POLICY

Extraction of De-Identifiable Data from the Information Exchange

Approved Proposal

Purpose

Create a governance process to manage requests to extract de-identified data from the Information Exchange (IE).

Background

Several use cases have been identified that require the extraction of de-identified data from the IE. Those cases are categorized as follows:

1) extraction required to use computational power or applications not available within the IE, or

2) extraction required to share with other data (either on a local or national basis).

To create efficiencies and better serve our PIs, we propose a process to govern these requests that would not require approval by the Executive Leadership Committee for each individual request. The process would be limited to allow extraction of de-identified data only for the use cases described above.

Process Requirements

1. Extraction Requirements
   • Requester must be an IE Authorized User (must have x500 id, have completed HIPAA training, and have IRB approval or waiver of approval).
   • The extraction must be required to 1) use computational power or applications not available in the IE, or 2) share with other data. If the extraction is for any other purpose,
the request must be forwarded to the Executive Leadership Committee for approval on an individual basis.

- Data will be de-identified according to the AHC IS process. This process meets requirements under HIPAA for de-identification of data. See attached for details of this process.

2. 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 extractions will be available to the data stewards of Fairview, UMP and UMN on demand through AHC-IS.

3. Data Request Form Modifications
   - The current request form to request data will be modified to support this process.
   - Expected changes to the request form will likely include adding fields to support additional information on:

     1) the need for the extraction of de-identified data (pull-downs will likely be added for needs related to computational power, other applications and sharing with other data, along with a free form response field to provide additional information or to describe another need)

     2) how the data will be handled after extraction, including who will have access to the data, where the data will be stored, how it will be secured, and when it will be disposed of
AHC IE De-Identification Process

This document outlines the approach taken by AHC IE developers for data de-identification of clinical data sets. According to HHS Guidance, de-identification can be obtained through either Expert Determination or through Safe Harbor by suppressing the 18 HIPAA identifiers.


AHC IE has chosen to use the Safe Harbor definition and will suppress all 18 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.  
http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/De-identification/guidance.html#supress

**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.

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

**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 AHC IE developers (table 1).

<table>
<thead>
<tr>
<th>REQUEST_ID</th>
<th>ID_TYPE</th>
<th>SRC_ID</th>
<th>NEW_ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>REQ23</td>
<td>MRN</td>
<td>00000123</td>
<td>REQ23---8989</td>
</tr>
<tr>
<td>REQ23</td>
<td>VISIT</td>
<td>333329200</td>
<td>REQ23---930094</td>
</tr>
<tr>
<td>REQ24</td>
<td>MRN</td>
<td>00000123</td>
<td>REQ24---1092</td>
</tr>
<tr>
<td>REQ24</td>
<td>VISIT</td>
<td>333329200</td>
<td>REQ24---739221</td>
</tr>
</tbody>
</table>

**Dates**

6/10/2014  Approved by Academic Health Center Information Exchange Executive Leadership & Governance Committee
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 AHC IE developers.

Table 2 --- Example

<table>
<thead>
<tr>
<th>REQUEST_ID</th>
<th>PAT_ID</th>
<th>DATE_SHIFT_SEED</th>
</tr>
</thead>
<tbody>
<tr>
<td>REQ23</td>
<td>REQ23---989</td>
<td>-4</td>
</tr>
<tr>
<td>REQ23</td>
<td>REQ23---3112</td>
<td>11</td>
</tr>
<tr>
<td>REQ24</td>
<td>REQ24---1092</td>
<td>22</td>
</tr>
<tr>
<td>REQ24</td>
<td>REQ24---9039</td>
<td>-30</td>
</tr>
</tbody>
</table>

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.

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 AHC 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 AHC IE Data Shelter. A copy of the original data set will be maintained within an AHC IE file share that is only accessible by AHC IE developers and system administrators.

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