Chapter 2: Patient Data

The SEER*DMS database houses traditional cancer registry data. These include consolidated patient data and standard source records that contribute to the consolidated data. SEER*DMS also supports the processing and storage of data from novel data sources that may be used to augment standard diagnosis, treatment, and staging information. In addition, SEER*DMS supports the processing and storage of data items obtained from linkages to national organizations.

- **Source records** - The record table houses standard types of data submitted to the registry. Records contain the original data submitted to the registry by hospitals and other facilities. Record data include medical data submitted by reporting facilities and follow-up information received from non-medical organizations. Each record contains a single report of data. A NAACCR record contains data from a single abstract, an HL7 E-path record contains data from a single pathology report, etc.

- **Consolidated data** - A Patient Set contains the data consolidated from a patient's source records and provides links to all data available for the patient (source records, novel data types, and linkage data). A registrar edits the Patient Set fields and, typically, does not change values in the original records. The records serve as a permanent reference of the original values sent to the registry; while Patient Set data are submitted by the registry to SEER, NAACCR, and other organizations.

- **Novel data types** - SEER*DMS supports novel data sources that are linked to other data for a patient but, currently, are not used for tumor level consolidation. These data are used for patient-level consolidation when appropriate. For example, some of the data types are used to update date of last contact. Novel data are not used to update standard data items related to diagnosis, staging, or treatment. Tumor level data from novel data sources are currently stored in non-standard fields defined in SEER*DMS. SEER*DMS currently supports insurance claims data, pharmacy claims, radiation oncology treatment data, genomic test result data, and physician reports (meaningful use data).

- **Linkage data –** Residential address history and other types of data obtained from linkages are linked to the Patient Set data and stored in separate SEER*DMS tables.

In this chapter, you'll learn about
- Records
  - Linked and Unlinked Records
- Patient Sets
  - Cancer/Tumor/Case (CTC)
  - Staging
  - Summary Treatment (Summary TX)
  - Course
  - Treatment Procedures (TX and TXr)
  - Facility Admission
  - Pending and Consolidated Records
  - Auto-building Patient Sets and CTCs
- Novel Data Types
- Linkage Data
- Audit Logs

Records

In the SEER*DMS database, the record table contains the original data submitted to the registry. Records are created in the database when data are loaded from files or entered manually (as described in *Chapter 5: Importing Data Files* and *Chapter 6: Data Entry*).

Each record travels through the workflow, triggering the automated and manual "tasks" that must be performed to process the data. The path of a record through the workflow is determined by registry rules and the record's type. *Chapter 4: SEER*DMS Workflow* contains diagrams showing the workflow path. The following are the record types supported by SEER*DMS; some record types may not be available in your registry's configuration.

Abstract Records:

- **Health Record** – An initial abstract record submitted by a medical facility in a registry-defined file format. This is a deprecated record type. All new registries use NAACCR Abstract for initial abstract records.

- **NAACCR Abstract** – An initial abstract record submitted by a medical facility in a NAACCR file format.

- **NAACCR Modified Record** – An abstract record that was modified and re-submitted to the registry. This record is identical in format to the NAACCR record except that the record type field is coded with an 'M'.

- **Short Health Record** – A SEER*DMS short health record is a partial or incomplete abstract. A partial abstract may include some patient demographic information, complete or partial histopathology information, treatment information, or other information not sufficient to complete a full case abstract. Typically, these records provide additional information regarding follow-up, treatment, and admissions related to a previously abstracted case. In rare situations, a short health record may contain the only data related to a case and may be used to build a new CTC.

Pathology and Imaging Reports:

- **HL7 E-Path** – Electronically transmitted pathology reports transmitted in an HL7 standard file format. These records typically trigger the creation of Abstract Facility Leads (AFL), an AFL indicates that an abstract is expected from the facility for this incident case. Disease type and registry workflow rules determine whether a CTC or Patient Set is built when the HL7 record is first processed or if an attempt is made to obtain an abstract first. Note: paper path reports received as scanned images are also stored in the record table with the HL7 E-path record type. They use a similar workflow as reports from HL7 files.

- **Pathology Report** – Pathology reports imported into SEER*DMS using a registry-defined file layout. This is a deprecated record type. Electronic path reports for new registries will all be stored in a single record type (HL7 E-Path), regardless of file format.

- **Imaging Report** – Text data, typically transmitted in an HL7 format, containing a physician's interpretation of a medical image (x-ray, PET/CT scan, MRI, mammogram, etc). These records are linked to the Patient Set. Imaging reports may be used for case finding if there are no other data for that case or patient.

Casefinding Records:

- **Casefinding** – Data obtained from medical reports which identify new potential cases. These records typically trigger the creation of Abstract Facility Leads. Registry-defined procedures determine whether a new CTC or Patient Set is built when the casefinding record is first processed or if an attempt is made to obtain an abstract first.

- **NAACCR Casefinding** – A casefinding record transmitted in the NAACCR file format. In some registry algorithms, the record is considered a casefinding record if Class of Case indicates that it is a report of pathology or other lab specimens.

Update Records:

- **Registry Update** – A registry-defined record for submitting field-specific corrections to data previously submitted to the registry. This record type is rarely used and not recommended.

Other Types of Records:

- **Death Certificate** – A record created from state death certificate data files. Death certificates are used to update vital status and cause of death information, and to identify potential new cases. If a death certificate indicates a new reportable cancer, an Abstract Facility Lead is created. If an abstract cannot be obtained, death clearance follow-back procedures are performed and a Death Certificate Only case may be created.

- **Death Notice** – Contains information published in public death notices. These records are used to update follow-up variables for existing patients. This is a deprecated record type.

- **Follow-up Only** – Data reported by a facility to provide follow-up information for existing Patient Sets. In the workflow, Follow-up Only records move through the automatic edit checks and the auto-matching task.

- **Health Index Record** - Data from hospital listings such as diagnostic index, discharge logs, or surgery logs. These data are used for quality assurance related to casefinding and may provide follow-up and treatment data that are consolidated into existing Patient Sets.

- **ME Cohort** – Data from the Hawai'i component of the US Public Health Service's Multi-ethnic Cohort (MEC) Study which are linked annually to the Hawai'i Tumor Registry data.

- **Reject Record** – Data from registry or hospital abstractors indicating that a case is not reportable or should be deleted for some reason. The workflow for this record type is defined, as needed, based on registry-specific requirements.

- **Supplemental** – Follow-up information from non-medical organizations. These data are matched to the database and used in passive follow-up procedures. The Supplemental record type is used for department of motor vehicle, SSA, CMS, NDI, and voter registration data. Specific file layouts are defined in registry-specific configuration files.

In SEER*DMS, records provide a mechanism for storing and accessing the original data values submitted to the registry. Record data do not have to be error-free; records may have null data fields or fields with errors. Problems with the data are generally resolved in the Patient Set data fields at the time of consolidation. In general, it is recommended that registry staff do not change record data fields except to correct errors that prevent the record from being processed in SEER*DMS (see *Chapter 7: Resolving Record Errors* for more information). All changes are documented in the record's audit log which includes the original data value for each field that was

modified. In addition, the original data files are archived and can be accessed by your registry's system administrators.

Linked and Unlinked Records

A "link" is an association between a record and the Patient Set. The link is established after the record is matched against the database. A record moves through the workflow and is either linked to an existing Patient Set that it matched, linked to a newly created Patient Set, or held in the database as an "unlinked" record.

- If an incoming record matches an existing Patient Set, the record is linked to that Patient Set and may be linked to one of the Patient Set's CTCs. The data from the record are consolidated with the data in the Patient Set. This may involve manual processes.

- If a record contains reportable data for a new patient, a new Patient Set may be auto-built using the record's data. The record is linked to the newly created Patient Set. Reportable abstract records are used to build Patient Sets as part of the standard workflow. Pathology reports, death certificate, case finding, and Short Health records can also build Patient Sets.

"Unlinked" records are records that have not been incorporated into a Patient Set, including:

- Records that are in the workflow and have yet to be matched.

- Records for a new patient that do not contain enough data to create a Patient Set.   These would include records that only contain follow-up data.

- Records for new patients that do have sufficient data to create a Patient Set, but the registry prefers to wait for an abstract or other record type. For example, some registries prefer to build a Patient Set from an abstract rather than a pathology report. If a pathology report is received before the abstract, the pathology report would exit the workflow and remain in the database as an unlinked record. If an abstract is not received, the pathology report can be used to create a Patient Set later.

The Patient Set Editor provides access to all linked records, enabling users to review the raw data that contributed to the consolidated Patient Set. The following information is recorded in the audit log entry when a Patient Set is updated via auto-consolidation: Import ID, name of the data file, line number in the file, original and new value of each changed field.
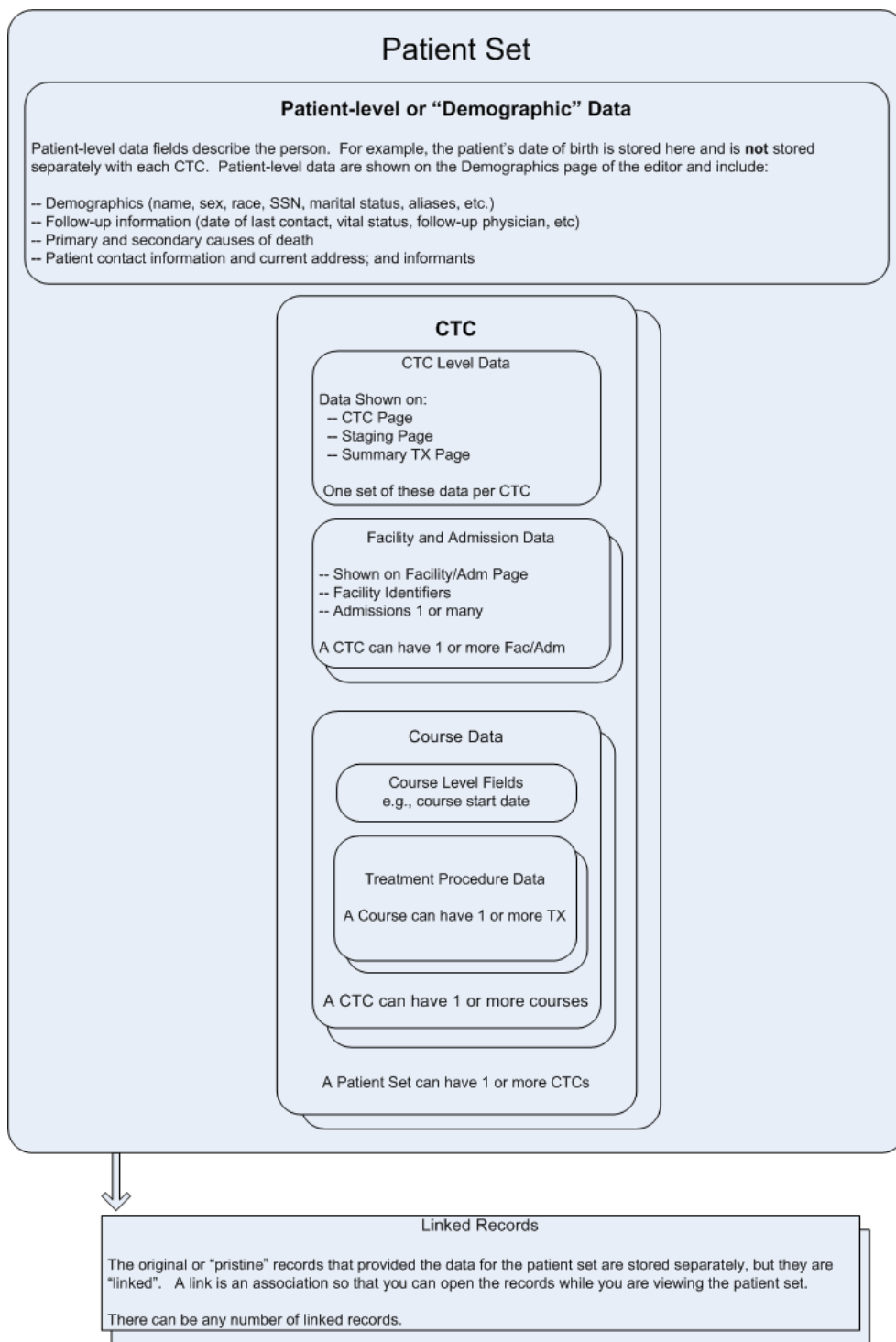
Patient Sets

A **Patient Set** is the packet of data in the SEER*DMS database which contains all data related to the patient. It includes demographic data, information on all reportable cancers, admissions information, diagnostic procedures, treatment information and text documentation. A Patient Set also includes direct links to the original records that contributed to the Patient Set data.

The diagram on the following page illustrates the basic structure of a Patient Set. There is one set of patient-level data, but there can be multiple CTCs – one to represent each of the patient's cancer diagnoses.   In the same regard, each CTC has one set of CTC-level data but may have multiple treatment procedures and admissions. The original records are linked to the Patient Set.

Note: This diagram provides an introduction to the Patient Set and is not comprehensive. There are other data entities associated with a Patient Set that are not noted in the diagram (images,

special studies, AFLs, etc). More complete data diagrams are provided in Appendix A of the SEER*DMS user's manual.

Patient Set

Patient-level or "Demographic" Data

Patient-level data fields describe the person. For example, the patient's date of birth is stored here and is **not** stored separately with each CTC. Patient-level data are shown on the Demographics page of the editor and include:

-- Demographics (name, sex, race, SSN, marital status, aliases, etc.)
-- Follow-up information (date of last contact, vital status, follow-up physician, etc)
-- Primary and secondary causes of death
-- Patient contact information and current address; and informants

CTC

CTC Level Data

Data Shown on:
-- CTC Page
-- Staging Page
-- Summary TX Page

One set of these data per CTC

Facility and Admission Data

-- Shown on Facility/Adm Page
-- Facility Identifiers
-- Admissions 1 or many

A CTC can have 1 or more Fac/Adm

Course Data

Course Level Fields
e.g., course start date

Treatment Procedure Data

A Course can have 1 or more TX

A CTC can have 1 or more courses

A Patient Set can have 1 or more CTCs

Linked Records

The original or "pristine" records that provided the data for the patient set are stored separately, but they are "linked". A link is an association so that you can open the records while you are viewing the patient set.

There can be any number of linked records.

Cancer/Tumor/Case (CTC)

CTC is the SEER*DMS data structure that contains consolidated, tumor-level data. The CTC acronym originally stood for Cancer/Tumor/Case but is often understood to mean consolidated tumor case. A CTC contains data for a neoplasm with topography, histology, and behavior codes that meet the SEER, local, or special study case reporting guidelines. CTC data are submitted to SEER, NPCR, and NAACCR. These are the data used in data statistics and products (eg, SEER*Stat, SEER Explorer, etc).

In the SEER*DMS database, CTC is the name of a data entity within a Patient Set. All diagnostic, staging, and treatment data for a reportable neoplasm are stored in a CTC. The CTC data include summarized treatment information, as well as data regarding each admission and treatment for that CTC. A Patient Set may include multiple CTCs, one for each primary cancer diagnosed for the patient.

Staging

Data fields describing the extent or severity of the cancer are displayed on the Staging pages of the Patient Set editor. These include fields from standard staging systems such as the TNM system developed by the American Joint Committee on Cancer (AJCC), SEER Summary Stage, and Extent of Disease (EOD). Staging data fields are stored in the CTC table of the database.

Staging software, such as the SEER Registrar Staging Assistant (SEER*RSA), are integrated into SEER*DMS. The SEER*RSA API provides coding information for the field lookups and calculates derived data items related to extent of disease, and important site-specific predictive and prognostic factors.

Summary Treatment (Summary TX)

Summary treatment fields are stored in the CTC table of the database and displayed on the Course page of the Patient Set editor. These fields contain the consolidated treatment values and associated dates for the treatments related to the CTC.

A group of SEER*DMS polishers set the values of Summary TX fields to likely or suggested values. A summarized date of therapy is also calculated by the polisher. A suggested summary value may or may not be the best value when combinations of treatment are being consolidated. For example, the correct consolidation for two single agent chemotherapy entries may be a single agent chemotherapy or multi-agent chemotherapy. If the value set by SEER*DMS is not the best, you may over-ride the value. Instructions for visually editing and consolidating treatment data are provided in *Chapter 11: The Patient Set Editor*.

Course

SEER requires that all treatment administered during the first course of therapy be included for each reported case. See the *SEER Program Coding and Staging Manual* for definitions of first course of therapy. Some registries also collect information on subsequent therapy. This could be treatment for recurrence of cancer or for disease progression. SEER*DMS allows the registry to identify the treatment course and will distinguish between treatment courses when deriving suggested best values for treatment. The default value for treatment course is first course of therapy.

Treatment Procedures (TX and TXr)

CTC treatment data are stored in the *treatment_procedure* table of the database and displayed on TX and TXr pages of the Patient Set editor. The Facility ID number associated with the place of treatment is listed next to the page link in the patient navigation box. For example, TX (2344) indicates that this treatment was reported by and administered at facility FAC-2344.

TXr is typically used to identify a treatment that was reported by one facility but given at a different facility. TXr may also be used for registry-specific designations, for example, it is also used to identify treatment procedures with Previous Flag = Yes in SEER*DMS MDCSS.

If the treatment facility is known, the facility number of both the reporting facility and the treatment facility will appear following the TXr. For example, TXr (6431~2344) indicates that according to a cancer report submitted by facility FAC-2344, this treatment was given at facility FAC-6431. A code of 9999 indicates that the facility is unknown. For example, TXr (9999~2344) indicates that according to a record submitted by facility FAC-2344, this treatment was given at an unknown facility. MDCSS Only -- if the reporting facility field is blank for TXr data and the Previous Flag is set to 1, '????' will be displayed in the navigation for reporting facility.

When treatment data are incorporated into a CTC from a NAACCR abstract record, both TX and TXr pages are created if there are treatment data in the NAACCR summary fields that are not included in the NAACCR hospital fields. A TX data page is created from the NAACCR RX hospital fields. A TXr data page would be created for the data stored in the NAACCR summary fields that is additional to the hospital information. The performing hospital for the TXr data is set to the "unknown" facility code which is usually FAC-9999.

Facility Admission

The Facility Admission data catalogues information gathered from each patient encounter at a given facility. If a patient has multiple admissions or encounters at the same facility, the facility may provide admission data for each encounter. These data are stored in the *facility_admission* tables of the database.

Pending and Consolidated Records

When a Patient Set is viewed during a consolidation task, records are listed as either "pending" or "consolidated".   A "pending" record is a record linked to the patient set that needs to be manually consolidated into the Patient Set data. Once it is consolidated by the user, the record will move from "pending" to "consolidated" in the left navigation box. See *Chapter 12: Consolidating Data* for more information on automatic and manual consolidation processes.

Auto-Building Patient Sets and CTCs

An entire Patient Set or components of an existing Patient Set must be created when new reportable data are incorporated into the database. For example, a new Patient Set must be created when it is determined that data for a reportable cancer are related to a new patient. A new CTC must be created within an existing Patient Set when it is determined that data matching an existing patient represents a new, reportable tumor.

SEER*DMS auto-builds Patient Set components based on record data. The transfer of record data to the Patient Set is automated. The values of the appropriate data fields are auto-filled, as determined by registry-specific algorithms which are specified for each record type. For some

fields, the value is copied directly from a record field to a corresponding Patient Set field. The values of other Patient Set data fields are imputed based on multiple fields in the record, or by recoding the value of a single field in the record. The field mappings and recoding algorithms defined for your registry are documented on the Auto-Build page of the Help menu.

These are examples of auto-build events:

- A reportable abstract record does not match an existing Patient Set. The system auto-builds a new Patient Set based on that record. A single CTC is created within the Patient Set. Within that CTC, the number of auto-built admissions and treatments will vary based on the record type and the amount of data available on the record.

- A reportable abstract record matches a patient already in the database but describes a cancer for which that patient has no CTC. During the Consolidation task, a CTC is auto-built when a user links the record to a "New CTC" in the Patient Set. Within that CTC, the number of auto-built admissions and treatments will vary based on the record type and the amount of data available on the record.

- A reportable record matches an existing patient and CTC but describes a new facility admission or treatment relevant to this CTC. In this case, the system attempts to auto-build new facility and treatment data fields for that CTC.

SEER*DMS auto-builds Patient Set data fields using data from a single record. If multiple records for a new patient are consolidated, the Patient Set data fields are auto-built from one of the records when the Consolidation task is started. SEER*DMS selects an abstract record (NAACCR Abstract, NAACCR Modified, or Health Record) to use for auto-building. If there are two or more records, one of the records is chosen as the best ("master") record for auto-building based on record type, reportability and date loaded. Some registries have custom logic for finding the master record. Other incoming records are then consolidated with the new Patient Set.

Audit Logs

Whenever a record or Patient Set data field is changed by a user or the system, an entry is made in an audit log. A separate audit log is maintained for each record and Patient Set in the system. The information in the audit log includes the user or process which modified the data, comments related to the change, the date and time of the modification, and the original and modified value of each data field that was changed. A user may provide a separate comment for each data item changed or one comment for all data items changed within a single session.

When the audit log is displayed in the record or Patient Set editor, the changes are grouped by events. Events include workflow tasks, system tasks, and other processes that modify patient data. Patient Set and Record audit log events are:

- **Ad-hoc Editing** – The changes were not made during a worklist task; a user opened the record or Patient Set via the patient lookup and modified one or more data items.

- **AFL Update** – Changes made to an AFL were incorporated into the record that triggered the AFL. At this time, data items related to reportability are the only fields changed via this process. The record's reportability is changed when the AFL status is set to "Not a Reportable Cancer".

- **AFUP Update** – Patient set data were updated when a response to Active Follow-up (AFUP) was entered.

- **Auto-Build** – The Patient Set or a CTC was auto-built from a record.

- **Auto-Coding** – A change was made to a record as it entered the workflow. Registry-specific polishers are executed in the auto-coding workflow task.

- **Auto-Consolidation** - The change was made during an automated process to consolidate incoming data.

- **Auto-Linking** – A record was linked at the patient-level or CTC-level during an automatic linkage process. The automated linkage task may have been triggered after an automatic matching task, a manual matching task, or by a user forcing a record to move from one Patient Set to another.

- **Automatic Path Matching** – Changes were made by the automatic path matching task in the pull list workflow.

- **Auto-Screening** – The change was made during the automated screening task. This task sets the reportability status of an incoming record.

- **Consolidate** - The change was made by a user during a manual consolidation task.

- **Consolidate FUP**- The change was made by a user during a Consolidate FUP worklist task. This is no longer an active task; however, this entry may appear in the audit log.

- **Data Entry** – One audit log entry is made when a record is saved in a data entry session.

- **Direct Database Update** – Record or Patient Set data were modified by an external script (audit log entries were not created by all external scripts in early versions of SEER*DMS).

- **Duplicate Record Processing** – Incoming abstract records are matched against other abstract records to determine if it is a subsequent abstract sent by the facility. A subsequent abstract for the same facility and case is processed as a "duplicate" abstract. The original abstract is updated with changes from the new abstract; and the new abstract is deleted.

- **Edit or Polisher System Tasks** – The change was made by a SEER*DMS "polisher" which executed after an Edit or Polisher system task completed. A polisher is a background utility that, according to SEER or registry-defined algorithms, imputes data values based on values in other fields.

- **Image Data Entry** – Record data were updated during a manual Image Data Entry task. An Image Data Entry task is created for each image record imported via the autoloader.

- **Import** – An image was linked to the record via an import process. The image file was imported with linkage information so that the image could be auto-linked.

- **Initial Conversion** – The change was made during an automated process to load data from the legacy system.

- **Legacy Audit Log** – Changes documented in this event were made on the registry's legacy system. The audit log information was transferred with the migration data when the registry first started using SEER*DMS.

- **Link AFL Source** – This is a deprecated automated task but will show in audit logs prior to 2020. The record was auto-linked to a Patient Set when the record's AFL matched an abstract. In this task, the record with the AFL is linked to the same patient set as the closing record (the abstract).

- **Link Modification** – This event is only used in the record audit log. It indicates that a user opened a Patient Set and changed the way a record was linked (e.g., changed a record from linked at patient level to linked to an existing CTC). The record itself was not opened directly but was modified via a change made in the Patient Set editor.

- **Mass Change** – Changes were made as specified in a mass change import. The Import ID is listed with the event information.

- **Path Auto-Linking** – A path report was auto-linked to a Patient Set after a Path Matching or Review Path Matching task. These tasks are part of the Pull List Workflow that is only used in the Seattle registry.

- **Pathology Screening** - The change was made by a user during a manual Pathology Screening task.

- **Pre-record Auto-Linking** – Novel data stored in the *pre_record* table were automatically linked to or unlinked from the Patient Set. Claims, pharmacy, EHR, and other types of novel data are stored in the pre_record table.

- **Pull List Algorithm** – The change was made by the automatic Pull List Algorithm Task. The Pull List Workflow is only used in the Seattle registry.

- **QC Patient Set** – The change was made by a user during a manual QC Patient Set task.

- **QC Record** – The change was made by a user during a manual QC Record task.

- **Rapid Follow-up** – This is a deprecated automated task but will show in audit logs prior to 2018. Patient set data were updated during an automated Rapid Follow-up worklist task.

- **Resolve Duplicate Patient Set** – A Patient Set data was updated during the manual task to review Patient Sets identified as potential duplicates.

- **Resolve Pat Set Errors** – The change was made by a user during a manual Resolve Patient Set Errors task.

- **Resolve Record Errors** - The change was made by a user during a manual Resolve Record Errors task.

- **Review Data** - The change was made by a user during a manual Review Data task.

- **Review Pull List** – The change was made by a user during a manual Review Pull List Task. The Pull List Workflow is only used in the Seattle registry.

- **Review Special Studies** - The change was made by a user during a Review Special Studies worklist task.

- **Routing Script** – A workflow routing script changed a field on the record. The workflow scripts update system fields that indicate the current consolidation status of a record (e.g., the flags indicate whether auto-consolidation was successful or manual review is required).

- **Screening** - The change was made by a user during a manual Screening task.

- **System Task** – Changes were made by a SEER*DMS system task.

- **Two-tier Admission** – Changes were made by the automatic task that applies the two-tier algorithm. This task is part of the pull list workflow used by Seattle.

- **Visual Edit Pat Set** – This is a deprecated automated task but will show in audit logs prior to 2018. The change was made by a user during a manual Visual Edit Patient Set task.