User Guide

IIS Deduplication

Prepared by STChealth
May 2020

STChealth
411 South 1st Street
Phoenix, AZ 85004
480.745.8500 phone
602.598.7712 fax
www.STChealth.com

Copyrights and Trademarks

© 2020 by STChealth. All rights reserved.

This document describes the following: IIS Deduplication procedures and logic.

No part of this publication may be altered, reproduced, transmitted, stored in a retrieval system, or translated into any human or computer language, in any form or by any means, electronic, mechanical, magnetic, optical or otherwise, without the prior written permission of the copyright owner.

This publication, along with any embedded media, is provided for informational purposes only. STChealth provides this material 'as is' and does not assert that this publication is error free. The information contained herein may not necessarily accurately represent the current or planned functions of the application, and may be subject to significant and frequent modification. STChealth may change this publication, any embedded content and the product described herein at any time. Any changes will be incorporated in new versions of this publication.

STChealth recognizes the rights of the holders of all trademarks used in its publications.

This publication may provide hyperlinks to third-party websites or access to third-party content. Links and access to third-party sites are provided for your convenience only. STC health does not control, endorse, or guarantee third-party content and is not responsible for any content, associated links, resources or services associated with a third-party site. STChealth shall also not be liable for any loss or damage associated with your use of any third-party content.

Table of Contents

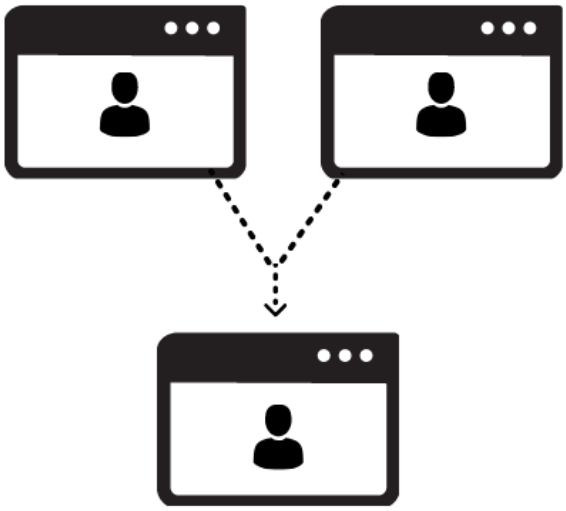
Introduction 5
Getting Started - Administrative Settings 6
Patient Settings 7
Deduplication Properties 8
Organization Settings 9
PHC Hub Settings 10
Deduplication in Action 11
Patient Deduplication 12
Database Tables and Deduplication Procedures for Record Storage 13
Deduplication Rules, Logic & Calculations 14
Match Codes/Match Numbers 14
Distinguishing Match Sets 15
Definite Match Sets 15
Automatic Master Duplicate Procedure (scanning): 16
Vaccination Deduplication 17
Administrative Deduplication Functions 19
Addendums 20
A. Patient Settings 20
B. Deduplication Properties 22
C. Organization Settings 26
D. PHC Hub Settings 27
E. Database Tables and Deduplication Processes for Record Storage 29
F. Deduplication Rules, Logic & Calculations 32
G. Deterministic (Default) Rules/ Logic 35
H. Administrative Deduplication Functions 44
H.1 Manual Deduplication 44
H.2 Ambiguous ID 44
Ambiguous ID Resolution Workflow 46
H.3 Master Deduplication by External Source 47
H.4 Merge History Report 47

H.5 User Reported Duplicates 47
H.6 Separate Bad Merges 48
H.7 Run Deduplication 48
H.8 Schedule Recurring Deduplication 48
H.9 Stop Scheduled Deduplication 48
H.10 Auto Deduplication 49

Introduction

Deduplication is one of the most fundamental functions within IWeb, a process that consolidates information coming from a variety of clinical sources. The IIS should contain complete and timely demographic immunization information for children, adolescents and adults residing or immunized within its jurisdiction.[1] The consolidation of information from a variety of contributing sources about the patient or the vaccination, as well as information submitted by the same organization on multiple different occasions, leads to a more accurate and complete reflection of reality in the IIS.[2] When the IIS contains complete or nearly complete records for every patient, the data quality is significantly increased, providers can feel confident in their decisions made at the point of care and population level coverage assessments can be regarded as accurate.

Deduplication occurs at both the demographic and the vaccination level, and includes automatic and manual procedures to ensure patient record accuracy. This document will describe the automatic procedures that occur, the various decisions that are made about a record, and the tables within the database that the records are stored in. This guide is intended to be used by programmatic and technical staff that are responsible for maintaining operations of the IIS.



[1] https://www.cdc.gov/vaccines/programs/iis/functional-standards/func-stds-v4-1.html

[2] Consolidating Demographic Records and Vaccination Event Records, 2017

Getting Started - Administrative Settings

Prior to deduplication occurring in the IIS, there are a number of settings/properties that must be defined according to the needs of the immunization program. Administrators have some control over how the patient and vaccine records are handled during the automatic deduplication process, the general volume of records that end up in manual deduplication and how deduplication behaves in general. These settings/properties are housed in different tables within the IIS Administration menu, and are defined below.

This guide accomplishes the following goals:

1. How to get started - describes how to configure the deduplication specific properties and settings to align with individual jurisdiction requirements
2. Deduplication in action - describes the automatic deduplication procedures for both demographic and vaccination information
3. Describes the manual deduplication tools and procedures that are available to the operators of the IIS

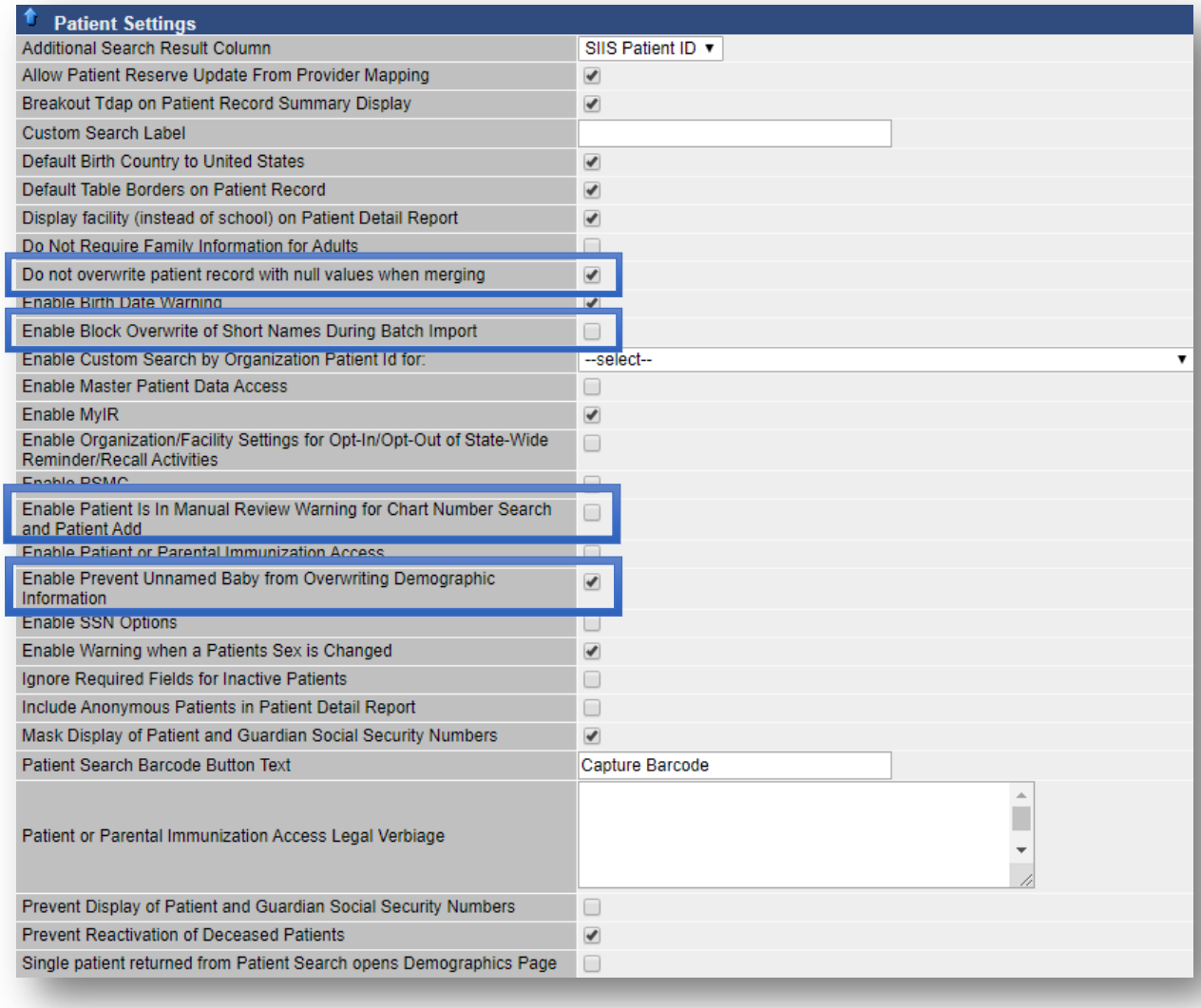
Patient Settings

Located towards the bottom left menu after logging into IWeb

1. Log into IWeb
2. Click **Administration**
3. Click **Properties**
4. Click **Patient Settings**

The following identified patient settings apply to deduplication:

See Addendum A for fully displayed descriptions

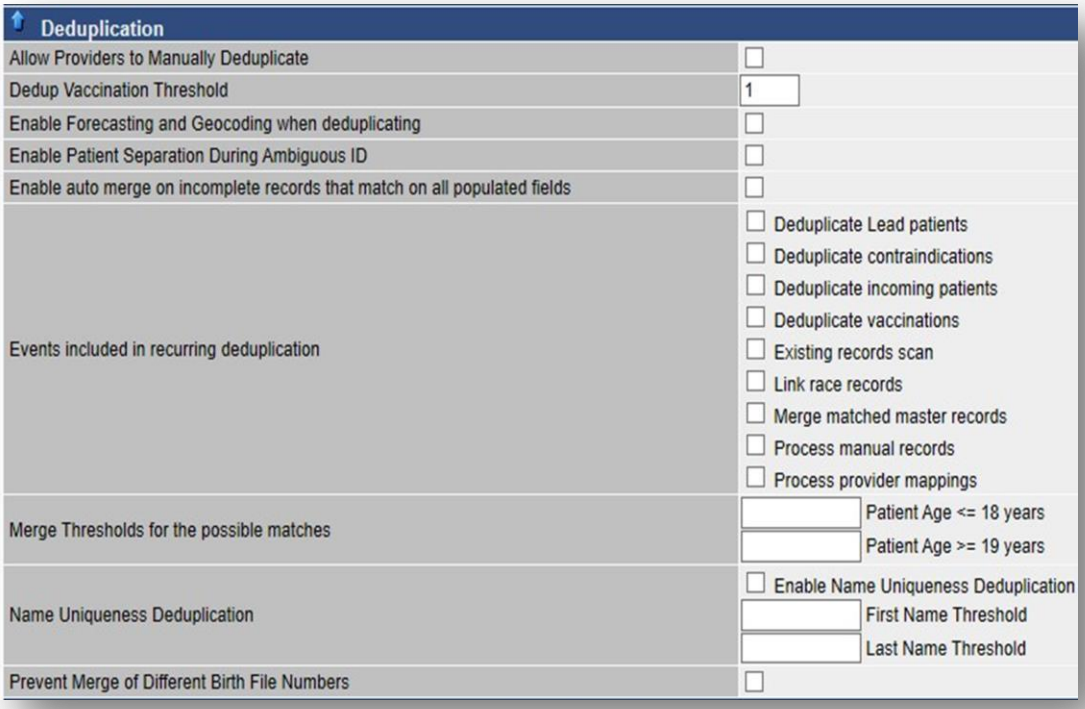


Deduplication Properties

Located towards the bottom left IWeb menu

1. Click **Administration**
2. Click **Settings**
3. Click **Properties**
4. Click **Deduplication Properties**

See Addendum B for fully displayed descriptions



Organization Settings

Located in the upper left IWeb menu

1. Click **Organization**
2. Click **Detail**
3. Click **Vaccine Management**

The following organization settings apply to deduplication:

See Addendum C for fully displayed description

Vaccine Management	
Certifying Provider:	
Certifying Provider Phone:	
Certifying Provider Fax:	
Certifying Provider Email:	
Agreement Signatory Name:	
Agreement Signatory Title:	
Original VFC Cert Date:	
Current Certification Date:	
Freezer average temperature of 5°F or colder:	N
Separate, insulated freezer door:	N
Freezer Measuring Device:	Standard fluid filled
Allow Online Orders:	
Allow Vaccine Transfers:	
Allow Vaccine Returns:	
Default physical inventory on order:	
Default doses used last month on order:	
Order Threshold Percent:	
Ordering Frequency:	
Order Timing:	
Order Frequency Start Month:	
Recommended Quantity Calculation:	BASED ON ADMINISTERED DOSES
Enable Auto-Receive:	
Enable Barcode Tools:	
Order Set / Order Type:	
Allow only one brand selection per vaccine:	
Require Manual Deduplication for Intra-Organization Merges:	Y
Bypass Dose-Level Accountability:	
Upload Temperature Data:	
Facility Required for this Organization:	Default
Organization Last Updated Date:	02/21/2020 07:20:30 PM
Organization Last Updated By:	ACHAN
Organization Created Date:	02/21/2020 07:08:09 PM
Organization Created By:	ACHAN

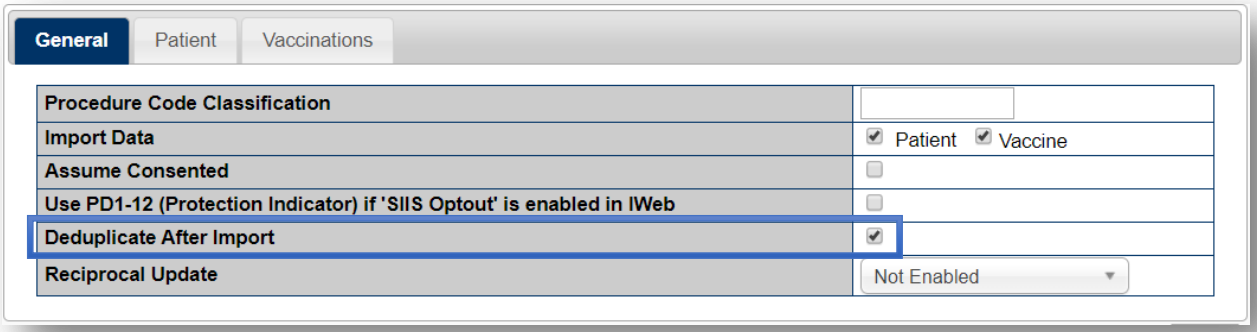
PHC Hub Settings

Located in the middle left menu after logging into PHC Hub

1. Click **Import Profile**
2. Click **Select Import Profile**
3. Select the appropriate profile
4. Click **Edit**
5. Click the **General** tab from the **Update Settings** menu

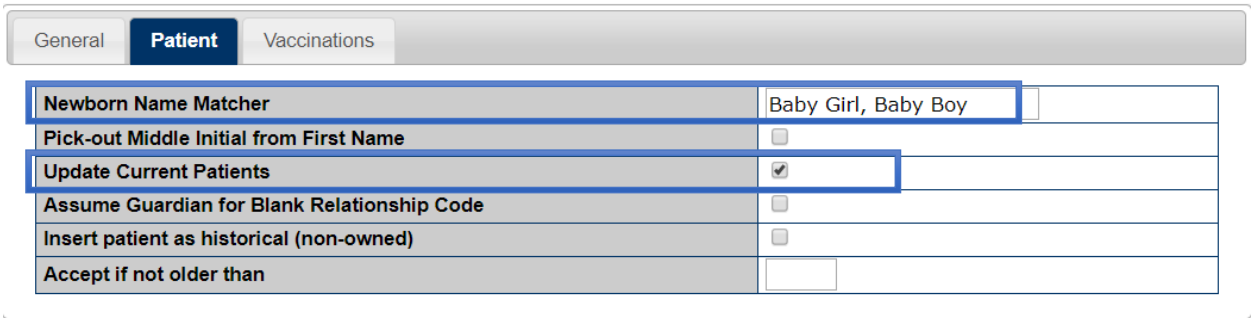
The following PHC Hub settings apply to deduplication

See addendum D for fully displayed descriptions



From the Update Settings menu

1. Click the **Patient** tab



Deduplication in Action

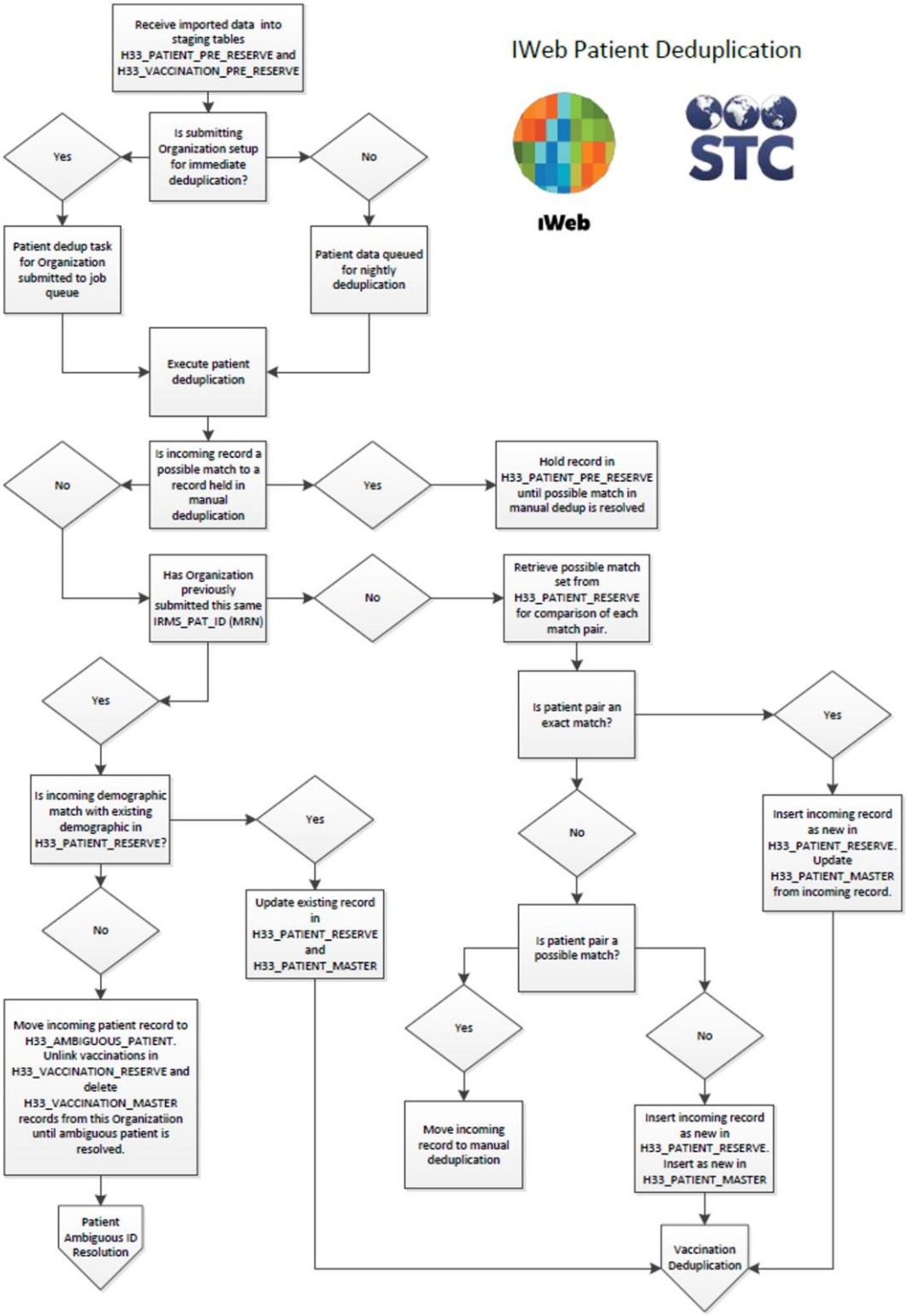
The registry is designed to minimize the number of duplicate patient and vaccination records within it, but it is still possible for duplicates to occur. As records come into the system either through manual entry, DTT or HL7, decisions are made about the data to determine whether or not there is already a match in the system. Those decisions are made using various deduplication mechanisms and result in one of the following three scenarios:

1. If there is a good match for exactly one patient record in the registry, the two records are automatically merged together
2. If there is not a good match with any patient records in the registry, it is automatically added as a new patient record
3. If there is a possible match with one or more patient records, the records must be manually reviewed before they can be merged

The following information describes the automatic deduplication processes that are in play when a record is submitted to the IIS. These processes are not observable to the human eye, but it is useful to understand the workflows, decisions made and database tables that information passes through.

Patient Deduplication

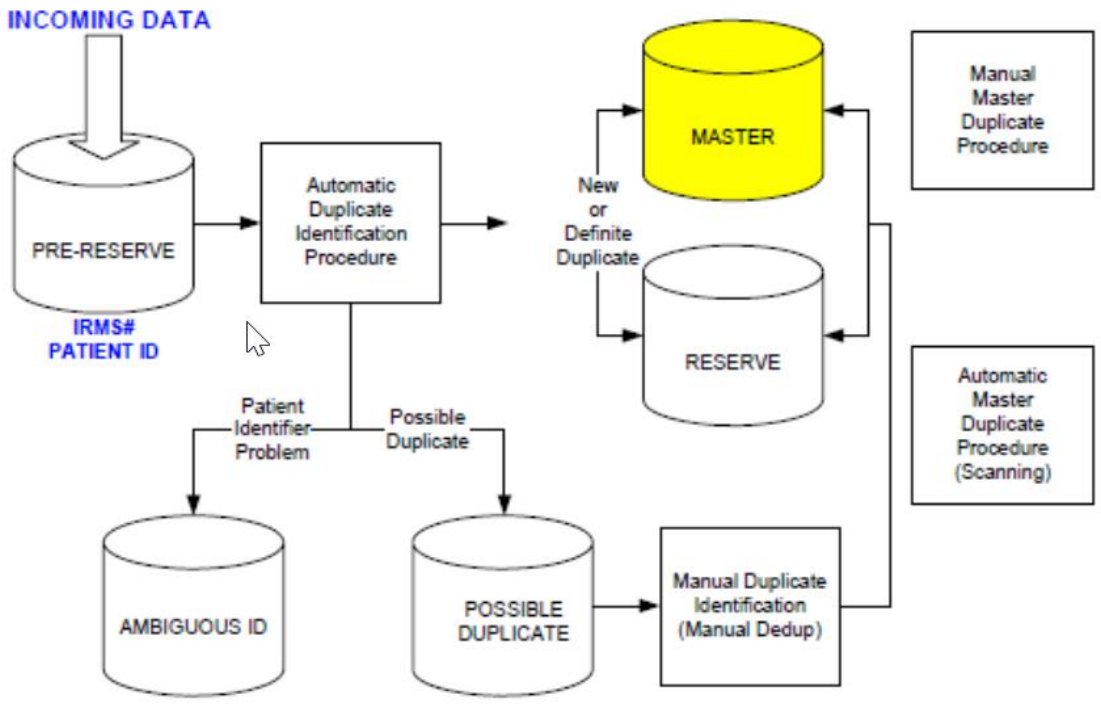
Below is a visual representation of the workflows that a patient record will go through



Database Tables and Deduplication Procedures for Record Storage

A patient record must pass through a number of database tables and deduplication procedures. There are four types of procedures used to determine whether or not a patient record is a duplicate. The image below shows each procedure, when it may run and the database tables that the record may encounter through its deduplication journey.

The procedures are represented in the squares, the database tables are represented in the cylinders.



See Addendum E for fully displayed descriptions of procedures and database tables that a record may encounter

Deduplication Rules, Logic & Calculations

See Addendum F for full descriptions of how rules, logic and calculations are applied

*** Attention IIS users: Refer to this section when patients end up in the manual merge queue, and you have questions regarding why they ended up there, and didn't automatically merge*.

The SIIS deduplication algorithm compares an incoming patient record to records in the patient reserve table that have similar data elements. To quickly perform the match, each incoming record is compared against the patient reserve using the match codes/match numbers model detailed below to create a subset of records for comparison. The incoming record is then evaluated with each record in the subset at a more finite level using groups of data elements called match sets.

Each match set is given a score based on the likelihood that it does not contain data elements that would be shared by more than one person. For example, a match on First Name, Last Name, Birthdate and Medicaid Number would belong to only one person and would be considered an *exact match*. A match on First Name, Last Name, Birthdate and city could belong to more than one person and would be a *possible match*. Key elements such as First Name, Last Name, Birthdate and another item such as Address or Guardian are most helpful in achieving an exact match. An exact match is given a score of 2 and is automatically merged. A score of 1 indicates a *possible match* and the record is sent to manual review. A score of 0 indicates no match and the incoming record is inserted as a new patient.

Match Codes/Match Numbers

Match codes and match numbers are generated when a patient record is inserted into the database. They are stored in the h33_patient_reserve and h33_patient_mater tables. These codes/numbers are used to return a subset of records that might be possible duplicates to an incoming record being processed through deduplication.

Match_Code5	Address Street, Address Short City Example: 100 Main, Phoenix, AZ Match_Code5 = 100MAIN PHOENIX
Match_Num1	Patient birth date Year, Month, Day and First Initial as an integer Example: Patient first name (Bob), Patient Birth Date 10/01/1980 Match_Num1 = 1980100101
Match_Num2	Patient birth date Year, Month, Day and Last Initial as an integer Example: Patient last name (Jones), Patient birth date 10/01/1980 Match_Num2 = 1980100110

Distinguishing Match Sets

Distinguishing match sets are a result of in-depth comparison of an incoming patient record with the records that have been selected based on the match codes/numbers described above. An incoming patient record is compared to records in the patient reserve table. Each match set receives a composite score based on the likelihood that it is a match to the record.

Score = 2	Most likely a match, automatic merge
Score = 1	Possible match, move to manual deduplication
Score = 0	No match

Definite Match Sets

Definite match set statements that are 'true' are assigned a composite score of 2 and are automatically merged. For example, in Set 100 below (see *Addendum G*), if both Medicaid Number and Birth File Number fields were matched on both patient records, then each field type would have been scored as 1. Therefore, 1 + 1 = 2, the statement is true, and the composite score would be a 2, resulting in an automatic merge.

Birth order and multiple birth count are used to negate a match set if these items are populated on both the incoming and the existing records but do not match.

Automatic Master Duplicate Procedure (scanning):

Automatic Master Duplicate Procedure is a nightly automatic process that reviews all of the records in the patient master table.

For example:

> A patient named John Aaron Smith, born on 1/1/1999, enters the registry and is displayed in manual review against a patient named John Aaron Smitty, born on 10/10/1999.

Since these records are not similar enough to be safely merged, the person performing the manual review decides that the records are different, thus resulting in two separate patient records. Weeks later, the sender of the record for patient Smitty corrects the record and changes it to John Aaron Smith, born on 1/1/1999. Now the records are similar enough to be merged. Since both records exist in the master tables as two different patients, the Master Deduplication process is used to merge them.

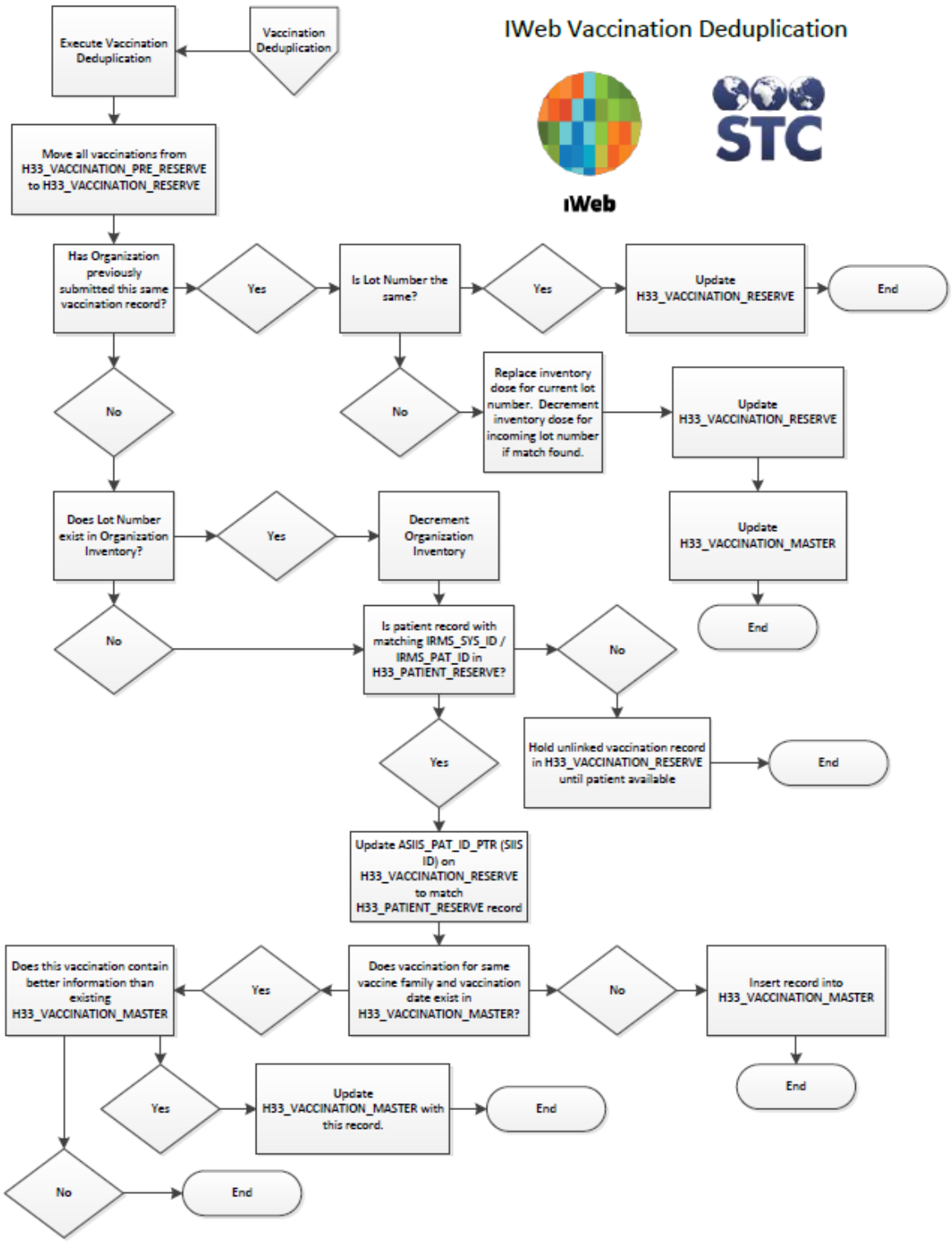
Vaccination Deduplication

For most patients, vaccination event information is being sent to IWeb from a variety of sources, whether from a physician's office, the pharmacy, the hospital or the billing department. This leads to multiple duplicate vaccinations for the same vaccine family and the same vaccination date may exist for more than one of the merged patient records. Many providers submit historical vaccinations (administered by the provider on past visits or administered by another provider), as well as administered vaccinations (administered by the current provider on present visit).

During the data entry process, a provider can view the existing vaccinations on a patient record and should not duplicate entries. However, a provider submitting batch data has no knowledge of the vaccinations already in the registry and may send records that have already been recorded.

Vaccinations submitted as batch data may be submitted as a separate file from the patient data. The vaccination record must contain a unique identifier linking it to the patient record to which it belongs. During the deduplication process, this identifier is maintained so that as patient records are merged or separated, the vaccinations are related to the correct patient. An identifier for the merged patient record is maintained with the vaccination to create a consolidated vaccination record.

This is a visual representation of the deduplication workflow for vaccines:



Administrative Deduplication Functions

See Addendum H for full descriptions of each function

Located in the bottom left IWeb menu

1. Click on **Administration**
2. Click on **Deduplication**

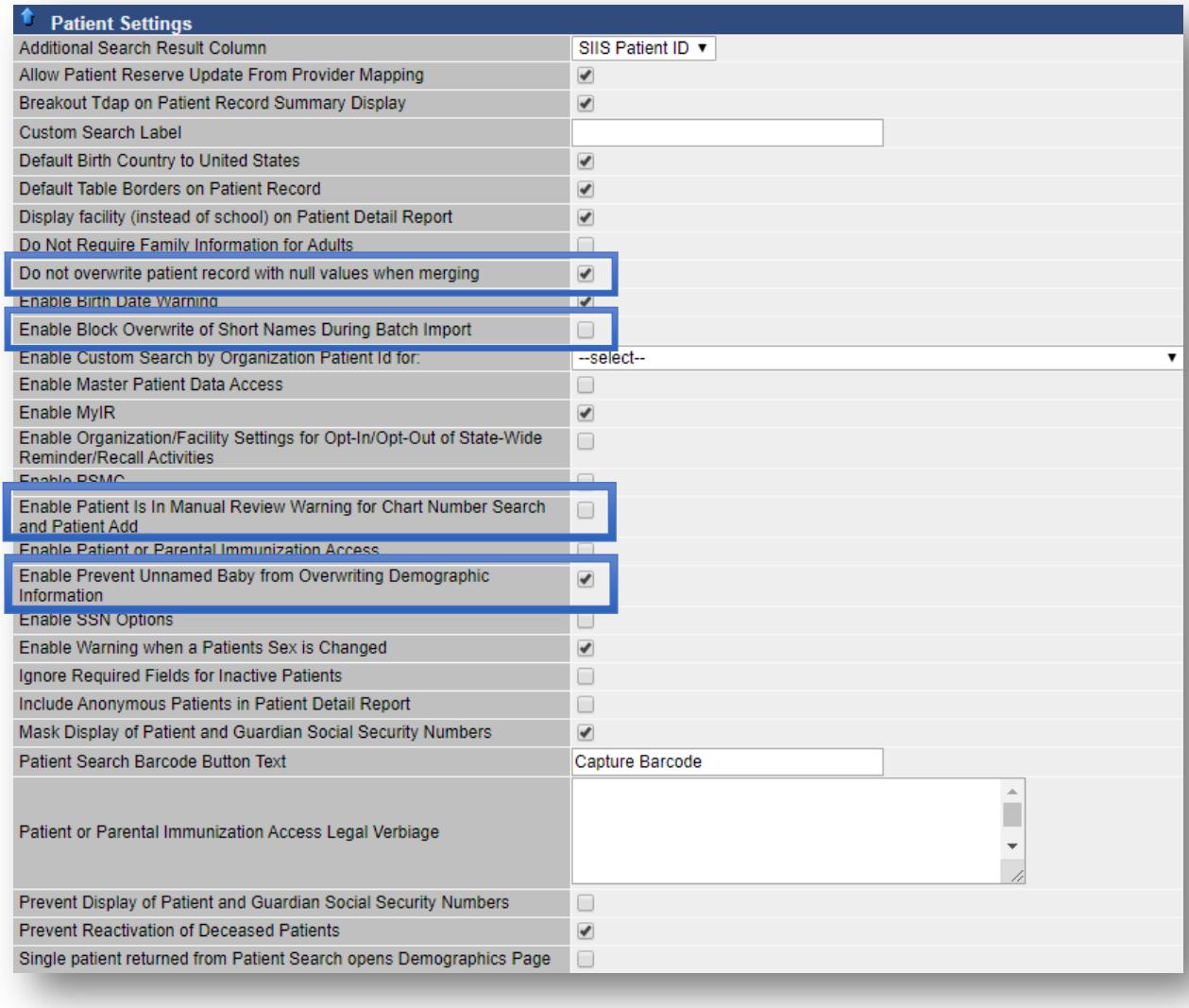
The following functions must be completed manually from within IWeb, and are only available to those Registry Client users with the permission of *System Administrator*, unless *Allow Providers to Manually Deduplicate* is checked in the *Deduplication Properties* table. Each function should be performed on an as needed basis.

- Manual Deduplication
- Ambiguous ID
- Master Deduplication by External Source
- Merge History Report
- Define Deduplication Reasons
- Separate Bad Merges
- Run Deduplication
- Schedule Recurring Deduplication
- Stop Scheduled Deduplication
- Automatic Deduplication

Addendums

A. Patient Settings

1. Click on **Administration**
2. Click on **Properties**
3. Click on **Patient Settings**



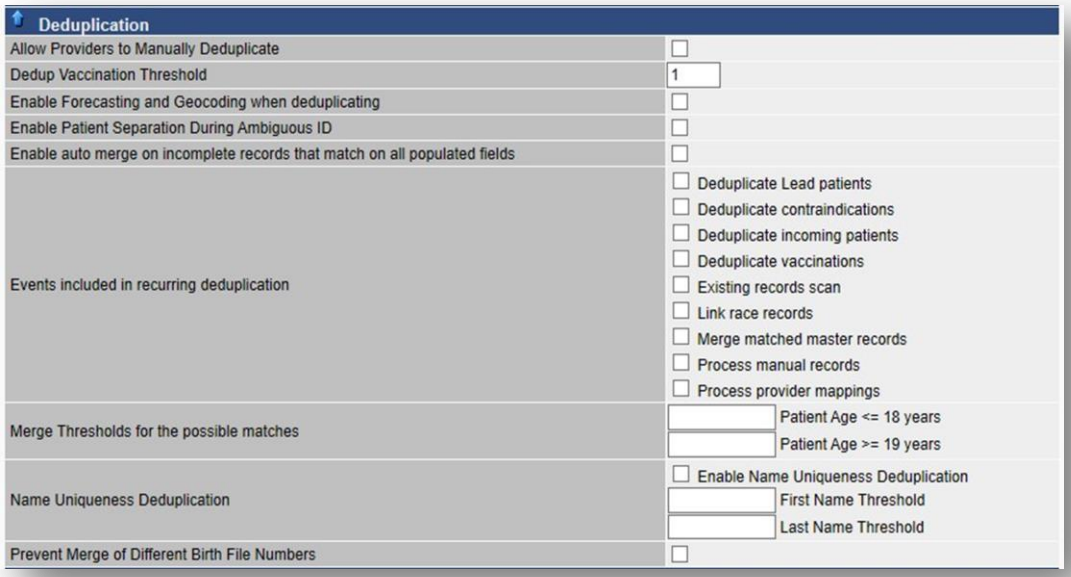
- *Do not overwrite a patient record with null value when merging* - If an incoming record with blank fields is received, the blank data will not overwrite existing data.
- *Enable Block Overwrite of Short names during Batch Import* - If you have a formal name, it won't overwrite it with a nickname in the name field.
- *Enable Patient is in Manual Review Warning for Chart Number Search and Patient Add* - If a patient gets stuck in manual deduplication, then a user can be warned that the patient result (In

patient search) is not showing up because the patient is being reviewed currently in manual deduplication.

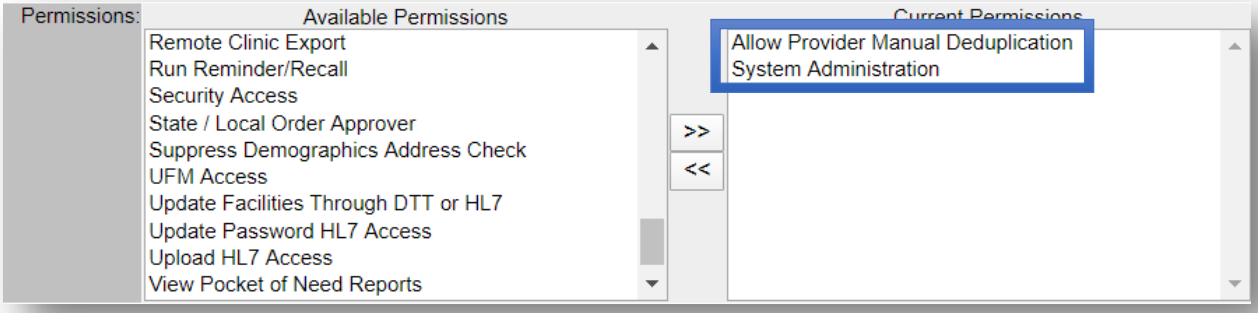
- *Enable Prevent Unnamed Baby from Overwriting Demographic Information* - When data is imported and the baby is not named such as Baby Girl, if it does match on a record in the system, it will not overwrite the actual name with the placeholder baby name.

B. Deduplication Properties

1. Click on **Administration**
2. Click on **Settings**
3. Click on **Properties**



- *Allow Provider to Manually Deduplicate* - check this box to give Organization Client users and Facility Client users access to review and process patients in Manual Deduplication and Separate Bad Merge queues that have been submitted by their organization / facility. Uncheck box to only allow State Administrator access to manual deduplication
    - Restriction: User must also be given the *System Administration* permission.



- *Dedup Vaccination Threshold* - when checked, evaluates duplicate vaccines that are sent for the same patient. The default is 1 day, so any incoming vaccination from the same family with a vaccination date of +/- 1 day from the previous vaccination is considered a duplicate. Each IIS administration should set threshold based on how deduplication is resolved locally. It is evaluated to determine if it will replace the current vaccination with more complete information, but it is not inserted as new.
    - Example:

Existing DtaP Vaccine	Incoming Infanrix Vaccine	Threshold	Outcome
01/01/2019	01/03/2019	1	Entered as two vaccines
01/01/2019	01/03/2019	3	Considered a duplicate. Most recent complete vaccine saved to the master

- *Enable Forecasting and Geocoding with deduplication* - enabling this property will cause both forecasting and geocoding to process when deduplication is run from the user interface. Forecasting will involve generating a vaccination forecast for all patients whose vaccination records have changed. If Geocoding is enabled in the registry, it will run as the last process since it can take a lengthy amount of time. Due to the lengthy process, a nightly backup may interrupt it. However, it continues where it left off during the next running of the procedure. Only patient records that are new or have had a change of address are geocoded.

- *Enable Patient Separation During Ambiguous ID - this property should NOT be enabled.* This property will be removed in time.

- *Enable auto merge on incomplete records that match on all populated fields* - when enabled, records are compared on a populated field basis. If all *populated* fields match, yet some fields are incomplete, records would auto merge. When not enabled, if one patient has a value populated in a field and the other patient does not, the field is not considered a match.
    - Example:

	EXISTING PATIENT	INCOMING PATIENT
First Name	Bob	Bob
Last Name	Smith	Smith
DOB	01/01/1990	01/01/1990
Mother's Maiden Name	Jones	
SSN		123-45-6789
	No Other Data Populated	No Other Data Populated

These two patients are merged automatically and not queued for manual review. If the existing record had a different SSN, these records would not automatically merge. Special street address processing logic is also used. If the existing address is longer than eight characters and the full string exists within the incoming street address string, the two addresses are considered a match and the address is updated with the incoming address (the longer address is more likely to be accurate). This process uses the portion of the "mangled" (normalized) address prior to the state/zip code.

Example: these addresses are considered a match and the address is replaced with the incoming address

Existing Address	1033 Pine Valley Road
Incoming Address	1033 Pine Valley Rd. #103

- *Events included in recurring deduplication* - select which deduplication processes will run when data is submitted using the HL7 Real time processor, and when deduplication is run from the UI. *These settings do not apply to nightly deduplication.*

Deduplicate lead patients	When checked, processes imported DTT lead records
Deduplicate contraindications	When checked, processes imported HL7 & DTT contraindications
Deduplicate incoming patients	When checked, processes imported HL7 & DTT patient records
Deduplicate vaccinations	When checked, processes imported HL7 & DTT vaccination records
Existing records scan	When checked, calls the Master Duplicate Scan procedure which reviews all of the records in the Patient Master Table and merges the similar records
Link race records	When checked, processes the imported HL7 & DTT race field on patient records
Merge matched master records	When checked, calls the Deduplication Simple Master procedure which merges matching reserve records into one record
Process manual records	When checked, processes the records that were addressed during manual deduplication
Process provider mappings*	When checked, processes records that need to be changed based on provider mapping

* Note: Provider Mappings are done when a third party enters in patient data and wants to map ownership to the correct provider. This prevents the third party from taking ownership of records that do not belong to them and maps them to the provider that the patient received vaccines from.

- *Merge thresholds for possible matches* - enter in values in order to define the total score required to consider the records an exact match and automatically merge. This is directly related to the weight stored procedure. There are two different age options since you might want to consider merges differently depending on the demographic, Patient Age <= 18 and Patient Age > 19. The

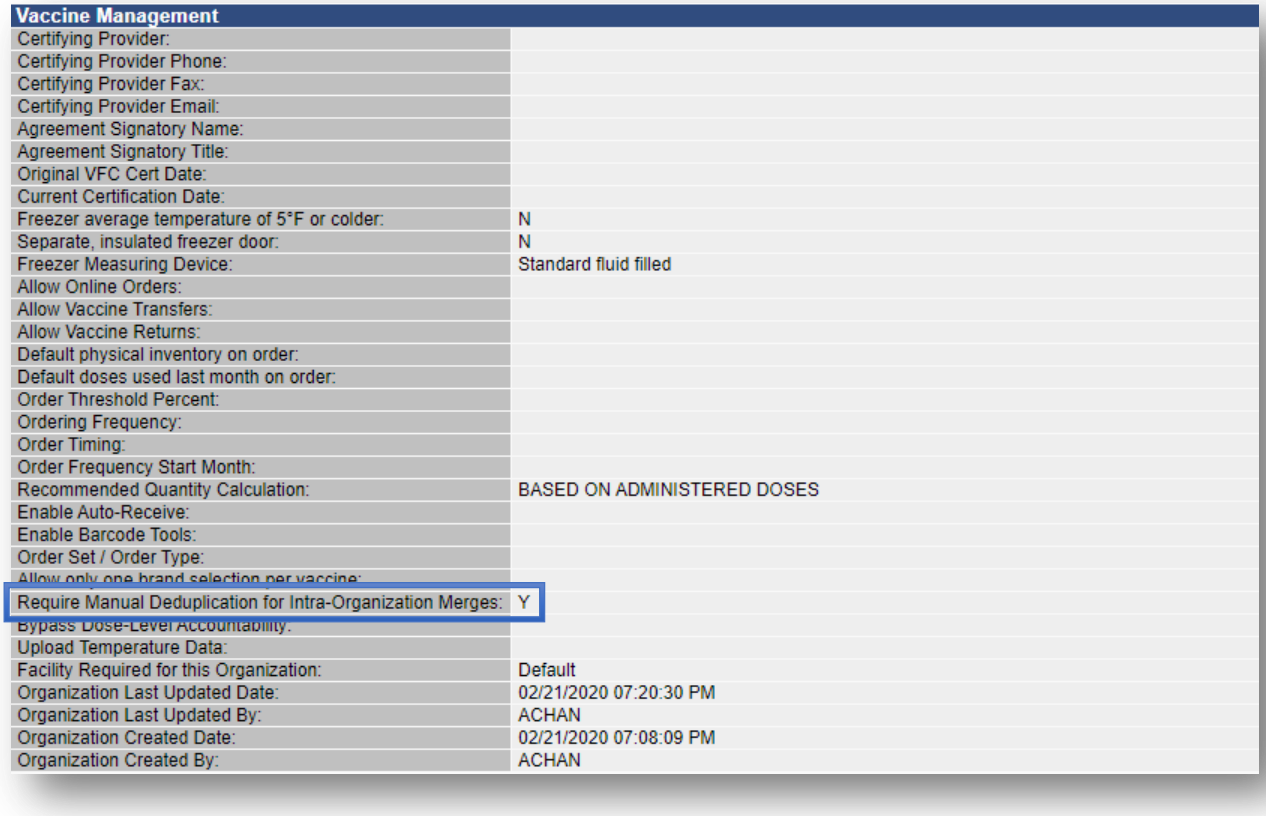
total approximate score if all fields match is 45. When looking at the scores, the items marked with an asterisk are grouped together and only one of the values would be counted if there was a match. The vaccinations are not used in the threshold matching.

IRMS SYS ID	1	bd only match	0
IRMS FAC ID	1	bd no match	-2
MEDICAID #	3	race match	1
BIRTH FILE #	3	gender match	1
SSN #	4	address, city, state and zip match	1
fn match &> name threshold*	1	address, zip and street 1 match	2
fn match &<= name threshold*	2	address, zip and first 7 characters of street 1 match	1
fn match, ln match & middle match **	2	phone, bd, fn & ln match	2
(fn or ln match) & middle match**	1	phone match only	1
ln match &> name threshold***	1	email match	2
ln match & <= name threshold***	2	guardian fn match	1
pt suffix match	1	guardian mn match	1
fn, mn, ln & birthdate match	4	guardian ln match	1
fn, ln & birthdate match	3	guardian 2 fn match	1
(fn or ln match) & birthdate match	2	guardian 2 ln match	1
		guardian 2 ssn match	1

- *Name Uniqueness Deduplication* – enable this property to evaluate name uniqueness based on the frequency of the name in the database. The threshold fields represent the percentage of frequency. For example, the incidences of 'James' divided by the total number of records. The threshold is a percentage defining the match level for uniqueness.
    - The deduplication logic is as follows: If the patient is otherwise a possible match and the first name and last name both match exactly, if the name is unique per the threshold configuration, allow an exact match on First Name, Last Name, and Birth Date.
        - Not to be set above 0.95
- *Prevent Merge of Different Birth File Numbers* – when enabled, the import logic evaluates each patient's birth file number and prevents a merge if they are different. If this property is enabled and the records are an exact match with different birth file numbers, the match is negated and goes into the manual deduplication queue. *This does not apply to master merges or user reported duplicates*
    - A birth file number match automatically sets the record as a possible match. After that, other fields are evaluated to determine if it moves to an exact match.
    - When not enabled, birth file numbers are not evaluated in the merge process.
    - When not enabled, record matches with different birth file numbers will be merged.

C. Organization Settings

1. Click on **Organization**
2. Click on **Detail**
3. Click on **Vaccine Management**



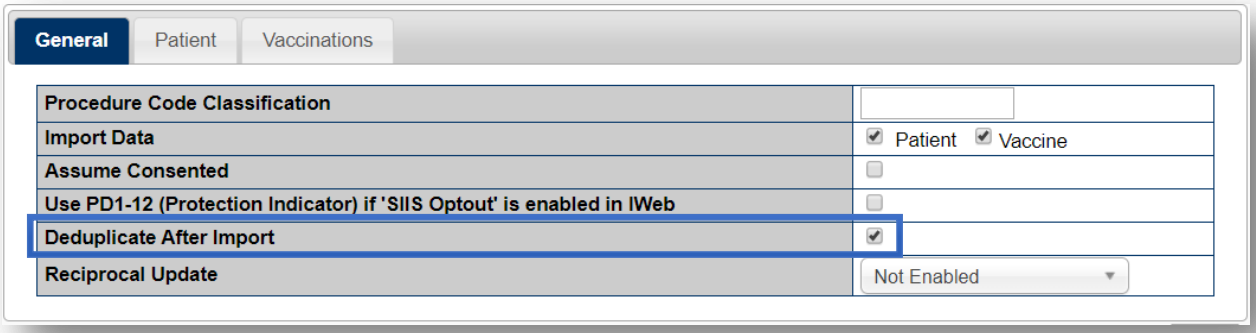
Vaccine Management	
Certifying Provider:	
Certifying Provider Phone:	
Certifying Provider Fax:	
Certifying Provider Email:	
Agreement Signatory Name:	
Agreement Signatory Title:	
Original VFC Cert Date:	
Current Certification Date:	
Freezer average temperature of 5°F or colder:	N
Separate, insulated freezer door:	N
Freezer Measuring Device:	Standard fluid filled
Allow Online Orders:	
Allow Vaccine Transfers:	
Allow Vaccine Returns:	
Default physical inventory on order:	
Default doses used last month on order:	
Order Threshold Percent:	
Ordering Frequency:	
Order Timing:	
Order Frequency Start Month:	
Recommended Quantity Calculation:	BASED ON ADMINISTERED DOSES
Enable Auto-Receive:	
Enable Barcode Tools:	
Order Set / Order Type:	
Allow only one brand selection per vaccine:	
Require Manual Deduplication for Intra-Organization Merges:	Y
Bypass Dose-Level Accountability:	
Upload Temperature Data:	
Facility Required for this Organization:	Default
Organization Last Updated Date:	02/21/2020 07:20:30 PM
Organization Last Updated By:	ACHAN
Organization Created Date:	02/21/2020 07:08:09 PM
Organization Created By:	ACHAN

- *Require Manual Deduplication for Intra-Organization (IRMS) Merges* - select to have the system queue records for Manual Review if a new patient is an exact match to an existing patient within the same Organization (IRMS). If this option is disabled, the current deduplication behavior is applied.
    - Note: If both the Organization (IRMS) system ID and the medical record number (or Organization patient ID) are the same for the patient, the record is considered to be an update and continues to be processed automatically.

D. PHC Hub Settings

From the left PHC Hub menu of the home page

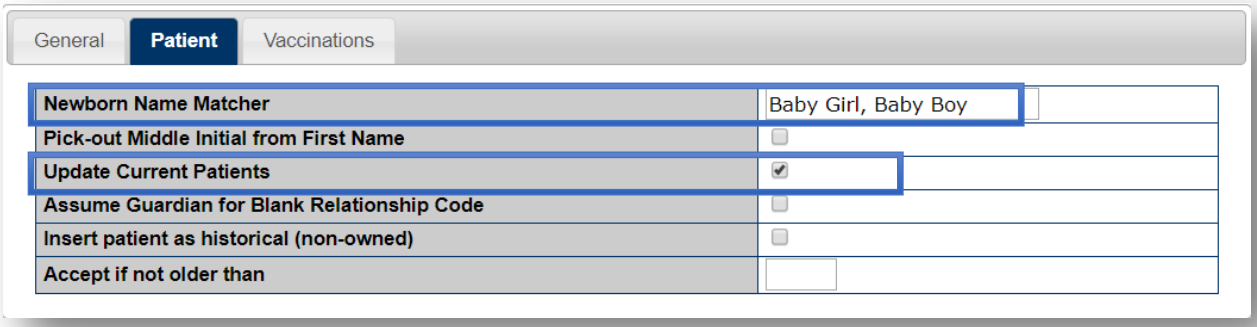
1. Click on **Import Profile**
2. Click on **Select Import Profile**
3. Select the appropriate profile
4. Click on **Edit**
5. Select the **General** tab from the Update Settings menu



- *Deduplicate After Import* - select to enable automatic, immediate deduplication of incoming data. Deselect for data to be duplicated later, generally after business hours

From the Update Settings menu

1. Select the **Patient** tab

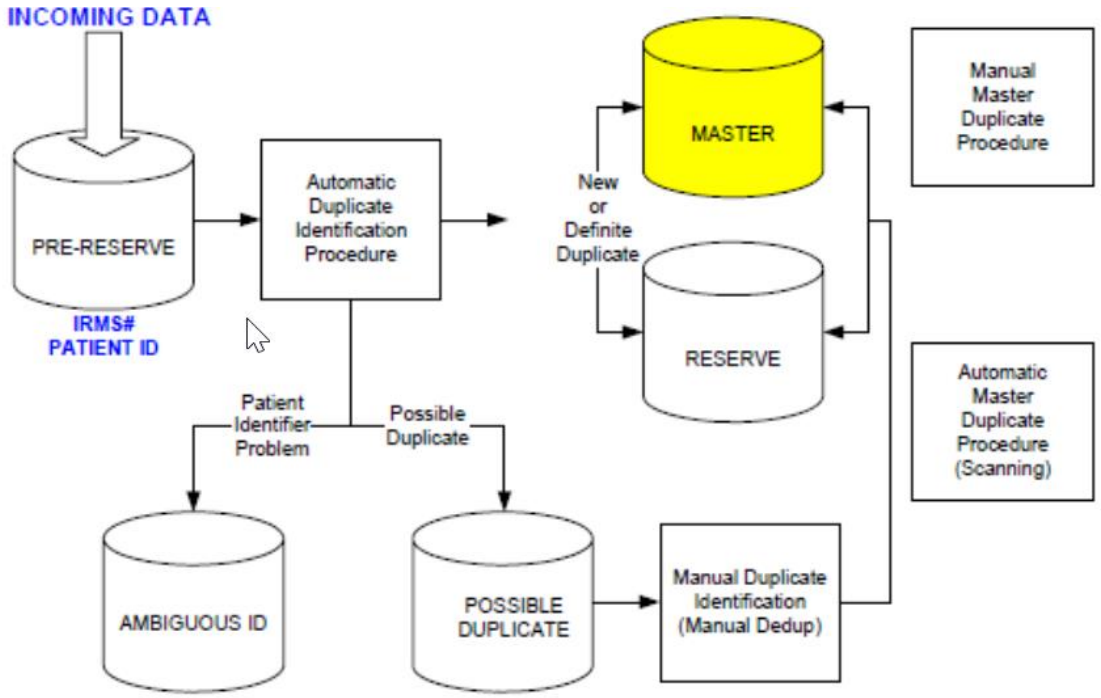


- *Newborn Name Matcher* - enter a placeholder name that sender uses for newborns who have not been named (i.e. Baby Girl, Baby Boy). If the value here matches the first name in a message, the first name is changed to *Newborn*, allowing for easy identification of newborn records.
- *Update Current Patients* - select to always update patient demographics. When unselected, the interface checks to determine if a patient with the same MRN has already been submitted by the

provider. If yes, the patient update is skipped and vaccinations included with the message are added to the record.

- Note - clear this option when old information is uploaded and the current demographics are already in IWeb for some or all patients. This keeps the current patient records from being overwritten.

E. Database Tables and Deduplication Processes for Record Storage



The following are descriptions of the deduplication processes and database tables that a record may encounter[3]:

Types of procedures to determine if patient record is a duplicate or not:

- *Automatic Duplicate Identification Procedure* - implements a rule based algorithm that looks at specific data in the incoming record from the pre-reserve table (IRMS#+ Patient ID = patient identifier), and attempts to locate a match or a similar record
- *Manual Duplicate Identification Procedure*- Manual deduplication process where a user compares the incoming record and the database record to make a decision.
- *Automatic Master Duplicate Procedure (scanning)* - a nightly automatic process that reviews all of the records in the patient master table and merges the similar records.

[3] *Note for Montana Users: For patient records coming through HL7 messages and DTT, the following changes were made in the May 2018 IWeb release:*

Consent Undetermined: If a record is inserted where the consent is undetermined, patient Active/Inactive status (PAIS) defaults to active. If a user (not including users from non-owning facilities) updates the consent status from Undetermined to Consented and the patient status is Inactive at the time, the status is automatically changed to Active (assuming ownership). The user can manually change the status back to Inactive if needed.

Consent Denied: Patients records where consent was denied have a patient status of Inactive upon insertion or saving. If a user (not including users from non-owning facilities) updates the status from Denied to Consented, patient status is automatically changed to Active (assuming ownership). The user can manually change the status back to Inactive if needed.

- *Manual Master Duplicate Procedure* - a manual review of the records in the patient master table, where user manually merges similar records

Database Tables

Pre Reserve	Holds patient records waiting to be processed into the *Reserve* & *Master* tables. When the data is loaded into the *Pre-Reserve* table, some of the fields are pre-processed. Both the before and after process records are stored in the table.*
Reserve	Stores the patient record exactly as it was sent from the Facility or Organization. The Facility/Organization's database's IRMS number and patient ID are stored in this table so that the record can be mapped back to the single patient master record.
Master	Stores the latest patient record sent from the owning Facility, IRMS database. Each reserve record that has been mapped to this Master record can reference it. The mapping of patient reserve records to patient Master records is used to create a composite vaccination record by combining all the vaccinations sent from the Facility/IRMS database.
Possible Duplicate	Holds incoming patient records that have possible duplicate matches in the *Reserve* table. Records only go into this table if the Automatic Duplicate Identification Procedure could not definitely identify a duplicate record. Records in this table require human review using the Manual Duplication Identification or Manual Duplication process
Ambiguous ID	Holds incoming patient records that have patient identifier problems. Records are moved from the Pre-Reserve tables to the Master table one record at a time. The journey of a record to the Master table depends on whether another patient record is found to be similar or a definite match. **

* Addresses are pre-processed to ensure conformity to postal standards and spaces are removed.

** *Ambiguous ID* example: an IRMS sent patient 5 twice - referred to as patient 5a and patient 5b. Patient 5a has a different name than patient 5b. Records in this table require human review using the Ambiguous ID application.

Names are pre-processed in three ways:

- Unnecessary spaces and special characters (such as hyphens) are removed
- Names are checked to see if FIRST (name) and MIDDLE (name) have been swapped, or if MIDDLE and LAST have been swapped.
- Names are checked to determine if FIRST and MIDDLE names were both entered into the FIRST name field.

The record match will result in one of the following outcomes:

- *New Record* - a match was NOT located.
- *Duplicate Record* - a match was located. For example, if the street address/state/zip code or street address/city matches the record in the Reserve table, the end result is a duplicate.
- *Possible Duplicate* - unsure if a match was located.
- *Patient Identifier Problems* - unique Patient Identifier is no longer unique. A separate module, Ambiguous ID, is used by the registry administrator to review and determine whether the record represents the same person or not. For version 4.2, if two records come in from the same IRMS, the newest record always overwrites the existing record, except in cases where the newest record has a null value (keeping what is in the original record rather than keeping the blank field).
    - The existing property *Enable Separate During Ambiguous ID* was disabled.
    - The logic was modified to process ambiguous ID decisions, so if the administrator has chosen to process the incoming records, the newest record will not overwrite any existing values with null (the newest record currently overwrites all existing values).
    - This new logic executes both when the user reviews the ambiguous ID records manually and selects to accept the new record, and when the user selects Process IRMS to process all records without looking at them individually.
    - If the incoming record is intentionally nulling a column, such as an instance when the SSN was incorrect and the correct one is not known, this logic prevents that from being corrected and the user must manually update it in IWeb.

F. Deduplication Rules, Logic & Calculations

The SIIS deduplication algorithm compares an incoming patient record to records in the patient reserve table that have similar data elements. To quickly perform the match, each incoming record is compared against the patient reserve using the match codes/match numbers model detailed below to create a subset of records for comparison. The incoming record is then evaluated with each record in the subset at a more finite level using groups of data elements called match sets.

Each match set is given a score based on the likelihood that it does not contain data elements that would be shared by more than one person. For example, a match on First Name, Last Name, Birthdate and Medicaid Number would belong to only one person and would be considered an exact match. A match on First Name, Last Name, Birthdate and city could belong to more than one person and would be a possible match. Key elements such as First Name, Last Name, Birthdate and another item such as Address or Guardian are most helpful in achieving an exact match. An exact match is given a score of 2 and is automatically merged. A score of 1 indicates a possible match and the record is sent to manual review. A score of 0 indicates no match and the incoming record is inserted as a new patient. Match sets used in the deduplication process are detailed in the following document.

Match Codes/Match Numbers

Match codes and match numbers are generated when a patient record is inserted into the database. They are stored in the h33_patient_reserve and h33_patient_mater tables. These codes/numbers are used to return a subset of records that might be possible duplicates to an incoming record being processed through deduplication.

Match_Code5	Address Street, Address Short City Example: 100 Main, Phoenix, AZ Match_Code5 = 100MAIN PHOENIX
Match_Num1	Patient birth date Year, Month, Day and First Initial as an integer Example: Patient first name (Bob), Patient Birth Date 10/01/1980 Match_Num1 = 1980100101
Match_Num2	Patient birth date Year, Month, Day and Last Initial as an integer Example: Patient last name (Jones), Patient birth date 10/01/1980 Match_Num2 = 1980100110

Additional criteria used to establish the subset of possible duplicates include:

- Social Security Number
- Phone
- Guardian Social Security Number
- Address Street Mangled (Composite address including street, city, state and zip modified for standardization)

Distinguishing Match Sets

Distinguishing match sets are a result of in-depth comparison of an incoming patient record with the records that have been selected based on the match codes/numbers described above. An incoming patient record is compared to records in the patient reserve table. Each match set receives a composite score based on the likelihood that it is a match to the record.

Score = 2	Most likely a match, automatic merge
Score = 1	Possible match, move to manual deduplication
Score = 0	No match

Field level comparisons

First Name	Social Security Number	Guardian SSN
Middle Name	Medicaid Number	Guardian 2 First Name
Last Name	Birth File Number	Guardian 2 Last Name
Birth Date	Birth Order	Guardian 2 SSN
Gender	Multiple Birth Count	Alias First Name
Address	Mother's Maiden Name	Alias Last Name
City	Guardian First Name	IRMS System ID
Phone	Guardian Last Name	IRMS Patient ID

In addition to exact field matches, comparisons are also made of mother's maiden name to patient last name, guardian last name and a comparison of patient last name to guardian last name. Approximate birth date is also compared to eliminate duplicates that might be caused by typographical errors on the birth date. A field level comparison that is determined to be a match receives a score of 1. This is not a "weighted" score, simply an indication that there is a match. For example, Medicaid Number = Medicaid Number would be scored as 1.

Records with first names designated as "BABY", "BABYBOY" and "BABYGIRL" are also compared to other first name and last name fields on matching records and designated as a "SPECIAL_BABY".

Definite Match Sets

Definite match set statements that are 'true' are assigned a composite score of 2 and are automatically merged. For example, in Set 100 below, if both Medicaid Number and Birth File Number fields were matched on both patient records, then each field type would have been scored as 1. Therefore, 1 + 1 = 2, the statement is true, and the composite score would be a 2, resulting in an automatic merge.

Birth order and multiple birth count are used to negate a match set if these items are populated on both the incoming and the existing records but do not match.

G. Deterministic (Default) Rules/ Logic

Deterministic logic uses the following rule-based algorithms to determine matches:

1. Likely Matches
2. Matches (auto merge)
3. Possible Matches
4. Separate Rules/Special Cases

Patient Subset for Matching to Incoming Records

The first step of deduplication is to obtain a subset of likely matches. These are existing patients in the database that match on any of the following field values:

- Birthdate and first initial
- Birthdate and last initial
- Street address, state and zip code
- Street address and city
- 7-digit phone number (excluding area code)
- Guardian social security number
- Social Security number

The next step is to compare each patient in that subset to the 'new' patient.

- If a match is found at any point, the program immediately exits and goes to the next patient
- If a match is not found, the records are flagged as a possible match or as separate patients

The following values are pre-processes:

- Addresses are pre-processed to conform to postal standards and to remove spaces
- Names are pre-processed to remove spaces and special characters (such as dashes in hyphenated names)
- Names are checked to see if First and Middle names have been swapped, or if Middle and Last names have been swapped
- Names are checked to see if First and Middle names were both entered into the First name field.

The deduplication logic examines the following fields:

Medicaid Number	Guardian SSN
Birth File Number	Guardian First Name
SSN	Guardian Last Name
First Name	Mothers Maiden Name
Middle Name	Sex
Last Name	Second Guardian First Name
Birth Date	Second Guardian Last Name
Street Address, State & Zip Code	Second Guardian SSN
Street Address & City	Birth Order
7 – Digit Phone Number (excluding area code)	Multiple Birth Count

Additional information is determined as follows:

If Field	Quantifier	Field/Value	Then Field	Result
First Name	EQUALS	BABY; BABY BOY; Babyboy; BABY GIRL; Babygirl; Babynb; BB; BB1; BB2; Boy; BOY; FC1; FC2; FC3; Female; GIRL; Male; NEWBORN; Newborn; Newborn Boy; Newborn Girl; MC1; MC2; MC3; TwinA; TwinA; Twin B; TwinB	First Name	Is identified as a special baby name match
Street Address, State & Zip Code	EQUALS	Street Address City	Address	Is identified as an address match
Guardian Last Name or Mother's Maiden Name	EQUALS	Patient's Last Name	Last Name	Additional weight is given to unique last names

Patient 1's Last Name	EQUALS	Patient 2's last name; Patient 2's Guardian's Last Name; Patient 2's Mother's Maiden Name	Last Name	Unique Last Names is incremented by 1
Patient 1's Guardian's Last Name	EQUALS	Patient 2's Last Name; Patient 2's Guardian's Last Name; Patient 2's Mother's Maiden Name	Last Name	Unique Last Names is incremented by 1
Patient 1's Mother's Maiden Name	EQUALS	Patient 2's Last Name; Patient 2's Mother's Maiden Name; Patient 2's Guardian's Last Name	Last Name	Unique Last Name is Incremented by 1
Patient 1's Last Name	EQUALS	Patient 1's Guardian's Last Name; Patient 1's Mother's Maiden Name	Last Name	Unique Last Name is incremented by 1
Patient 1's Guardian's Last name	EQUALS	Patient 1's Mother's Maiden Name	Last Name	Unique Last Name is decremented by 1

Note: Since the unique last name is calculated in this manner, its value can by anywhere from 0 to 3

Note: Approximate birth date signifies the birthdate is close. It is calculated as follows:

- If the birthdates are less than nine months apart, birth date is approximate if any of the following are true:
  - The day and month are swapped (i.e. 10/11/2003 and 11/10/2003)
  - The year and day are the same, but the months are different
  - The year and month are the same, but the days are different

- If the birth dates are less than ten years apart birth date is approximate if the day and month match;
- First Guardian Name and Second Guardian Name are examined to check for swaps (i.e. if the father was the First Guardian on one record and the Second Guardian on the other record);
- If both Birth Order and Multiple Birth Count are present and different, and the record would have otherwise matched, the record is sent to Manual Deduplication;
- If Mother's Maiden Name is 'ADULT' or 'SELF,' these values are ignored.

Patient deduplication logic uses a rule based algorithm to identify a match. These merge (match) rules are identified by a rule number. The rules, descriptions and fields for a matched record are listed below, where each Match Set Score = 2, and therefore will automatically merge. Once all of the exact matches have been identified, the remainders of the patient records are run through the rules for *Possible Matches.*

Rule #	Description	Fields: Column 1	Fields: Column 2	Match Set Score
100	Medicaid Number + Birth File Number	N/A	N/A	2
110	SSN + Birth File Number	N/A	N/A	2
120	First Name and at least one of the fields from columns 1 & 2	SSN; Birth File Number; Medicaid Number;	Birth Date; Middle Name; Unique Last Name; Guardian First Name; Address; Phone	2
121	Nickname + Sex + At least one of the fields from columns 1 & 2	SSN; Birth File Number; Medicaid Number	Birth Date; Middle Name; Unique Last Name; Guardian First Name; Address; Phone	2
122	First Name + approximate Birth Date + at least one of the fields from columns 1	SSN; Birth File Number; Medicaid Number		2
125	Passport Number + Birth Date	N/A	N/A	2
130	Special Baby Name + Birth Date + at least two of the fields from column 1	Unique Last Name; Address; Phone; Guardian First Name; Medicaid Number; Birth File Number; SSN; Guardian SSN;		2
160	Birth Date + First Name + at least one of the fields from column 1	Address; Phone; Guardian SSN		2
161	Birth Date + Nickname + Sex + at least one of the fields from column 1	Address; Phone; Guardian SSN		2
190	Birth Date + First Name + at least two fields from column 1	Guardian First Name; Unique Last Name		2
191	Birth Date + Nickname + Sex + at least one of the fields from column 1	; Guardian First Name; Unique Last Name		2
192	Birth Date + Alias First Name + Sex + at least one of the fields from column 1	Guardian First Name; Unique Last Name		

200	First Name + Middle Initial + Last Name + approximate Birth Date + at least one field from column 1	Address; Phone; Guardian SSN		2
201	First Name + Middle Initial + Last Name + approximate Birth Date + at least one field from column 1	Mother's Maiden Name; Guardian First Name;		2
202	First Name + Last Name + approximate Birth Date + at least one field from column 1 & 2	Address; Phone; Guardian SSN	Mother's Maiden Name; Guardian First Name;	2
203	First Name + Middle Name + approximate Birth Date + at least one field from column 1 & 2	Address; Phone; Guardian SSN	Guardian First Name; Unique Last Name	2
204	First Name + Last Name + approximate Birth Date + at least two of the fields from column 1	Address; Phone; Guardian SSN		2
205	First Name + Last Name + approximate Birth Date + Unique Last Name has at least two matches	Guardian First Name		2

The Possible Match rules are not examined until all of the Match (merge) rules for definite matches have been examined. (Note that these rules do not have rule numbers.)

Understanding the Match Set Score: Possible Match negation sets lower a possible match composite score if there are certain fields that are definitely not matches and prevents the records from going into manual deduplication.

NOTE: States can optionally define a list of Organizations whose patients should never be sent to manual deduplication. If the patient is a possible match with any patient that is currently in the registry, the incoming patient record is deleted rather than sent to manual deduplication. The list of Organizations must be provided and sent to STC's Help Desk to be set in the database.

Description	Fields: Column 1	Fields: Column 2	Match Set Score
First Name, Middle Initial, Last Name, Birth Date	N/A	N/A	1
First Name, Middle Name, Last Initial, Birth Date	N/A	N/A	1
SSN	N/A	N/A	1
Medicaid Number	N/A	N/A	1
Birth File Number	N/A	N/A	1
At least one field from columns 1 & 2	First Name; Birth Date	Address; Phone; Guardian SSN	1
First Names are like sounding + approximate Birth Date matches + at least one field from column 1	Address; Phone; Guardian SSN		1
Birth Date + at least two fields from column 1	First Name; Middle Name; Guardian First Name; Unique Last Name		1
Birth Date + First Names are like sounding, & Last Names are like sounding	N/A	N/A	1

After the *possible matches* are determined, additional rules that will negate the match score are applied to remove matches based on other information that would make them an unlikely match.

If the rules for *possible matches* have flagged too many records, the following additional, separate rules apply for special cases. There are no rule numbers for these.

Description	Fields: Column 1	Fields: Column 2	Match Set Score
Birth Dates are very different: Day, Month, and Year do not match and Birth Dates are more than ten years apart.	N/A	N/A	Deduct 1
Twins (Birth Dates are the same and Family/Address information is the same). If First Names do not sound alike and First/Middle/Last Names were not swapped, the records are "twins" if any of the fields in column 1 are true.	Sexes are opposite; SSNs are not null and not equal; Provider's ID Number is sequential; First Initials do not match & First Names, excluding First Initials, do not sound alike; both the First and Middle Names are populated; the first three characters of both Names are difference		Deduct 1
Birth Date does not match and it is not an Approximate Birth Date, and the First Name does not match and the First Name does not match as a Nickname	N/A	N/A	Deduct 1
Middle Initials are present and different in both names and there are no swapped initials	N/A	N/A	Deduct 1
Birth Date does not match and it is not an Approximate Birth Date, and the First Initial is different and the Last Name does not match	N/A	N/A	Deduct 1
Birth Date and Phone Number match but SSN, Address, and Last Initial do not match, and neither First Name or Nickname match, and First Names do not sound similar (this handles the special case of a Phone Number being reissued)	N/A	N/A	Deduct 1
To handle twins with the same First Name, if the records otherwise match and the Middle Initial is different, or the Middle Name is populated and sounds different, the record is sent to manual review	N/A	N/A	Deduct 1
If First Name, Middle Name, and Birth Date match, but none of these match: Last Name, SSN, Mother Maiden Name, Guardian First Name, Address, the records are not a match	N/A	N/A	Deduct 1
If Phone Number and First Name match, but	N/A	N/A	Deduct 1

Birth Date, Approximate Birth Date, SSN, Address, Last Initial, Guardian First Name, and Middle Initial do not, the records are not a match (it otherwise would have been sent to manual). This addresses a corner-case of a Phone Number being reissued to someone with the same First Name			
If Address and First Name match, but Birth Date, Approximate Birth Date, SSN, Address, Last Initial, Guardian First Name, and Middle Initial do not, the records are not a match (it otherwise would have been sent to manual). This addresses a corner-case of a person with the same First Name moving to an address that someone else had previously	N/A	N/A	Deduct 1
If a record is a possible match and 3+ vaccinations match, the record is automatically processed as an exact match			Deduct 1

H. Administrative Deduplication Functions

H.1 Manual Deduplication

Manual deduplication allows an administrator to manually view duplicate patient records and make a decision on whether to merge the two (or more) records into one, or leave them as separate if they are not similar enough. The administrator reviews both the incoming patient record and the existing record from the Reserve table, determining whether to create a new patient record or merge the patient records.

> Note: For master and user-flagged deduplication, the patient owner is determined by the selected patient if there are no vaccinations. When you can select which one of the records you want as the master record, regardless of update stamps, the ownership will be the organization of the record manually selected. If there are vaccinations and the Patient Master Update on the Vaccination registry option is enabled, the owner changes to the Organization (IRMS) with the most recent administered immunization. The patient demographics information on the master record is updated to the owning provider's selected patient.

When records are imported and the record is a possible match with a queued duplicate record, the new record is staged and not processed until the previous possible match is processed through Manual Deduplication. This can occur when records from a high-priority Organization, such as Death Records, are staged rather than processed. By enabling the logic, the state can prioritize a particular Organization's records in order to process them first.

H.2 Ambiguous ID

This function is used to evaluate records when unique patient IDs from incoming Organizations no longer appear to be unique. Medical Record Numbers submitted by an Organization should be unique to a patient to ensure vaccinations are properly linked to the patient. The administrator can decide if the record actually does represent the same person or if the submitting site has a corrupt ID system. These records are stored in the Patient Ambiguous ID table.

> Note: This is a manual deduplication operation. As a result, it locks deduplication while it processes (*Incoming is Different*, *Bypass* button clicked). Due to the locking of records, only one person should process ambiguous IDs at a given time. Additionally, if there are other deduplication operations running, ambiguous IDs cannot be processed.

The Ambiguous ID process is described here, with a visual representation below.

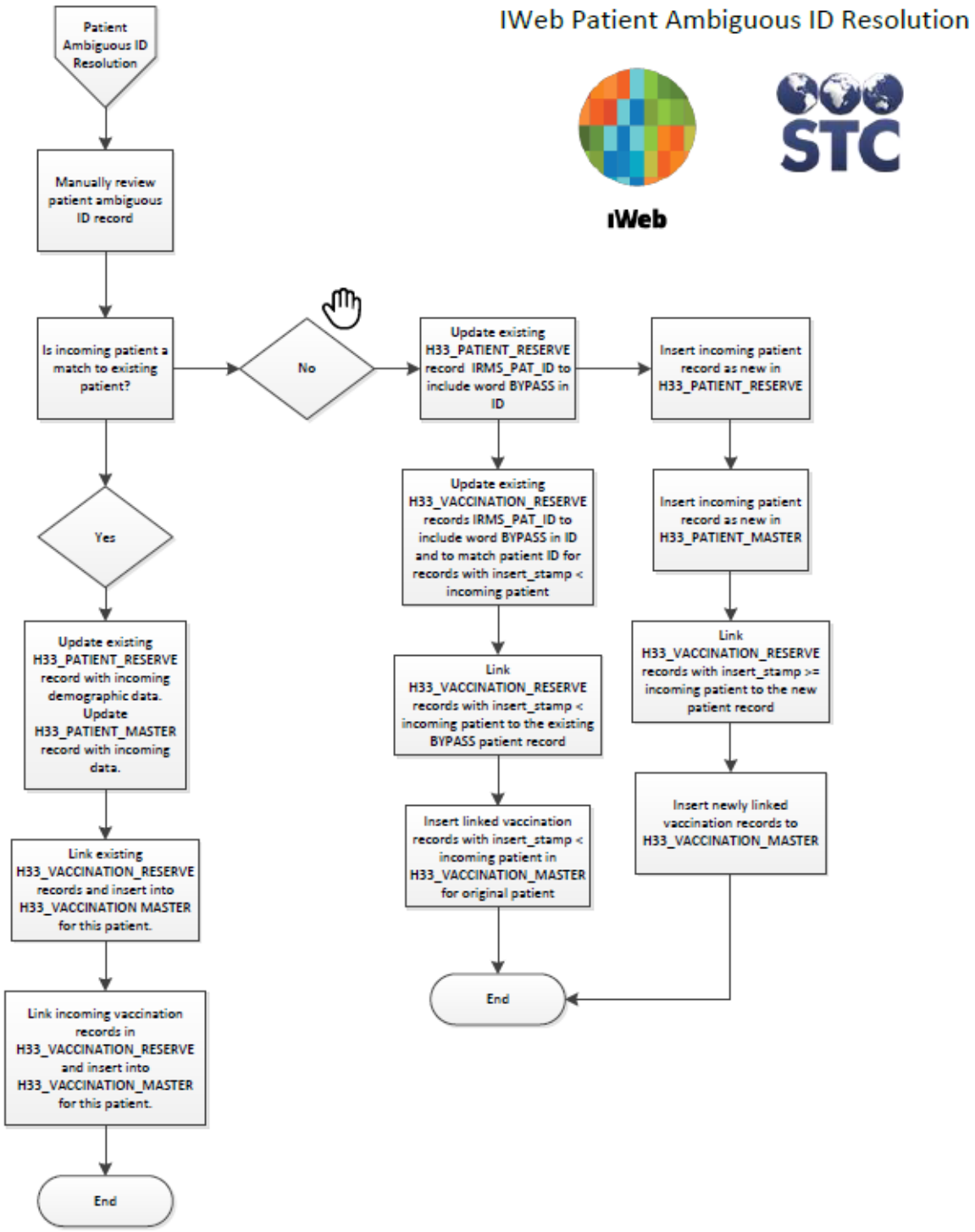
1. The incoming record is staged into the h33_patient_pre_reserve database table.
2. The deduplication logic processes the incoming record.
3. The first deduplication step is to look for an existing record with the same irmsSysId/irmsPatId. If found, the deduplication runs for that pair (the existing record and the incoming record with the same irmsSysId/irmsPatId). If either a definite or a possible match is located, the record merges the new data and deduplication is finished for that patient. If the record pair are determined to not be a match, the patient is staged to Ambiguous ID. Immunizations that were associated with the existing patient record are unlinked from the Master patient record until the Ambiguous Id is resolved to ensure they are associated to the correct patient.

4. If the user reviews the records individually (i.e., the user clicks the *Examines Records* button), and
    - The *Incoming is Different, Bypass* button is clicked, then the existing irmsPatId is changed (this is the ###.BYPASS.### that can be seen at times), the incoming record is placed back into the h33_patient_pre_reserve table for reprocessing, and when deduplication runs, it does not find the matching patient. The incoming record is processed through deduplication. Immunizations are linked to the correct patient record based on the time and date they entered the IIS.
    - The *Incoming is Same, Update* button is clicked, and then it depends on the scenario.

        i. Scenario 1: Only one reserve record is the matching irmsSysId/irmsPatId: The existing patient is overwritten with new patient data; the old patient name, etc., is completely gone (except in the history table); and this is treated as an update. For example, the deduplication logic determined that the incoming irmsSysId/irmsPatId record matched the existing one exactly except the health plan number was different, so the old health plan number is overwritten and only exists in the history.

        ii. Scenario 2: Patient has multiple reserve records: The matching record is separated (a registry-level property enabled/disabled with a script called Separate Patients in Ambiguous ID) first, if enabled, when the button is clicked. If disabled, the matching record is not separated and the associated reserve record is updated. The master updates (or not) based on the normal rules (ownership, historical, etc.) The script can be obtained by logging a help desk ticket. The existing patient data for the separated record is overwritten with the new patient data, and all remaining reserves stay with the original record.

Ambiguous ID Resolution Workflow



H.3 Master Deduplication by External Source

Master Deduplication by External Source allows an administrator to view duplicate patient records and make a decision on whether to merge the two records into one or leave them as separate if they are not similar enough. The goal is to group records belonging to the same patient so a complete picture of the patient's vaccination history can be viewed.

Note: When using this option, the Master Patient IDs are required.

H.4 Merge History Report

The *Merge History Report* lists automatic and manual merges that were performed in IWeb. The types of merges included are:

- User-Reported Duplicates
- Master merges (completed via Master Deduplication by External Source)
- Master Duplicate Scan (a nightly run job that scans for duplicates)
- Third Record Source (duplicates that are found based on new incoming records)
- Manual merges (merges that are reviewed on the Manual Deduplication page)

The report output includes the report parameters used to generate the report, as well as SIIS Patient IRMS ID and IRMS Patient Detail ID.

The report output omits one of the patient's SIIS Patient IDs when the merge is a Manual merge. The reason for this is because when a patient is queued for Manual Review, the SIIS Patient ID is not created until after the patient is reviewed.

The report is based on three tables: H33_MERGE_AUDIT_HISTORY (contains data for user-reported duplicate merges), H33_MERGE_LIST (contains data for Master merges, Master Duplicate Scan, and Third Record), and H33_MANUAL_DEDUP_DECISION_LOG.

H.5 User Reported Duplicates

When administrators define values under *Define Deduplication Reasons*, users will be able to see a *Report Duplicates* button on the Patient search page when there are two or more search results returned.

After selecting the *Report Duplicates* button, user will have the following workflow:

- user selects possible matches,
- user enters in comment and reasons why they believe there are duplicates,
- user specifies recommended master,
- patients then do NOT merge, but are staged for administrator review and are accessible from the *Manual Deduplication* page

Define Deduplication Reasons is used to define the reasons that a user is reporting a duplicate record. Defining reasons will populate when a user reports duplicates after searching patients through *Patient Search/Add*. Some examples include:

- First Name spelling
- Last Name spelling
- First Name hyphenation
- Last Name hyphenation

H.6 Separate Bad Merges

Separate bad merges is used to separate patient records that were erroneously merged. The Master patient ID is entered to view the reserve record mapping and it will show all the records that were merged together. You can separate out which record that should not have been merged. An indicator that a bad merge has occurred is when the record contains vaccinations that are known to be false.

H.7 Run Deduplication

Run Deduplication is used to start the deduplication process, including concurrent multiple runs. The scheduler utilizes all five threads for deduplication. If deduplication is running for all Organizations only one deduplication process can run at a time. If deduplication is running for one specific Organization, multiple deduplication processes can run, but only for one Organization each. The progress tables add the Organization and the audit tables log the progress tables.

The deduplication logic checks for deduplication processes for the specific Organization. If there is no Organization specified, no other deduplication processes are allowed to start.

H.8 Schedule Recurring Deduplication

Schedule recurring deduplication allows an administrator to set up a schedule to automatically run the deduplication process. Select the *Schedule recurring deduplication* link to access the schedule parameters. User should define the following, and then save:

- frequency
- hour
- minute

H.9 Stop Scheduled Deduplication

Stop scheduled deduplication allows an administrator to stop a scheduled deduplication process. Select the *Stop scheduled deduplication* link, and if there is a procedure currently running, the page will open with the header *Currently Running Procedure*. Click *Refresh* to refresh the page if necessary. Click *Run Stopall* to immediately terminate any database procedures. Click the *Back* button to return to the Administration Main Menu page. It is important that the Master Duplicate Scanner not be executed during regular business hours as it will cause a system slow down.

H.10 Auto Deduplication

Auto Deduplication is the ability to process one or all current records in manual deduplication in order to quickly auto merge. The *auto deduplication* logic loops through the records flagged for manual deduplication and automatically clicks the *merge* or *skip* button on the screen and proceeds to the next record. The business logic behind when to merge an incoming record or skip is as follows:

> The incoming patient record is compared against the matched record. Each of the matched fields (first name, last name, DOB, gender, race, address, guardian name) are given a match index based on the match. When a match index is 5 or greater the incoming record is automatically merged.

Similar to manual deduplication, users can run auto deduplication by Organization or all records. The auto deduplication screen and filters exactly match the manual deduplication screen.

Note: *auto deduplication* is only run for the batch imports. Auto deduplication does not process any user flagged records.