitted under HIPAA;

AND

2. Extraction of data is necessary due to a need to:
   - use computational power or applications not available in the IE data shelter; or
   - compile with other data not in the IE (such as with data at another research institution, data already existing in another physical location that cannot reasonably be moved, reporting to a data registry);

AND

3. Requester must be an Authorized User per the AHC-IE Data Access and Use Policy (at a minimum, requester must have an x500 User ID and must have completed HIPAA Training)

Evaluation and Oversight

Requests for data extractions will be tracked as part of the overall data request process. Information tracked includes all data fields entered for the request, as well as total number of requests, categories of request (computation, data sharing), and number of requests fulfilled.

Reporting on data extraction will be presented to the Committee by staff from the Best Practices Integrated Informatics Core (BPIC) of the University's Clinical and Translational Science Institution on a quarterly basis, in the form and format requested by the Committee, and will also be available at other times upon request by the data stewards of participating organizations (currently Fairview, UMP and UMN).

This *Extraction of De-Identified Data from the Information Exchange* policy will be presented to the Committee by BPIC on a regular basis for review.

REASON FOR POLICY

To create a governance process to manage requests to extract de-identified data from the AHC Information Exchange.

PROCEDURES

Requests to extract data from the IE secure data environment will be submitted to the Data Shelter File Transfer application.

1. Login to the AHC-IE Data Shelter
2. From any browser, go to the URL: https://portunus.ahc.umn.edu
3. Sign in using your x500 username and password
4. Approve dual authentication
5. Select your project
6. Specify the reason for extracting the file
7. Upload file to be extracted*

*Note: Files leaving the data shelter will need to be approved before they are available to download

FORMS/INSTRUCTIONS

Instructions can be accessed here - https://confluence.ahc.umn.edu/display/POR/File+Transfers

APPENDICES

Appendix A: IE De-identification Process

FREQUENTLY ASKED QUESTIONS

There is no FAQ associated with this policy.

ADDITIONAL CONTACTS

Subject	Contact	Phone	Email
Primary Contact	**Gretchen Sieger**	**612.626.7495**	**siege022@umn.edu**
Alternate Contact	Steve Johnson	612.625.7940	joh06288@umn.edu

RELATED INFORMATION

University of Minnesota Administrative and Board of Regents Policies:

- Protected Health Information
- Protection of Individual Health Information
- Research Data Management: Archiving, Ownership, Retention, Security, Storage, and Transfer

University of Minnesota Administrative Procedures:

- Destruction of University Records

Clinical and Translational Science Institute Policies and Procedures:

- AHC-IE Data Access and Use Policy
- BPIC Data Request Procedures

HISTORY

Policy Created: 9/9/2014

Policy Revised: 6/16/2020

APPENDIX A

IE De-Identification Process

Data extracted from the IE will be de-identified in compliance with HIPAA and IE procedures, as described in below.

According to HIPAA and HHS Guidance, de-identification can be achieved by one of two methods: 1) Expert Determination or 2) the removal or suppression of 18 specific identifiers, referred to as the Safe Harbor method (see: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#safeharborguidance).

The IE uses the Safe Harbor method and will suppress all 18 identifiers (referred to as HIPAA identifiers):

1. Names (see **Names**)
2. All geographical subdivisions smaller than a state, including street address, city, county, precinct, zip code (see **Zip Codes**), and their equivalent geographical codes
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date and date of death (see **Dates**)
4. Telephone numbers
5. Email addresses
6. Fax numbers
7. SSNs
8. MRNs
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/License numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. URLs
15. IP addresses
16. Biometric identifiers, including finger and voice prints
17. Full-face photographs and any comparable images
18. Any other unique identifying number, characteristic, or code, unless otherwise permitted by the Privacy Rule for re-identification

Names

Only names of the individuals associated with the corresponding health information (i.e., the subjects of the records) and of their relatives, employers, and household members will be suppressed. There is no explicit requirement to remove the names of providers or workforce members of the covered entity or business associate.

Zip Codes

We will supply the first three digits of a patient's zip code. However, to produce a de- identified data set utilizing the safe harbor method, all three-digit zip codes with a population of 20,000 or fewer persons must have the zip code changed to 000.

According to 2010 Census data, the following three---digit zip codes must be set to 000: 036, 059, 102, 202, 203, 204, 205, 369, 556, 692, 753, 772, 821, 823, 878, 879, 884, and 893.

https://www.census.gov/geo/maps-data/data/gazetteer2010.html

Dates

If birth date, death date, or any service related dates (e.g. admit date, discharge date, appointment date, order date, result date, diagnosis date) are requested, each date will be offset by +/- 30 days. This offset, or "seed", is randomly generated for each patient within the request. All dates are shifted by the same seed so relative time is maintained for the patient throughout the data set. The date shift will only be applied to the date, not the timestamp. Timestamps maintain their original values. The date shift seed is stored within a database table only accessible by IE developers.

Table 1 --- Example

REQUEST_ID	PAT_ID	DATE_SHIFT_SEED
REQ23	REQ23---8989	-4
REQ23	REQ23---3112	11
REQ24	REQ24---1092	22
REQ24	REQ24---9039	-30

Birth date: AHC- IE developers will flag all patients who are of the age 89 (and over) at the point in time the original query is executed (+ 30 days). All of those patients will then be applied the same birthdate which will be -89 years from the date the query is executed.

Pseudo identifiers

Pseudo identifiers are generated for each unique patient and service record. This is done so the de-identified data can be linked back to the identifiable patient information if needed and allowed. The pseudo IDs are randomly generated for each request/patient/visit. The relationship to the real patient ID and visit ID are stored in a database table only accessible by IE developers.

Available Clinical Data

Clinical data values such as vital signs, labs, medications, procedures, and clinic/hospital location are available within a de-identified dataset. However, the IE does not have the software in place that would be needed to de-identify free text such as clinical notes and patient social history notes. Therefore, until such a system is in place, we will not provide free text clinical data as part of de-identified datasets.

Delivery

The most common format for the data output is Excel spreadsheets. The spreadsheet will be delivered to the end user(s) via the IE Data Shelter. A copy of the original data set will be maintained within an IE file share that is only accessible by IE developers and system administrators.

Final output must contain more than 10 patient records to be delivered.