NIST Immunization Data at Rest (DAR)

Data Extraction Specification

Jan 2, 2018

**Introduction**

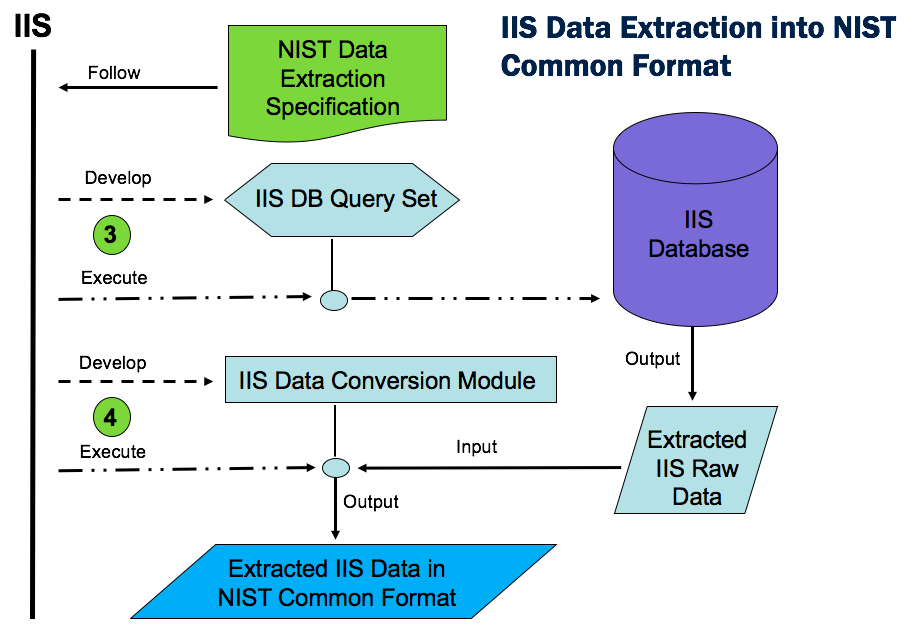
This specification describes the data elements to extract from an IIS, and how to store them in an external file for the purposes of data quality analysis by the NIST DARQ (Data At Rest Quality) analysis tool. The data elements extracted may come directly or be derived from IIS data store(s). The scope/quantity of the data extracted by the process (e.g. query) is beyond the scope of this document and shall be provided in a separate external guidance document.

**Overview**

Key steps (refer to Figure 1):

1. Develop/implement process to extract specified data elements
2. Execute extraction process (ex. DBMS queries)
3. Store the extracted Patient data elements and Vaccine data elements into a separate ASCII text files (Tab separated values) with a unique Patient ID used to associate data between them.

TBD: Text providing guidance on repeating elements (“only extract the latest, best known, etc.”).

****

**Figure 1 - Example Usage**

**Output Format for Extracted Data Elements**

All data shall be output to the relevant Extract File as UTF-8 encoded text strings (see: <http://wiki.hl7.org/index.php?title=Character_Set_used_in_v2_messages)> without change from the local representation except for

* Elements which represent Dates (e.g. Date of Birth) shall be stored using the date string format: *yyyy-MM-dd*
* Elements represented using a coding system (e.g. Product Type may use CVX codes, local codes, etc.) shall be stored according to the coding system prescribed in HL7 V2.5.1.

**Patient and Vaccine Extract File Format**

Extracted patient and vaccine-related data elements (see tables 1 & 2) shall be stored in two separate files, a Patient Extract File and a Vaccine Extract File. A Patient ID data element is used to link every patient record to its corresponding vaccine record(s). The structure of both files, referred to below simply as “Extract File,” are identical.

Each Extract File is a text file containing lines of tab[[1]](#footnote-1) separated values (TSV), where the first line consists of Data Element Name values from the tables below (Patient or Vaccine) - i.e. the column headings. Each subsequent line represents a logical “record,” consisting of tab separated values containing the extracted values corresponding to each “column.”

Note that it is possible for some extracted data elements themselves to contain one or more tab (ASCII 09) characters. In such instances, any tab characters *appearing* *in data elements* must be mapped to space (ASCII 32) characters in the Extract File.

Prior to data extraction, the IIS will be instructed[[2]](#footnote-2) as to which patient and vaccine-related data elements are “in scope” and should be extracted versus those elements which should not be extracted (i.e. excluded). Additionally, prior to data extraction, the IIS shall determine which of the above “in scope” elements may or may not be extracted for any of the following reasons:

* the data element is never collected/stored by the IIS
* local policies prohibit release of the element, and any metadata associated with it (ex. presence, length, etc.)
* local policies only permit release of meta data only

The determination of which data elements shall or shall not be extracted must be consistent across all lines (i.e. records) of the Extract File. Thus, each data element shall either *always* be extracted (and therefore output to the Extract File), or *never* extracted, in which case a predefined string indicator (see Figure 2 below) shall be output to the Extract File in the respective column for that element.

Figure 2 provides a flow chart of how to determine what should be output to the Extract File for any given data element. Step 1 is for the IIS to consult the external guidance document to see which elements are “in scope” for the extraction. Elements not in scope shall be represented by the special string “[[EXCLUDED]]” (quotes not included) in the Extract File.

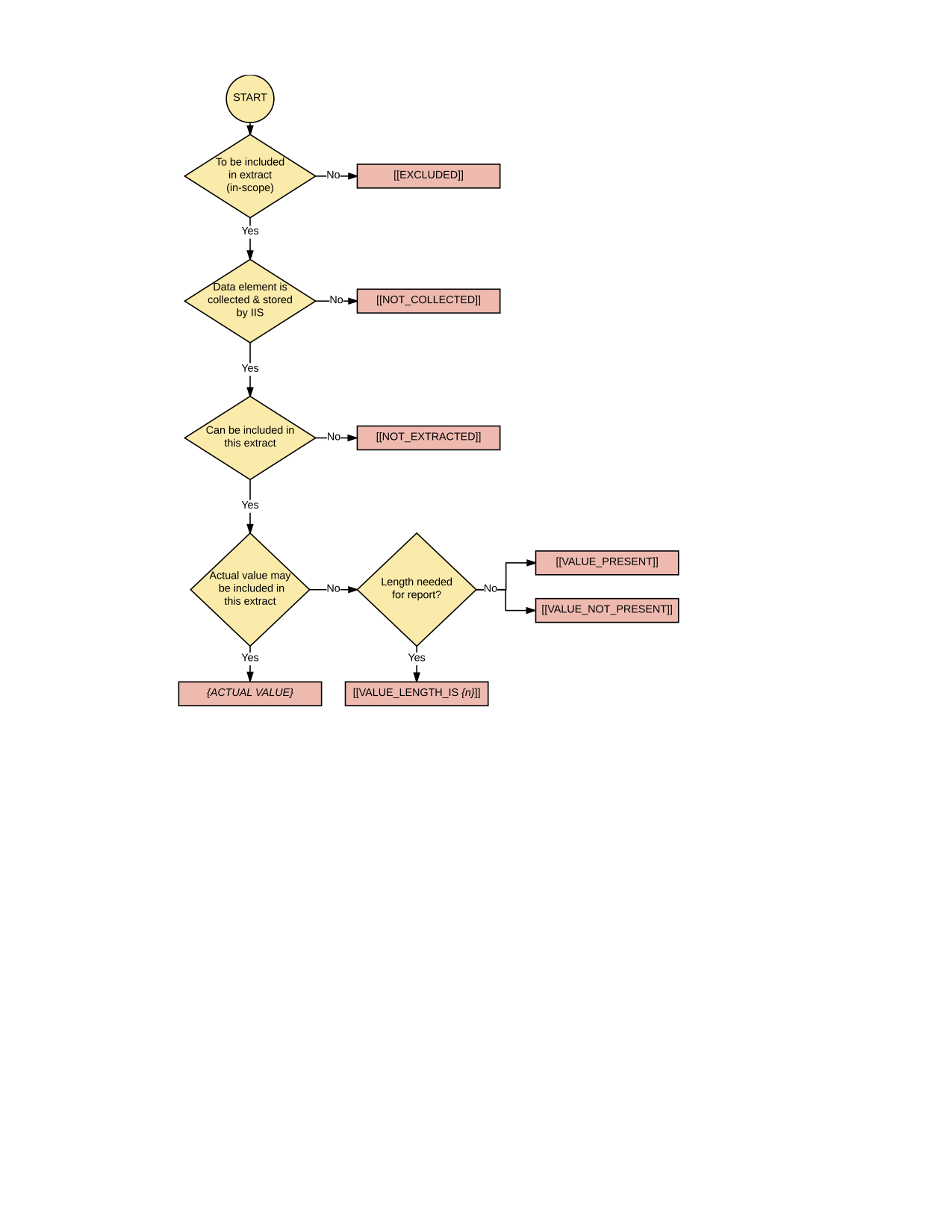
Step 2 is for the IIS to determine which of the “in scope” data elements are actually extractable (i.e. are they in the IIS database?). Elements which cannot be extracted because they are not collected and/or not stored in the IIS database shall be represented by the special string “[[NOT\_COLLECTED]]” (quotes not included) in the Extract File.

The third step is for the IIS to determine, based on their local policy, which of the remaining data elements (after steps 1 & 2 above) they are *willing* to extract. For example, if the data quality analysis will take place offsite (e.g. because IIS policy prevents local installation of the NIST DARQ analysis software) the IIS may choose *not* to extract the element out of privacy concerns. In that case the element shall be represented by the special string “[[NOT\_EXTRACTED]]” (quotes not included) in the Extract File.

Some IIS, however, may determine that although they are not willing to extract the actual values of particular data elements, they *are* willing to indicate whether or not the element is or is not present in the IIS database (such information is essential to the quality analysis for measurement of data “completeness”). In such cases the special strings “[[VALUE\_PRESENT]]” and “[[VALUE\_NOT\_PRESENT]]” (quotes not included) may represent the element in the Extract File.

Similarly, some IIS may determine that although they are not willing to extract the actual values of a particular data element, they *are* willing to extract its length (e.g. it may be useful to the quality analysis to know the length of some data elements such as Patient Name). In such cases the special string “[[VALUE\_LENGTH\_IS {n}]]” (quotes not included) shall be used to represent the data element in the Extract File, where {n} is an integer representing the string length of element’s value in the IIS database. Note that data elements which contain empty/null (ex. note present) values in the database should result in {n} being set to 0.

Note that the above cases where data elements *are* *not* extracted is different from the cases where data elements *are* being extracted but contain empty/null (ex. not present) values in the IIS database. The latter shall result in <tab><tab> being output to the Extract File.



**Figure 2: Determining special character-string values for data elements not extracted**

**Patient Data Elements**

The patient-related data elements to extract to the Patient Extract File are listed in the table below.

Note that all of the patient records in the Patient Extract File shall have at least one record in the Vaccine Extract File (see next section) unless the patient has no immunizations. Thus, if a patient record does not have a corresponding immunization record the analysis will assume this patient has received no immunizations.

Also note that patient information is assumed to be “consolidated” by the IIS prior to extraction, and thus a single set of the elements listed below are expected to be extracted for each Patient.

Table 1: Patient Data Elements

|  |  |  |
| --- | --- | --- |
| Data Element Name | Type | HL7 Example |
| Patient ID | | String | PID-3 |
| Name - First | | String | PID-5.2 |
| Name – Middle | | String | PID-5.3 |
| Name - Last | | String | PID-5.1 |
| Mother’s Maiden Name | | String | PID-6.1 |
| Mother’s Name - First | | String | NK1-2.2 |
| Mother’s Name - Middle | | String | NK1-2.3 |
| Mother’s Name - Last | | String | NK1-2.1 |
| Date of Birth | | yyyy-MM-dd | PID-7 |
| Sex/Gender | | String | PID-8 |
| Address - Street | | String | PID-11.1 |
| Address - City | | String | PID-11.3 |
| Address - State | | String | PID-11.4 |
| Address - Country | | String | PID-116 |
| Address – Zip/postal | | String | PID-11.5 |
| Number of Race Codes Stored | | String | PID-10 |
| Number of Ethnicity Codes Stored | | String | PID-22 |
| Phone | | String | PID-13.6  PID-13.7 |
| Email address | | String | PID-13.4 |
| Primary Language | | String | PID-15 |
| Patient Alias Name: First | | String | PID-9.2 |
| Patient Alias Name: Middle | | String | PID-9.3 |
| Patient Alias Name: Last | | String | PID-9.1 |
| Responsible Person Name: First | | String | NK1-2.2 |
| Responsible Person Name: Middle, | | String | NK1-2.3 |
| Responsible Person Name: Last | | String | NK1-2.1 |
| Responsible Person : Relationship to Patient | | String (code) | NK1-3.1 |
| Birthing Facility Name | | String | PID-23 |
| Patient Multiple Birth Indicator | | String | PID-24 |
| Patient Birth Order | | String | PID-25 |
| Patient status indicator—Provider facility level | |  |  |
| Patient status indicator—IIS level | |  |  |

**Vaccine Data Elements**

The vaccine-related data elements to extract to the Vaccine Extract File are listed in the table below.

Note that all immunization records in the Vaccine Extract File shall have a corresponding patient record in the Patient Extract File (see section above). If an immunization record does not have a corresponding patient record in the immunizations will be noted and discarded from analysis.

Table 2: Vaccine Data Elements

|  |  |  |
| --- | --- | --- |
| Data Element Name | Type | HL7 Example |
| Patient ID | | String | PID-3 |
| Vaccine Type (CVX) | | String | RXA-5 |
| Vaccine Type (NDC) | | String | RXA-5 |
| Administration Date | | yyyy-MM-dd | RXA-3 |
| Manufacturer | | String | RXA-17 |
| Lot Number | | String | RXA-15 |
| Vaccination Event Record Type (administered/historical)  AKA: Vaccine Event Information Source  AKA: Administration Notes (RXA-9) | | String | RXA-9.1 |
| Administering Provider | | String | RXA-10 |
| Administered at Location | | String | RXA-11 |
| Administration Route | | String | RXR-1 |
| Administration Site | | String | RXR-2 |
| Expiration Date | | yyyy-MM-dd | RXA-16 |
| Dose Volume/Unit | | String | RXA-6  RXA-7 |
| Ordering Provider Name - First | | String | ORC-12.3 |
| Ordering Provider Name - Middle | | String | ORC-12.4 |
| Ordering Provider Name - Last | | String | ORC-12.2 |
| VIS Information - Type | | String |  |
| VIS Information - Publication Date | | yyyy-MM-dd |  |
| VIS Information - Date Given to Patient | | yyyy-MM-dd |  |
| Vaccine Eligibility at Dose Level | | String |  |
| Complete Status | | String | RXA-20 |

1. “tab” refers to the ASCII HT (Horizontal Tab) character, which has a decimal value of 09, and is the same character typically generated by pressing the ‘Tab’ key on most keyboards. [↑](#footnote-ref-1)
2. Guidance will be provided {reference?} on “in scope” elements for the extraction. [↑](#footnote-ref-2)