Integrating the Healthcare Enterprise



IHE IT Infrastructure

Technical Framework Supplement

De-Identification for Family Planning

(De-ID for FP)

Draft for Public Comment

Date: December 14, 2015

Author: IHE ITI Technical Committee

Email: iti@ihe.net

**Please verify you have the most recent version of this document.** See [here](http://ihe.net/Technical_Frameworks/) for Trial Implementation and Final Text versions and [here](http://ihe.net/Public_Comment/) for Public Comment versions.

**Foreword**

This is a supplement to the IHE IT Infrastructure Technical Framework V12.0. Each supplement undergoes a process of public comment and trial implementation before being incorporated into the volumes of the Technical Frameworks.

This supplement is published on December 14, 2015 for public comment. Comments are invited and can be submitted at [http://www.ihe.net/ITI\_Public\_Comments](http://www.ihe.net/ITI_Public_Comments/). In order to be considered in development of the trial implementation version of the supplement, comments must be received by February 5, 2016.

This supplement describes changes to the existing technical framework documents.

“Boxed” instructions like the sample below indicate to the Volume Editor how to integrate the relevant section(s) into the relevant Technical Framework volume.

Amend Section X.X by the following:

Where the amendment adds text, make the added text bold underline. Where the amendment removes text, make the removed text bold strikethrough. When entire new sections are added, introduce with editor’s instructions to “add new text” or similar, which for readability are not bolded or underlined.

General information about IHE can be found at: [http://ihe.net](http://ihe.net/).

Information about the IHE IT Infrastructure domain can be found at: [http://ihe.net/IHE\_Domains](http://ihe.net/IHE_Domains/).

Information about the organization of IHE Technical Frameworks and Supplements and the process used to create them can be found at: [http://ihe.net/IHE\_Process](http://ihe.net/IHE_Process/) and [http://ihe.net/Profiles](http://ihe.net/Profiles/).

The current version of the IHE IT Infrastructure Technical Framework can be found at: <http://ihe.net/Technical_Frameworks/>.

CONTENTS

[Introduction to this Supplement 4](#_Toc437851710)

[Open Issues and Questions 4](#_Toc437851711)

[Closed Issues 4](#_Toc437851712)

[**Volume 4 – National Extensions 6**](#_Toc437851713)

[4 National Extensions 6](#_Toc437851714)

[4.R2 De-Identification for Family Planning data 7](#_Toc437851715)

[4.R2.1 Algorithms for the de-Identification of Family Planning data 7](#_Toc437851716)

[4.R2.1.1 Facility Identifier Mapping Table 9](#_Toc437851717)

[4.R2.1.2 Clinical Provider ID Mapping Table 9](#_Toc437851718)

[4.R2.1.3 Patient Identifier ID Mapping Table 10](#_Toc437851719)

[4.R2.1.4 Visit Date 10](#_Toc437851720)

[4.R2.1.5 Administrative Sex 10](#_Toc437851721)

[4.R2.1.6 Limited English Proficiency (Language) 10](#_Toc437851722)

[4.R2.1.7 Race 11](#_Toc437851723)

[4.R2.1.8 Contraceptive Method 11](#_Toc437851724)

[4.R2.2 Example of De-Identified Family Planning Data 11](#_Toc437851725)

# Introduction to this Supplement

This supplement provides the volume 4 US realm specific de-Identification algorithms for the IHE QRPH Family Planning CDA® data elements.

For an understanding of how these algorithms were selected, please see the supporting whitepaper entitled “IHE ITI Whitepaper Analysis of optimal De-Identification algorithms for Family Planning data elements”.

## Open Issues and Questions

* Is there a problem with the length of Universally Unique Identifiers (UUIDs)? Probably not, but the Comma Separated Value (CSV) rows will end up being fairly long.

## Closed Issues

* What format should be used to publish the de-Identified data? The input data will be received in Clinical Document Architecture (CDA®) format. Is CDA® format preferred for the de-Identified output?
* For data elements where the de-identification algorithm transforms the data element away from its original data type, is it possible to transmit the new data type in CDA®? *This is not possible using base CDA®, an extension would need to be defined.*
* Defining a CDA® extension for this data is not worth the effort that this will impose on users and implementers. Due to the small number of de-Identification points anticipated, and the use of CSV formats for analysis, CSV format is preferred.
* For data elements that may be either a string or a number, can we leave the format as “String or Number” or is that too difficult for implementers? I.e., for visit date where the value may either be “42” for the 42nd week of the year, or “3 visits in week 42”:
* Do you prefer that we leave this as String or Number; or
* Define this as a String; or
* Another solution?
* This issue is closed, as there should only be one Family Planning CDA® document per visit, and therefore only one visit date per input CDA® document.
* For administrative sex, what happens if a patient’s sex changes between encounters as a result of the generalization of “other” sexes to either male or female? Is this too identifiable? Should the CDA® entry for “other” simply be redacted?
* Changing “other” to “Female” only will not significantly impact statistical distribution or any of the performance measures that rely on Administrative Sex. Conclusion: Change all entries of “Other” to “Female” when de-Identifying.
* Where should we put the minimums and maximums for height and weight?
* Min/Max Height is 59 inches to 76 inches
* Min/Max Weight is 100-299lbs
* Decisions based on average height and weight data listed here: <ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2010/samadult_freq.pdf>

Volume 4 – National Extensions

Add appropriate Country section

4 National Extensions

# 4.R2 De-Identification for Family Planning data

De-Identified family planning data elements are required for performance measurement and federal reporting uses in the US realm. The users involved in those uses are:

* Clinicians who deliver services
* Quality managers or administrators at the site level
* Program managers
* Grant managers
* Regional monitors
* Office of Population Affairs (OPA) Health IT subject matter experts
* 3rd party analysts under contract to OPA

Analysis of these de-identification algorithms indicates that while they substantially reduce the risk of individual disclosure, it is not sufficient to allow the resulting data to be disclosed to a large group of stakeholders. As a result, there will need to be access and security controls on the resulting dataset to limit access to only authorized users.

If a dataset is to be made public and published, additional de-identification steps will be needed.

## 4.R2.1 Algorithms for the de-Identification of Family Planning data

The information elements in the Family Planning Clinical Document Architecture (CDA®) document shall be processed as shown in Table 4.R2.1-1. Each CDA® document describing an encounter shall result in a single line in a Comma Separated Value (CSV) file. CSV column and format assignments are described below.

Table 4.R2.1-1: De-identification Algorithms for Family Planning Data

| CDA® Element | De-identification Algorithm | CSV column number | CSV column format |
| --- | --- | --- | --- |
| Facility Identifier | Mapping table (see Section 4.R2.1.1). | 1 | String |
| Clinical Provider ID | Mapping table (see Section 4.R2.1.2). | 2 | String |
| Patient Identifier | Mapping table (see Section 4.R2.1.3). | 3 | String |
| Visit Date | Generalized to week of year (see Section 4.R2.1.4). | 4 | String or Number |
| Date of Birth | Convert to age in whole years, with no rounding. For clients under 18, grouped and mapped to “under 18”. For clients over 50, grouped and mapped to “over 50”. | 5 | String or Number |
| Administrative Sex | For values of “Male” or “Female” forward the data unchanged.  For Administrative Sex values of “other” change them to “Female” (see Section 4.R2.1.5). | 6 | String |
| Pregnancy History | Redacted. | - | - |
| Language of Communication | Collapse all forms to Limited English Proficiency (LEP) TRUE or LEP FALSE (see Section 4.R2.1.6). | 7 | String |
| Language Proficiency |
| Preferred Language |
| Ethnicity | Only the values “2186-5 Not Hispanic or Latino” or “2135-2 Hispanic or Latino” may be used. Any other input value must be converted to “2186-5 Not Hispanic or Latino”. | 8 | String |
| Race | Collapse to 5 OMB categories plus Other.  For each county, establish which races are below the threshold of 50 people per county. For those races, group them into “Other” (see Section 4.R2.1.7). | 9 | String |
| Annual Household Income | Convert to percentage of Federal Poverty Level (FPL) percentage. | 10 | Number |
| Household Size |
| Visit Payer | Convert to Public Health Information Network (PHIN) Vocabulary Access and Distribution System (VADS).  See the mapping table posted on the FTP site here: <ftp://ftp.ihe.net/IT_Infrastructure/iheitiyr13-2015-2016/Technical_Cmte/Workitems/DeIndentification%20of%20Family%20Planning/ReferenceCodes/> | 11 | String |
| Current Pregnancy Status | Convert to YES/NO/Unknown | 12 | String |
| Pregnancy Intention | Unchanged. | 13 | String |
| Sexual Activity | Unchanged. | 14 | String |
| Contraceptive Method at Intake | Use the appropriate supergroup for low probability types, otherwise pass the original data element through without changes (see Section 4.R2.1.8). | 15 | String |
| Reason for no contraceptive method | Use “the appropriate supergroup” for low probability types, otherwise pass the original data element through without changes (see Section 4.R2.1.8). | 16 | String |
| Contraceptive method at Exit | Use the appropriate supergroup” for low probability types, otherwise pass the original data element through without changes (see Section 4.R2.1.8). | 17 | String |
| Date of Last Pap test | Redact the day of the month, and use Week and Year only. | 18 | String |
| HPV Co-test Ordered | Redact the day of the month, and use Week and Year only. | 19 | String |
| CT Screen ordered | Redact the day of the month, and use Week and Year only. | 20 | String |
| GC Screen Ordered | Redact the day of the month, and use Week and Year only. | 21 | String |
| HIV Screen Ordered | Redact the day of the month, and use Week and Year only. | 22 | String |
| HIV Rapid Screen result | Delete  HIV reporting will be handled separately. | - | - |
| HIV Supplemental Result | Delete  HIV reporting will be handled separately. | - | - |
| Referral Recommended Date | Delete  HIV reporting will be handled separately. | - | - |
| Referral Visit Completed Date | Delete HIV referrals.  For non-HIV referrals redact the day of the month, and use Month and Year only. | 23 | String |
| Systolic blood pressure | Unchanged. | 24 | Number |
| Diastolic blood pressure | Unchanged. | 25 | Number |
| Height | Unchanged, except for values below 59 inches or above 76 inches.  For values below 59 inches, convert to 59 inches.  For values above 76 inches, convert to 76 inches. | 26 | Number |
| Weight | Unchanged, except for values below 100lbs or above 299lbs.  For values below 100lbs, convert to 100lbs.  For values above 299lbs, convert to 299 lbs. | 27 | Number |
| Smoking Status | Unchanged. | 28 | String |
| ***All other elements and attributes.*** | CDA documents permit additional elements and attributes beyond the minimum specified in a profile. If any such elements or attributes are present, they shall be removed. | - | - |

### **4.R2.1.1** Facility Identifier Mapping Table

A mapping table shall be maintained by the de-identifier that associates a real facility identifier with a pseudonymous identifier. These pseudonyms shall be created as version 4 (random) Universally Unique Identifiers (UUIDs). The Facility Identifiers from the inputted CDA® documents shall be converted to the UUIDs from the mapping table and output in the Facility Identifier column in a row in the CSV file.

The de-identifier shall maintain this mapping table under strictest access and security controls. There is no need to share it with any other party.

### 4.R2.1.2 Clinical Provider ID Mapping Table

A mapping table shall be maintained by the de-identifier that associates a real Clinical Provider identifier with a pseudonymous identifier. These pseudonyms shall be created as version 4 (random) UUIDs. The Clinical Provider Identifiers from the inputted CDA® documents shall be converted to the UUIDs from the mapping table and output in the Facility Identifier column in a row in the CSV file.

The de-identifier shall maintain this mapping table under strictest access and security controls. There is no need to share it with any other party.

### 4.R2.1.3 Patient Identifier ID Mapping Table

A mapping table shall be maintained by the de-identifier that associates a real Clinical Provider identifier with a pseudonymous identifier. These pseudonyms shall be created as version 4 (random) UUIDs. The Clinical Provider Identifiers from the inputted CDA® documents shall be converted to the UUIDs from the mapping table and output in the Facility Identifier column in a row in the CSV file.

The de-identifier shall maintain this mapping table under strictest access and security controls. There is no need to share it with any other party.

### 4.R2.1.4 Visit Date

Visit dates shall be transformed into an Integer denoting which week (out of 52 or 53, see ISO 8601) of the year the visit date took place on.

Note: This approach relies on there being a separate Family Planning CDA® document for each visit, even if there are multiple visits in a day or a week.

### 4.R2.1.5 Administrative Sex

Administrative Sex is not a clinical or genetic statement; it is used for administrative purposes.

Where Administrative Sex is Male or Female in the input CDA® document, this value shall be forwarded without modification.

Where Administrative Sex is listed as “other” this value shall be de-Identified by converting the values to “Female”.

### 4.R2.1.6 Limited English Proficiency (Language)

All three CDA® entries for language (Language of Communication, Language Proficiency and Preferred Language) shall be collapsed into one value, either LEP TRUE or LEP FALSE.

The value shall be LEP TRUE for Limited English Proficiency in the US according to the following derivation rules:

1. IF LanguageCommunication.LanguageCode=Eng AND LanguageCommunication.LanguageProficiency=Poor THEN LimitedEnglishProficiency=TRUE

or

1. IF LanguageCommunication.LanguageCode=Eng AND LanguageCommunication.PreferredIn ^= True THEN LimitedEnglishProficiency=TRUE

In English terms, this means:

1. If the Language of Communication is English AND the Limited English Proficiency is true, then the LEP value is TRUE; or
2. If the Language of communication is English, AND the preferred language is NOT English, then LEP value is also TRUE.

Otherwise, LEP FALSE shall be used.

### 4.R2.1.7 Race

All values for Race from the Input CDA® document that are not one of the 5 OMB categories below shall be converted to the most appropriate of the following categories:

* 1002-5 American Indian or Alaska Native
* 2028-9 Asian
* 2054-5 Black or African American
* 2076-8 Native Hawaiian or Other Pacific Islander
* 2106-3 White

Where one of the above categories contains fewer than 50 clients per county, convert all values for that category to:

* 2131-1 UNK Other Race

Please note that CCDA® allows for reporting of two or more races. If two or more races are reported, de-identify each one as above.

### 4.R2.1.8 Contraceptive Method

In regions where certain contraceptive methods are rarely used and therefore highly identifiable, change those methods to the appropriate grouping based on the efficacy of the method; “Highly Effective” “Moderately Effective”, or “Less Effective”.

## 4.R2.2 Example of De-Identified Family Planning Data

JB is a 16-year-old G-0 P-0 in the clinic for STI screening and well woman exam. Last menstrual period (LMP) was 3 weeks ago. No history of STI. BP: 110/75. Height: 157.5 cm. Weight: 58 kg. Intermittent condom use. Last unprotected sex was 2 weeks ago after which she used oral emergency contraception. Wants to have children “at some point, but no time soon”. Wants to use pills for contraception going forward. Non-smoker. Rapid HIV test is negative. Post visit, chlamydia results are positive and gonorrhea results are negative. No insurance can be billed at the time of the visit. Demographics: White, native US English speaker. Since 16 year olds seldom know their family income, JB’s FPL is calculated based on her own income from a part-time job, and her household size of 1. Income for JB is therefore approximately 44% of the Federal Poverty Level (see ASPE here: <http://aspe.hhs.gov/2015-poverty-guidelines#guidelines>).

Visit date: 22 Dec 2014

Geographic location: HHS Region 4 (Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, and Tennessee)

| Data Element | Original Data | Data after application of de-identificatioN |
| --- | --- | --- |
| **Patient Identifier** | [patient ID from SUT] | [Mapped patient ID=333-333] |
| **Date of Birth** | 5 June 1998 | Under 18 |
| **Administrative Sex** | Female | Female |
| **Language of Communication** | en-US | LEP FALSE |
| **Language Proficiency** | NULL |
| **Preferred Language** | True |
| **Race** | White=2106-3 | 2106-3 |
| **Ethnicity** | Not Hispanic or Latina=2186-5 | 2186-5 |
| **Clinical Provider** | [provider ID from SUT] | [Mapped Provider ID = 222-222] |
| **Visit Date** | 22 Dec 2014 | W52 2014 |
| **Facility identifier** | [facility ID and address from SUT, but from HHS Region 4] | [Mapped facility ID = 111-111] |
| **Number of Total Pregnancies** | 0 | DELETED |
| **Current Pregnancy Status** | Not pregnant, by test=2 | NO |
| **Pregnancy Intention** | No, but maybe in the future= N | N |
| **Sexual Activity** | True | True |
| **Contraceptive Method at Intake** | EC=19 | Moderately Effective |
| **Reason for No Contraceptive Method at Intake** | NULL | NULL |
| **Last Cervical Cancer Screen (Date of last Pap test)** | NULL | NULL |
| **HPV Co-Test** | 22 Dec 2014 | W52 2014 |
| **Contraceptive Method at Exit** | OCP=7 | 7 |
| **Reason for No Contraceptive Method at Exit** | NULL | NULL |
| **Chlamydia trachomatis Screen Order** | 22 Dec 2014 | W52 2014 |
| **Neisseria gonorrhoeae Screen Order** | 22 Dec 2014 | W52 2014 |
| **HIV Screen Order** | 22 Dec 2014 | W52 2014 |
| **HIV Rapid Screen Result** | HIV Rapid Screen Result, Negative=NEG | DELETED |
| **HIV Supplemental Result** | NULL | DELETED |
| **Referrals Planned** | NULL | DELETED |
| **Referrals Completed** | NULL | NULL |
| **Height** | 157.5 cm | 62 inches |
| **Weight** | 58 kg | 128 |
| **Systolic Blood Pressure** | 110 | 110 |
| **Diastolic Blood Pressure** | 75 | 75 |
| **Smoking Status** | Never smoker=266919005 | 266919005 |
| **Annual Household Income** | $5,000 | FPL 44% |
| **Household Size** | 1 | DELETED |
| **Insurance** | No Insurance=NA | NA |

In an excel spreadsheet, the de-Identified row for the above encounter would look like this:



The corresponding comma-delimited (CSV) row for JB’s de-Identified family planning encounter is:

111-111,222-222,333-333,W52 2014,Under 18,Female,LEP FALSE,2186-5,2106-3,44,NA,NO,N,True,Moderately Effective,NULL,7,W52 2014,NULL, W52 2014,W52 2014,W52 2014,NULL,110,75,62,128,266919005

Note: UUIDs for the Facility, Provider and Patient ID are provided as an example only. Correct UUIDs are hexadecimal numbers that are 32 characters long separate by dashes.

Note: The above example should be only one line long, but document formatting splits inappropriately.