

Chapter 10: Matching Records to Existing Data

This chapter describes the processes to match an incoming record to other data in the database. Data for the same person are consolidated into a single Patient Set. CTC-level matching algorithms are also used in SEER*DMS on a limited basis (expansion of tumor-level match and consolidation rules are scheduled for the future).

In this chapter, you will learn about

- Patient-Level Matching:
 - Overview of Patient-Level Matching
 - Completing a Manual Match Task
 - Match Review Task
 - Automatic Match Task for Patient-Level Matching
- CTC-Level Matching
- SEER*DMS Workflow Diagram - Matching

Overview of Patient-Level Matching

Each source record is matched against the database as the record moves through the workflow. The automatic match task applies registry-specific algorithms to compare patient information on the record to data in the database (to view your registry's algorithms, go to **Help > Matching** in SEER*DMS). The record is matched against consolidated data (Patient Sets) and may also be matched against other source records that are unlinked. SEER*DMS uses probabilistic matching algorithms that are customized for each registry. If the match score is at or above the 1000 point threshold, then it is considered a perfect match. In most registries, the record would be auto-linked to a Patient Set that is a perfect match, and an attempt would be made to auto-consolidate the data.

A manual match task is created to resolve problems identified by the automated match task. For example, a manual Match task is created if the incoming record is a perfect match to two or more Patient Sets. Manual Match tasks may also be created to review probable matches that do not meet the threshold for auto-linkage. The Match Patient Level (Post) workflow routing script determines when a manual task is created. The script can be edited by IMS staff without requiring a release of the SEER*DMS software.

If the record did not match a Patient Set, then SEER*DMS will attempt to build a Patient Set from the record (along with any matched unlinked records), assuming it is an appropriate record type. Typically, a Patient Set is built immediately from a reportable abstract that does not match a Patient Set. Registry rules determine when other record types are used to build Patient Sets (e.g, pathology reports, death certificates).

A record that matches a Patient Set is auto-linked at the patient level. First, the registry's auto-consolidation rules are applied for the demographic (or patient-level) fields. Then, the registry's CTC-level rules for auto-matching, auto-linkage and auto-consolidation are applied. Each rule includes logic that defines when a discrepancy requires a manual review.

Completing a Manual Match Task

When you open a Match Task, the auto-match algorithms are re-executed. This guarantees that recently loaded or modified data are included in the match results. The screen shot below shows

the screen used for a manual Match task. This is a partial screen shot; some of the data columns are not shown.

SEER*DMS
Match (TSK-4564370)

Facility FAC-387085: Death Certificate Only ⓘ City TBD State GA Tumor Registry No

Score	Type	ID	Last	First	Middle	Suf	DOB	SSN	Sex	Race	DOLC	VS	Site	Morph
Incoming Record														
	DC [N]	REC-100830959 ⓘ	SQUAREPANTS RAY	SPONGEBOB			02-14-1923	123-45-6789	2	88	01-10-2014	4		
Matches (2 selected)														
<input checked="" type="checkbox"/>	☆ 1870 Pat Set	PAT-10193502 ⓘ	SQUAREPANTS	SPONGEBOB			06-18-1951	123-45-6789	1	01	11-29-2015	4	C809	8140/3
<input checked="" type="checkbox"/>	☆ 1870 HL7 [N]	REC-100812492 ⓘ	SQUAREPANTS	SPONGEBOB			08-10-1952	123-45-6789	1	99	01-16-2014	1		
<input type="checkbox"/>	990 Pat Set	PAT-10106171 ⓘ	THESNAIL	GARY	S		01-11-1938	123-45-6789	2	02	06-03-2015	1	C180	8140/3

Matched in 1s. [View Matching Algorithm](#)

Complete Complete & Follow Send for Review Cancel

To complete a Match task: (the match permission is required)

- When the Match task is opened, the automatic match is refreshed so that the match results are current.
- In this task, you will add a check mark to any row that has data for the same patient as the incoming record.
- Matches with a score of 1000 or higher are checked automatically. Review matches that are checked. You can de-select any match regardless of the score. Use the **arrow keys** to move up and down the rows in the table.
- Press the **space bar** or use your **mouse** to add or remove check marks from the rows.
- Verify that all true matches are checked.
- Use the **right arrow** or **click complete** to complete the task. The record will move forward in the workflow. The record will either be auto-consolidated with other matching data; or it will be used to create a Patient Set; or it will exit the workflow. The outcome will vary based on record type and registry rules.
- Other options for completing the Match task:
 - Complete & Follow** – Not recommended. This button is only shown if you have the *match_follow* system permission. Click Complete & Follow to follow the record through the workflow. *This option is not recommended because you will need to wait for the incoming record and matching records to complete all automated tasks. The record may exit the workflow, in which case either the Patient Set editor (if record was auto-linked) or the record editor (if record was not auto-linked) will open.*
 - Send for Review** – Some match tasks are difficult to complete because of missing data items or possible duplicates. Many registries designate a single person to complete difficult matches (for example, a person who has access to external data sources). Click **Send for Review** if you cannot complete a match task. The task will be re-assigned to the registry's designee for difficult match tasks. Note to registry managers: this is controlled by the assignment script for the Match task. The script checks the `task.isMatchReassign` function. That function will return true if a user clicked Send for Review. Please contact IMS if you need assistance with workflow scripts.
 - Cancel** – do not complete the current task and return to the worklist.

The following describes other options and information displayed on the Match task screen:



- **Task Bar** – The Task Bar (pictured above) is shown under the SEER*DMS header bar.
 - Task ID – shown on the left side of the task bar: Match (TSK-5144548).
 - Flags – worklist flags assigned to the task will be shown in the Task Bar. In the screen shot above, there is a “New Patient Set” flag assigned to the task. You can click the flag section of the task bar to toggle the flags. If flags are shown, hover over a flag to view a description of the flag.
 - Task Menu – each item in this menu is controlled by system permissions. You will only see a menu item if you have the permission. The Task menu items are:
 - Skip – exit current task without making changes and go to the next Match task. The skipped task will still be assigned to your account.
 - Worklist Flags – this is the same as clicking the flag section next to the Task menu. You may toggle flags on or off for the task.
 - Reassign – assign the task to a user that you select from a list. (Permission *worklist_action_reassign_mine* or *worklist_action_reassign_all*)
 - Release – set the task to unassigned. (Permission *worklist_action_release_mine* or *worklist_action_release_all*)
 - Restart – send the incoming record to an earlier step in the worklist. (Permission *worklist_task_restart*)
 - Terminate – terminate the match task. The incoming record will exit the workflow. It will remain in the database as an unlinked record. (Permission *worklist_task_terminate_manual*).
 - Other menus – the other menus (View, Manage, etc) shown in the task bar are not directly related to the Match task. Different menus will be shown depending on your system permissions.
- **Facility Information** – The Incoming Record’s facility is shown in two places on the Match screen. There is a column with the Facility ID in the row labeled Incoming Record. There is also a line above the Incoming Record row that shows the Facility ID, City, State, and the value of the Tumor Registry flag.
- **Incoming Record** – The Incoming Record row shows data items from the record that triggered the match. This record is moving through the workflow and stopped at this manual Match task.
- **Matches** – The Matches section shows Patient Sets and unlinked records that are likely or possible matches to the incoming record. A match with a score greater or equal to 1000 is considered a perfect match and is automatically selected. A match with a score less than 1000 is a possible match. You may click the match score to view the matching criteria associated with the score. Click “View Matching Algorithm” to view the complete matching criteria associated with all possible scores.



Select **Help > Shortcuts** or press **F1** to view keyboard shortcuts. The Matching section will only be shown when you are in a Match task.

Matching	
Select next row	Down Arrow
Select a row on the next page	Page Down
Select the last row	End
Select previous row	Up Arrow
Select a row on the previous page	Page Up
Select the first row	Home
Complete task and load next	Right Arrow
Complete task and follow	Alt+Right Arrow
Check/Uncheck the selected row	Space
Open/Close the Record Info Popup for the selected Row	I
Open/Close the Facility Info Popup	F
Open/Close the Score Popup for the selected Row	S
Open/Close the Linkage Algorithm Popup	L
Reload page and re-match	F5
Cancel	Alt+C
Skip	Alt+K

Match Review Task

A Match Review task is created if there is an attempt to create a new Patient Set that is a possible duplicate of another Patient Set in the database. For example:

- The record's Match task is completed. It may have been completed in a manual or automated process.
 - Example of a manual task: A user completes a manual Match task for a pending NAACCR abstract. A Patient Set is listed as a possible match, but the user does not select that Patient Set. The record moves forward to step 2 below.
 - Or the NAACCR abstract is matched against the database in the automated task. A matching Patient Set is not found by the automated match task. The routing script does not result in a manual Match task and the record moves forward to step 2 below.
- An attempt is made to create a new Patient Set from the abstract. At this point in the workflow, SEER*DMS executes the duplicate patient matching algorithm. If the new Patient Set is a possible duplicate, then processing stops and a Match Review task is created. The Match Review task is completed in the same manner as a Manual Match task.

Match Review (TSK-4750486)

Facility FAC-670010: Kaiser Permanente Network City State Tumor Registry No

Score	Type	ID	Last	First	Middle	Suf	DOB	SSN	Sex	Race	DOLC	VS	Site
Incoming Record													
NA		REC-100901194	SQUAREPANTS	SPONGEBOB	T		04-11-1944	123-45-6789	1	06	01-13-2016	1	C619
Matches (3 selected)													
<input checked="" type="checkbox"/>	☆ 1910	Pat Set	PAT-10193502	SQUAREPANTS	SPONGEBOB		06-18-1951	123-45-6789	1	01	11-29-2015	4	C809
<i>The original match task triggered a warning for this patient set. Please review and determine if it matches the incoming record.</i>													
<input checked="" type="checkbox"/>	☆ 1910	HL7	REC-100812492	SQUAREPANTS	SPONGEBOB		08-10-1952	123-45-6789	1	99	01-16-2014	1	
<input checked="" type="checkbox"/>	☆ 1870	DC	REC-100830959	SQUAREPANTS	SPONGEBOB		02-14-1923	123-45-6789	2	88	01-10-2014	4	
	990	Pat Set	PAT-10106171	THESNAIL	GARY	S	01-11-1938	123-45-6789	2	02	06-03-2015	1	C180
Matched in < 1s. View Matching Algorithm													
Duplicate Patient Set Warning A request was made to build a new patient set from the records. The new patient set would be a potential duplicate. Please review.													
ID	Last	First	Middle	Suf	Maiden	Sex	DOB	SSN	VS	DOLC	Address		
PAT-10193502	SQUAREPANTS	SPONGEBOB				1	06-18-1951	123-45-6789	4	11-29-2015			

Complete Complete & Follow Cancel

Automatic Match Task for Patient-Level Matching

“Automatic matching” refers to the execution of registry-specific algorithms to compare data items on an incoming record to data stored in the database. The primary objective of the auto-match task is to find a matching Patient Set so the record can be linked to that Patient Set. If no Patient Set exists, the record may be used to create a new Patient Set. All of the patient’s unlinked records will be pulled together into the new Patient Set.

The algorithms used by the auto-match task are registry-specific and are documented on the Matching help page. The names of the algorithms used for the auto-match task all start with the term “Record Linkage” and typically include:

- **Record Linkage: Primary Algorithm for Automatic Match Task.** This algorithm is used by the auto-match task for nearly all record types.
- **Record Linkage: Supplemental Records.** Many registries have a separate algorithm to match supplemental records against the database. This algorithm is often used for passive follow-up records. These records are matched against Patient Sets in the database and are not matched against other unlinked records.

The matching scores are defined in the Matching table. The highest score is 2000. All scores of 1000 and greater are considered “perfect” matches, but some “perfections” are recognized as stronger matches than others.

Workflow routing scripts are executed before and after the auto-match. Routing scripts provide a mechanism to define a workflow that is customized by record type, facility, site, and other attributes.

- **Pre-Matching Routing Script.** In the current version of SEER*DMS, the only purpose of the pre-match script is to send a record through the Pull List workflow before it goes to the auto-match task. This is a specialized workflow developed in 2015. For more information, refer to *Chapter 4: SEER*DMS Workflow*. Contact SEER*DMS Technical Support if your registry is interested in using the Pull List workflow.
- **Post-Matching Routing Script.** The post-matching routing script evaluates the match results and determines whether the record is sent to a manual Match task, auto-linked to a Patient Set (existing or new), or sent out of the workflow and retained in the database as an unlinked record.

CTC-Level Matching

After a record is automatically or manually matched, linked and consolidated at the patient level, the record usually must be matched and linked to one of that Patient Set's existing CTCs. In a manual consolidation task, this is done when the user links the record to a CTC.

Automatic CTC-level matching is implemented using code based on the Multiple Primary and Histology Rules. The code is currently under development but is being used by some registries. The script uses the function `.isSamePrimary`, which uses the following record and CTC properties:

- `primarySite` (#400)
- `histologyIcdO3` (#522)
- `behaviorIcdO3` (#523)
- `histologyIcdO2` (#420)
- `behaviorIcdO2` (#430)
- `laterality` (#410)
- `dateOfDiagnosisYear` (#390)
- `dateOfDiagnosisMonth` (#390)
- `dateOfDiagnosisDay` (#390)
- `rxSummTreatmentStatus` (#1285)

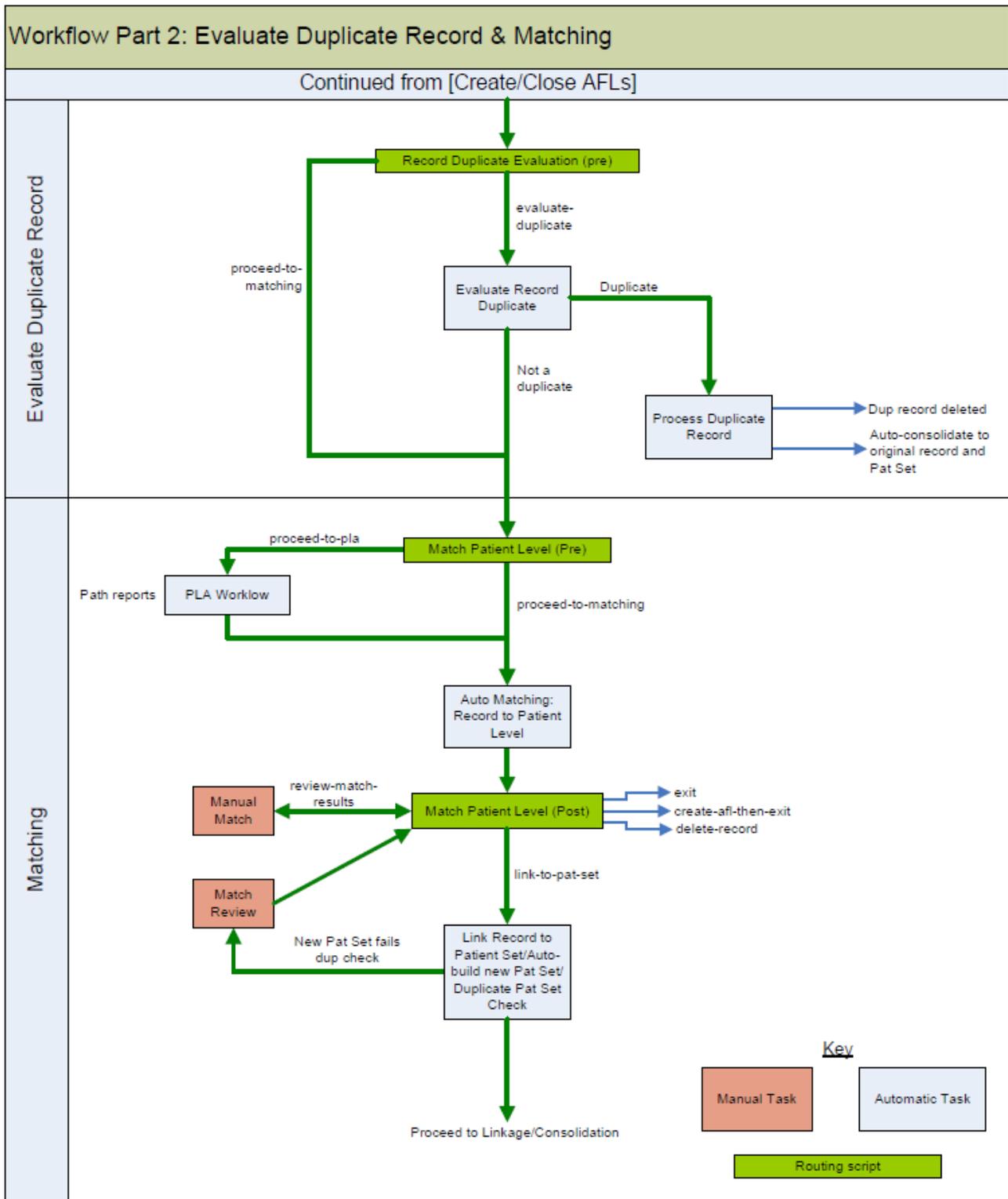
The Multiple Primary Rules can also be called for assigning "C" and "P" flags (System > Workflow > Task Manager > Flags).

If the record does not match any of the existing CTCs (through automatic or manual processes), a new CTC will be built according to your registry's Auto-Build rules.

SEER*DMS Workflow Diagram - Matching

A partial workflow diagram is shown in Figure 10-1 below.

Figure 10-1



The diagram above highlights the Matching section of the workflow. Complete workflow diagrams are provided in Chapter 4. This diagram is generic; that is, it indicates points which use registry-specific logic but registry algorithms are not documented.

The following table describes each process and decision point in Figure 10-1.

Process	Description	Workflows
Record Duplicate Evaluation (pre) Routing Script	This script is used to decide if a given record should be evaluated for record duplicates, or follow the regular workflow path.	
Evaluate Record Duplicate	Compares a record to another record of the same type	All record types
Process Duplicate Record	<p>If the incoming record is an updated version of a record in the database, auto-consolidate the two records. One record is updated and retained; the other record is deleted. If the original record was linked, then the Patient Set is automatically updated.</p> <p>The Duplicate Record matching algorithms are used to identify duplicates.</p>	
Match Patient Level (pre) Routing Script	This script is used to decide if a given record should go through the Pull List Algorithm (PLA), or follow the regular workflow path.	
PLA Workflow	The Pull List Algorithm (PLA) workflow was designed by Seattle for Pathology Reports. Currently, it is only used by the Seattle registry. This workflow can be used if the registry's operations include the import of "reject" records – records that indicate that a case finding request has been "rejected".	
Auto Matching Task	Match the record against data in the database. The matching algorithm determines whether it is matched against all patient sets and unlinked records; limited to patient sets, etc. As an example, in most registries, abstract records would be matched against patient sets that are not marked as deleted and all unlinked records. Supplemental records would only be matched against patient sets.	All record types
Match Patient Level (post) Routing Script	This script runs after automatic and manual matching; it determines if a Manual Matching task needs to be created, or if the matching results/selection should be processed.	

Process	Description	Workflows
Linkage Record to Patient Set/Auto-build/Duplicate Patient Set Check	<p>Automatic task that either:</p> <ul style="list-style-type: none"> • links the record to an existing matching Patient Set • Auto-builds a new patient set from the auto-build rules and checks for duplicate patient sets <p>If a duplicate patient set is found during auto-build, then the record is sent for a Match Review and is sent back to the Match Patient Level (Post) Routing Script.</p>	<p>Abstract workflow, short health record workflow, case finding and pathology report workflow, and death certificate workflow</p> <p>Auto-build rules are documented on each registry's help page.</p>