Chapter 5: Importing Data Files

Data files may be imported using the SEER*DMS interface or auto-loaded from a specified network location. All data files must adhere to file specifications documented in the Imports section of the Help menu. You may use the Import Manager to review summary information about each import and to access the source data files.

Typically, imported data are stored in the record table of the SEER*DMS database and can be viewed within the SEER*DMS record editor. Data that are stored in the record table are said to be “persisted” in the database. You may specify the Import ID in the Patient Lookup, Worklist, or Import Manager to search for persisted records or to determine whether the records have completed the workflow. Most supplemental imports are not stored in the database after they are processed. Registry configurations determine whether supplemental data are stored in the database and the methods used to process these data (see the *Processing and Storage of Imported Data* section of this chapter for more information).

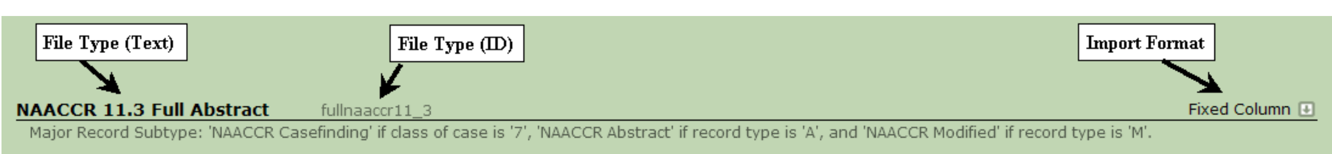
Mass change imports are unique because these data are used to make direct updates to the database. An audit log entry is made in the patient set, record, or other entity that is updated via mass change. Data in mass change files are not stored in the record table or processed via the workflow. Mass change imports are fully documented on the Imports help page within SEER*DMS.

In this chapter, you'll learn about

- Import Specifications and Algorithms
- Processing and Storage of Imported Data
- Import Manager
- Importing Data Files within SEER*DMS
- Auto-Loading Data Files
- Reviewing Status and Outcomes of Imports:
    - Import Information Page
    - Import Information – Files Tab
    - Reviewing Field Frequencies – Statistics Tab
    - Consolidation Summary – Updates Tab
- Import Review Task
- Finding Records and Tasks Related to an Import
- Reports Related to Data Imports

Import Specifications and Algorithms

The file layout and configuration settings for imports are documented on the Imports help page. The File Type (in full text and as a system ID), Import Format, and a brief description are displayed for each import type. SEER*DMS supports the following Import Formats: Fixed Column, HL7, Delimited, Mass Change, and Image Files.



Double-click the File Type or click the down arrow on the right-hand side to expand the documentation section. The Mass Change Import type is unique; its help section contains a detailed description of the file format, validation, and entities that can be updated via Mass Changed. The help sections for all other import types include the following sections:

- **Major Record Subtype -** The record subtype determines the routing of the data in the workflow and other processes. For example, NAACCR abstract is the major record subtype for all versions of NAACCR abstract data (NAACCR 10.0, 10.1, 10.2, 11, etc). The record types are defined in the Records section of *Chapter 2: Records and Patient Sets*. The types of records actually used in your registry are listed in the *lkup_record_major_subtype* database table.
- **Line Length** – The number of characters per line in fixed column imports. If there are multiple lines per record, this is the number of characters in the first line of the record (e.g., NAACCR Extended imports use one line of codes followed by multiple lines of text). There is no length restriction for the text lines in a multi-line import like NAACCR Extended.
- **File Layout for Delimited Files** – This section defines the number of fields, the method used to embed the delimiter character and quotes in a value (Microsoft Excel or standard), the character used as the delimiter, and whether a single or double quote is used as the quote character. If the file includes a header line then "Ignores First Line" is set to true. This describes the two methods for handling special characters in delimited files:
    - Excel
        - All values may be enclosed in quotes, but quotes are only required if the delimiter character, newline characters, or quotes are embedded in the value.
        - To embed the delimiter character or newline within the field, surround the entire value in quotes.
        - To embed a quote within the field, surround the entire value by quotes and any quote in the value must be double-quoted
        - If you enclose a field in quotes then you may not have any characters between the delimiter and the quotes.
    - Standard
        - To embed the delimiter character within the field, precede the character with a backslash
        - The quote character is not a special character, all quotes between delimiters are considered part of the value
        - All values will be trimmed
- **Import Review Task** – A manual Import Review task may be limited to data that generate errors, warnings, or have duplicates. Or an Import Review task may be created for all imports of this type.
- **Processing** – This configuration setting determines whether the imported data are processed in the workflow or in import routines external to the workflow. Refer to the *Processing and Storage of Imported Data* section of this chapter for more information.
- **Stored in Record Table –** This configuration setting determines whether data of this type are stored in the record table of the SEER*DMS database (see the *Processing and Storage of Imported Data* section of this chapter for more information).
- **Algorithmic Duplicate Checking –** A matching algorithm that identifies records which may not be an exact duplicate of previously loaded records, but provide the same data as previously loaded records. The criteria used in the matching algorithm are documented on the Matching help page. This is one method for identifying records that are not exact duplicates but do not provide any more data than a previous loaded record; partial hash checking is a second and more commonly used approach to the same problem.
- **Fields Ignored for Partial Hash –** The fields listed here are ignored when determining whether a record is a duplicate of another record within the same import or previous import.

- **Pre-processing Matching Algorithm -** A matching algorithm executed as the data are imported; typically, a pre-processing matching task is only used for supplemental imports. If there is no Patient Set that meets one of the match criteria, the supplemental record is not imported and is counted as ignored. The matching criteria are documented on the Matching help page.
- **Table of Import Fields –** This table shows the mapping from the import file to the SEER*DMS database. The columns of the table vary by import type, as described below.
  - Import Field – This is shown for Fixed Column and Delimited imports. The text in the Conversion Rule identifies the field for HL7 imports.
  - File Position – Start Column, End Column, and Length are shown for Fixed Column imports. Index and Max Length are shown for Delimited files.
  - Database Field – The database field (defined by *table.fieldname*) to which the import field is mapped.
  - Conversion Rule – Processes performed during the import are documented here. This includes rules to "convert" (reformat) the data and rules to validate the fields. Warnings and errors generated during the import and reported in the Import Review task are also documented. Conversion rules for specific fields are documented in the Conversion Rule column of the table; conversion rules that do not apply to a specific field or are applied to multiple fields may be documented below the table.

Processing and Storage of Imported Data

Each import algorithm includes a setting for "Processing"; this parameter determines whether the imported data are processed in the workflow or in import routines external to the workflow. The "Processing" field has two possible values as described below; the settings used in your registry are documented for each import on the Imports help page.

- **Processing = Workflow**. By definition, certain data types must be processed through the workflow. Abstracts, casefinding records, death certificates, and other types of records that need to be screened for reportability must be processed via the workflow. You may specify the Import ID in the Patient Lookup, Worklist, or Import Manager to determine whether the records have completed the workflow. You may also use the Import Manager to review a worklist summary for an import.
- **Processing = Non-workflow Import Routines**. When this setting is used, data are only sent into the workflow if manual review is required. The majority of supplemental imports are processed in procedures external to the workflow (e.g., the records are matched using a pre-processing matching algorithm that is not part of workflow processes). This reduces the load on the system at the time of the import. .
- **Mass Change**. Mass change files allow you to make direct database updates in a manner that generates an audit trail (an audit log entry is added for every change). Mass change processing does not involve the workflow.

Typically, supplemental import data are not stored in the record table unless a manual task is necessary to review the results of the automatic match task or resolve conflicts encountered during auto-consolidation. If SEER*DMS is able to match and consolidate the data without a manual task, the patient set is updated; and the Import ID, filename, and line number are recorded in the patient set's audit log. However, SEER*DMS does support other options related to data storage. The settings used for your registry's imports are documented on the Imports help page. System reports can be used to view changes made to data based on supplemental imports. In addition, a summary of updates related to a supplemental import is available on the Import Info page in SEER*DMS. Instructions for accessing the Import Info page are provided in the *Import Information* section of this chapter.

- **Stored in Record Table = All data**. Imports processed via the workflow must use this setting (this includes abstracts, casefinding, death certificates, and other records that are screened for reportability). This setting is rarely assigned to supplemental imports.
- **Stored in Record Table = Not stored unless manual review required**. This is the recommended setting for supplemental imports.
- **Stored in Record Table = Matches and data that require manual review**. This option may be used for supplemental imports if all data items must be retained as a reference. The import data are matched against the database in import routines. If a match is found, the data are auto-consolidated into the patient set, a record is created and the record is linked to the matching patient set. In addition, a record is created if the results of the automatic matching or consolidation tasks require review.
- **Stored in Record Table = Non-matches and data that require manual review**. This option may be used for supplemental imports, however, it may result in a large number of unlinked records that would be viewed during manual match tasks and database searches. The import data are matched against the database in import routines. A record is created for each of the non-matches. Matching data are auto-consolidated; records are not created for data that can be matched and consolidated in automatic tasks. As always, a record is created if the results of the automatic matching or consolidation task require review.
- **Mass Change**. Direct database updates made via mass change are documented via audit log entries. The original mass change file can be accessed via the Import Info page.

As noted above, supplemental data are stored in the database if a manual matching or manual consolidation task is created. Deterministic algorithms are recommended for supplemental matching (criteria resulting in “possible” matches are rarely defined for supplemental data). Therefore, in most registries, a manual supplemental matching task is only created if the incoming data is a “perfect” match to two or more patient sets. The matching algorithms used in your registry are documented on the Matching help page. A manual consolidation task is created if the incoming data conflict with the patient set’s data and an auto-consolidation “failure” is created. Auto-consolidation rules and failures are documented on the Auto-Cons help page.

Import Manager

Requires system permission: *import_electronic* or *import_manual*

Use the Import Manager to locate records loaded in a particular import, to view or download a copy of the source data files, or to determine the current workflow location of each record in an import.

To access the Import Manager, select **System > Imports**. The following data columns are shown in the Import Manager:

- **Import ID** – A unique ID assigned to each import. Click the ID to view details about the import as described in the *Import Information Page* section of this chapter.
- **Import Date** – Date and time that the import was initiated. This is the timestamp recorded when the data are analyzed and may be earlier than the date and time that the records are loaded.
- **Facility** – The facility selected by the user or set in the autoloader configuration file for this import. This facility is used by default as the record’s reporting facility when the reporting facility is not available on the record.
- **User** – The user who initiated the import. “seerdms” will be listed as the user for autoloader imports and data migrated at the time of deployment.
- **Records** – The number of records imported. This value will be zero until the Import Review task is completed and the data are successfully loaded.

- **Type** – The import's Media Type is displayed in this column. The possible values are Files, Data Entry, or Migrated Data. If Type is Files, the number of files in the import will be displayed in parentheses.
- **Comment** – The import-wide comment entered on the Import Source page when starting a new import or entered in the <comments> tag in the autoloader XML.
- **Action** – Click **Lookup** in the Action column to search the database for records loaded in the import. A maximum of 500 records will be listed. Use RPT-009A to generate a complete list of the records loaded in an import.

Use the filters at the bottom of the screen to search for an import based on data items displayed in the Import Manager. You may also filter by data **Type** or search for a particular file using the **File Name** filter.

Importing Data Files within SEER*DMS

Requires system permission: *import_electronic* and *contact_view*

You may use the SEER*DMS interface to import data files from any local or network disk drive that you can access. The files will be uploaded through your Web browser and are subject to HTTP size limitations. Use the auto-load feature to import data files that exceed these limitations.

*To import data files using the SEER*DMS interface:*

1. Select **System > Imports** to access the Import Manager.
2. Click **New Import**.
3. Specify the import **Facility** by typing a facility ID (the FAC prefix is not required) or by using the lookup to select a facility from the Facility List. You also have the option of specifying a **Contact** at that facility.
4. Set the **Media Type** to *Files*.
5. You may enter a **Barcode** or import reference number, if applicable.
6. You may enter notes regarding the import in the **Comments** field.
7. Click **Next**.
8. If a Data Exchange Agreement (DEA) between the registry and facility is defined, you will have the option to indicate whether the data in the import are related to the agreement. The DEA fields in Contacts and Imports are for tracking purposes only. Use these fields according to your registry's policies.
9. In the **Add File** section, click **Browse**. Explore to the appropriate folder and select a file.
10. Select the **File Type**. Click the information icon to review the layout for that file type.
11. You may enter a **Comment** regarding this specific file.
12. Click **Add**. SEER*DMS will determine if the file was loaded before. If the file is a duplicate, a warning message will be displayed. To remove the file from the import, click the **Remove** link in the **Action** column. (SEER*DMS will allow you to over-ride the warning and load a duplicate file. Use this feature cautiously and only after consultation with registry managers and technical support staff).
13. Repeat steps 9-12 for each file which is being imported from this facility. Each file will be listed in the **Current Files** section as it is added.

14. Click **Import**.

15. SEER*DMS will analyze the data files. If the results of this analysis require manual review, an Import Review task will be created and must be completed in order to load the data. A manual review is required if there are errors, warnings, or duplicates in an import; and a manual review is always required for some import types (this is determined by the Import Review Task setting for the import; parameter settings are shown on the imports Help page). Refer to the *Import Review Task* section of this chapter for instructions on completing this task. The Import Review Task provides an error summary and frequencies of all data fields. Once you review these statistics, you will have the option of loading the data or terminating the import for some or all data files.

Auto-Loading Data Files

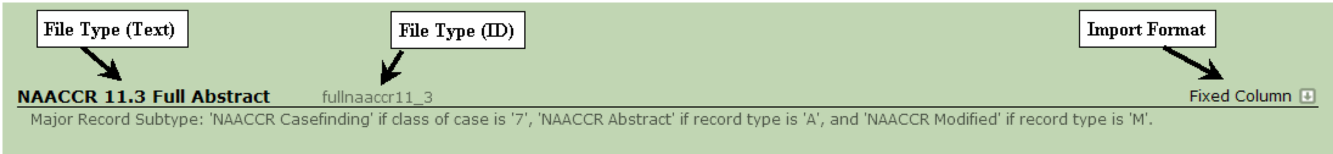
SEER*DMS can be configured to automatically load data files stored in designated locations on the network. On a regular basis, an automatic workflow task imports all files in the main auto-loader directory and the sub-directories within the autoloader directory. Each folder must contain its own copy of a configuration file (autoloader.xml). This file defines all import settings (Import Facility, Contact, Comments, File Type for each file, and information related to data exchange agreements). The Import Facility and File Types are required; all other elements are optional.

This is the Document Type Definition (DTD) for the autoloader.xml:

```
<?xml version="1.0" encoding="UTF-8"?>

<!ELEMENT autoloader (organization,representative?,comments?,meets-dea?,file-type)>
    <!ELEMENT organization (#PCDATA)>
    <!ELEMENT representative (#PCDATA)>
    <!ELEMENT comments (#PCDATA)>
    <!ELEMENT meets-dea (#PCDATA)>
    <!ELEMENT file-type (when*)>
    <!ATTLIST file-type default CDATA #REQUIRED>
        <!ELEMENT when (#PCDATA)>
        <!ATTLIST when ext CDATA #REQUIRED>
```

- **organization** – This element is required. Specify the Import Facility using this element. You may specify a complete facility ID or the organization name. It is recommended that you use the facility ID. If you specify the name then the text must be an exact match to a name in the SEER*DMS Facility List.
- **representative** – This element is optional. You may specify a complete contact ID or the contact's name. If you specify a name, it must be an exact match to a name in the SEER*DMS Contact List.
- **comments** – This element is optional. These are import-wide comments that are displayed in the Comment column of the Import Manager.
- **meets-dea** – This element is optional. If a Data Exchange Agreement (DEA) between the registry and facility is defined, you may specify whether the data in the imports are related to the agreement. The DEA fields in Facilities and Imports are for tracking purposes only. Use these fields according to your registry's policies.
- **file-type** – This section is required. Use this section to specify the File Type of files to be loaded via this folder. To do this, you must associate an import File Type ID with one or more extensions. The File Type ID is case sensitive but the file extension is not (*.pdf and *.PDF are equivalent to the autoloader). You must also assign a default File Type for unknown extensions. As shown below, the import help page includes the File Type ID for each import available in your registry.



In the following example, all files placed into this autoloader folder will be autoloaded as Linked Image files.

```
<?xml version="1.0"?>
<!DOCTYPE autoloader PUBLIC "-//IMS//DTD Autoloader Definition//EN"
"autoloader.dtd">
<autoloader>
   <organization>FAC-0001</organization>
   <comments>Autoloaded via the linked image folder for FAC-0001.</comments>
   <meets-dea>false</meets-dea>
   <file-type default="linked-image">
   </file-type>
</autoloader>
```

The example below allows different file types in the same folder and uses a default file type for undefined extensions. A *.pdf file will be uploaded as a linked image file. A *.txt file will be uploaded as an HL7 Epath file (note that Epath files use registry-specific algorithms). All other files will be uploaded as NAACCR 11.3 data.

```
<?xml version="1.0"?>
<!DOCTYPE autoloader PUBLIC "-//IMS//DTD Autoloader Definition//EN"
"autoloader.dtd">
<autoloader>
   <organization>FAC-0001</organization>
   <comments>This file was autoloaded.</comments>
   <meets-dea>false</meets-dea>
   <file-type default="fulllnaaccr11_3">
      <when ext="pdf">linked-image</when>
      <when ext="txt">la-epath-2.3</when>
   </file-type>
</autoloader>
```

Technical Specifications

- All subdirectories must have an autoloader.xml file to support automated pickup. If a directory is missing the autoloader.xml file, no file will be imported from that directory (but any directory within that directory will still be processed).
- On a regular basis, the Quartz scheduler fires off an automatic workflow task to gather any files residing in the autoloader folders and import them into the system. The frequency of the directory scan is determined by the *system.schedule.autoloader* configuration parameter. This uses a standard crontab string to set the schedule.
- The autoloader will wait for a stable directory tree before it starts processing any files and directories under that tree. If the autoloader detects a change in the directory tree, it will sleep for n-seconds before trying to scan the directory tree again (as specified in the *importer.autoloader.quiet_time* configuration parameter). This technique prevents the autoloader from starting to load files before all the import files have been copied into the proper directory. A single Import ID is assigned to all files loaded in a single task.

- If an error is generated, the file is not loaded and it is copied to the folder defined by the *importer.autoloader.failure.dir* configuration parameter (typically, the folder is named *.errors*). Email notification is sent to the email address defined by the *system.email* configuration parameter.
- Imports loaded via the autoloader are attributed to the user defined by the *importer.autoloader.user_name* configuration parameter.

Autoloader Folders

You should create autoloader folders to meet your needs based on these considerations:

- The main autoloader directory is defined by the *autoloader.dir* configuration setting.
- Any number of subdirectories can be created within the main autoloader folder.
- A folder is ignored by the autoloader if its name starts with a period (".").
- Copies of files should be placed in the autoloader folders. The registry should maintain an archive of the data files in another location.
- An import facility must be specified for each import in SEER*DMS. When using the autoloader, one facility must be specified in each autoloader.xml configuration file. Therefore, you must create a separate folder for each Import Facility (or even contact, if your registry tracks to that level of detail). If you want to specify different facilities (or a different contact at the same facility) then you can create folders by facility.
- The level of complexity that you create is up to you. Registries are encouraged to experiment and devise a plan during the beta-testing period or on the registry test server.

Import Information Page

Requires system permission: *import_electronic* or *import_manual*

To access the Import Info page, select System > Imports and click the ID of an import shown in the Import Manager. The information provided at the top of the screen is described below. The tabs displayed in the bottom half of the page are described separately in subsequent sections of this chapter.

- **Import Overview:**
    - **Import ID** – Unique ID assigned to the import.
    - **Facility** – The facility selected by the user who initiated the import or set in the autoloader configuration file. This facility is used as a default value for reporting facility when reporting facility cannot be set using data fields on an imported record.
    - **Contact** – This is an optional field that may be used to identify a representative at the import facility who may be contacted about data provided by that facility.
    - **Media Type** – Possible values are: Files, Data Entry, or Migrated Data.
    - **Barcode** – Barcode or reference number entered by the user who imported the data.
    - **Import Date** – Date and time that the import was initiated. This is the timestamp recorded when the data are analyzed and may be earlier than the date and time that the records are loaded.
    - **User** – The user who initiated the import. “seerdms” will be listed as the user for autoloader imports and data migrated at the time of deployment.
    - **Comments** – The import-wide comment made by the user who started the import or entered in the <comments> tag in the autoloader XML.
- **Import Summary:**
    - Number of files in the import.
    - Number of records in all files.

    - Number of valid records in all files. This is the number of records loaded during the import process. Valid = (Records – (Errors + Warnings + Duplicates + Ignored))
- **Current Status:**
    - Number of records in the database. This is the number of records that are persisted. This number will be less than the number of valid records in the Import Summary if not all records are persisted for this import type or if imported records were deleted. Click the link to view a sample of the records via the Patient Lookup. A maximum of 500 records will be listed.
    - Number of records involved in workflow tasks and the number of records per workflow task. Click the link for each task type to view the tasks in the worklist.

Import Information – Files Tab

The following information is listed for each file on the Files tab that is displayed on the Import Info page and in Import Review tasks.

- **Filename** – Click the filename to open or download the file. Note: if you are logged into your test server, the filename will only be a link if the file was imported on the test server.
- **Records** – The total number of records in the data file. The word "stopped" will be shown in this column if the import was terminated because the number of errors exceeded 200.
- **Errors** – The number of errors in the file. An import error is generated when a field fails import validation or when the record is not formatted correctly. A limited number of fields are validated during the import process; the field-level validation for each import is documented on the Import Help page. The number of errors will be a hyper link if the value is greater than zero. Click this value to review the error messages. SEER*DMS stops processing the import file when the error count exceeds 200.
- **Warnings** – The number of values Import warnings are generated for a limited number of data fields. These include invalid values for some date fields, and warnings when a text field is truncated in order to meet database constraints. Refer to the Import Help page to see which fields are validated; and whether that validation results in an import error or warning. The number of warnings will be a hyper link if the value is greater than zero. Click this value to review the warning messages.
- **Duplicates** – If two records within the same import are exact duplicates, one will be loaded and the other will be counted here. This count also includes records which are not loaded because they are exact duplicates of records loaded in an earlier import (based on hash maps); and records which are not exact matches but provide the same data as previously loaded records. Non-exact duplicates may have differences in non-essential fields but provide the same data in key fields. Refer to the Imports Help page to determine if non-exact duplicate checking is performed for a specific import type; if used, it will be implemented as "Algorithmic Duplicate Checking" or as "Fields Ignored for Partial Hash". To determine the specific types of imports, refer to the system log. The log messages recorded at the time of the import analysis include these counts:
    - File-dup – records that were not loaded because they were exact duplicates of records in the same import (the duplicate may be in the same file or in a different file within the same import)
    - Hash-dup – records that were not loaded because they are full or partial hash duplicates of records loaded in an earlier import
    - Alg-dup – records that were not loaded because they are a duplicate as determined by Algorithmic Duplicate Checking defined for this import.
- **Ignored** – To reduce the number of extraneous supplemental records in the database, some data types are evaluated to determine if they provide useful data. A record that is deemed as not useful is "ignored" (it is not processed further and is not loaded into the

record table). Rules to ignore records are defined in importer algorithms of certain import types. For example, some supplemental record types are ignored if social security number is null (e.g., this rule applies to CMS records in some registries). Importer rules are documented on the Imports Help page.

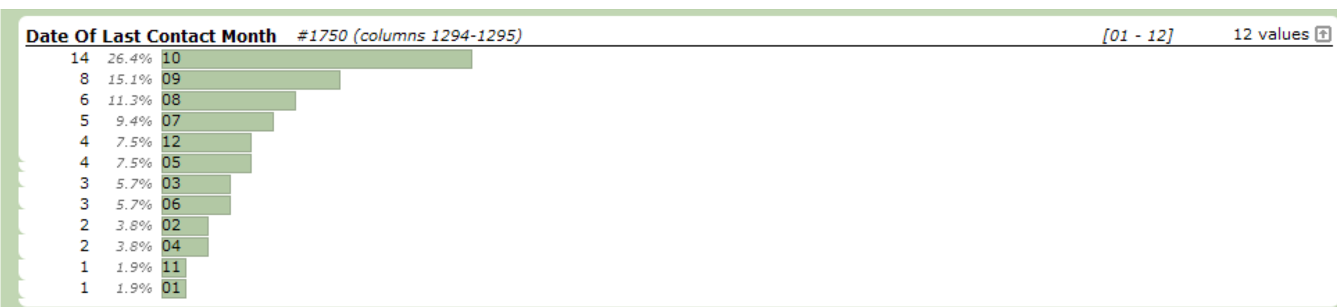
- **Valid** – The number of records which can be loaded. Valid = (Records – (Errors + Duplicates + Ignored)). If you complete the import and this is a data type that is persisted, this is the number of new rows that will be created in the record table and the number of records that will enter the workflow.
- **File Type** – The import file type.

Reviewing Field Frequencies - Statistics Tab

Requires system permission: *import_resolution*

It is critical that data which are imported into the system are valid. Some import files need to be pre-processed, while others can be imported as-is. If there were problems with the data preparation, either at the sending facility or within the registry, an invalid import file could incorrectly modify a large volume of system data. This type of problem is very difficult to remedy once the data are loaded into SEER*DMS; each record immediately begins to move through the workflow and affect changes in patient set data. It is the responsibility of registry staff to define and implement comprehensive quality control procedures to prevent import errors. This should include but not be limited to the review of input field frequencies to identify invalid or unacceptable values. The field frequencies are shown on the Statistics tab in Import Review tasks and on Import Info screens.

The fields listed on the Statistics tab are database fields mapped directly to import fields and database fields assigned via import conversion rules. If multiple files are included in the import, a drop-down list allows you to switch from one file to another. The fields are listed in alphabetical order. You may use the linked letters at the top of the list to view fields which begin with that letter or use Firefox search tools to search for a fieldname or column position.



The figure above shows the type of information shown for each data field. The following describes the information that is displayed and the text actually used in the example above.

- **Field name** - shown in bold above the line: *Date of Last Contact Month*
- **NAACCR Item Number and/or File position –** If this is a NAACCR data item, the item number will be displayed (#1750 in the example above). If the field is mapped directly to an import field, the import field's position will be displayed as column position for fixed column imports or field index for delimited imports. (File position is not displayed for fields that are assigned based on conversion rules; and file position is not shown for HL7 fields.)
- **Range of values** - shown in brackets above the line in center: *[01 – 12]*
- **Number of different values** – only shown if there are more than five different values for the field. The number of different values ("*12 values*" in the example) is displayed on the far right of the page, just above the line. Up and down arrows enable you to toggle the

display to show all values or hide all but the top five values. The figure below shows all twelve values for the example.

- **Statistics and bar chart -** For each value encountered for the field: the number of records, percent of records, and value. The value is shown within a histogram bar. In the example, 14 records have a value of “10” for DOLC Month. 14 equals 26.4% of the number of valid records in the file. Note: frequencies are not shown if the number of differing values exceeds display limitations. For example, it would be impractical to show a frequency for street address which would be unique for every record.

Consolidation Summary – Updates Tab

The Updates Tab is displayed if the imported data are processed in import routines rather than the workflow. Typically, Supplemental imports are the only imports processed in import routines (this is described in the *Processing and Storage of Imported Data* section of this chapter). This tab shows a summary of the import’s match results and of updates made to patient sets via automatic and manual consolidation of the import’s data. The values shown on the Summary Tab are:

Patient Sets Updated. This is the total number of Patient Sets modified via the consolidation of this import’s data. This includes updates made in automated and manual consolidation tasks.

No Matches. This is the number of records in the import file with a match score of 0 for all patient sets. These records were not used in consolidation and did not create a manual task to review possible or multiple matches.

Manual Match Tasks. This is the number of records in the import file that triggered manual match tasks (Supplemental Match tasks). This count and the number of manual consolidate tasks give an indication of the amount of manual labor associated with an import. A manual match task is required if, based on registry algorithms, a person must review the results of the automated match. Registry management should reconsider the matching algorithms if supplemental data result in large numbers of manual match tasks. Deterministic algorithms are recommended for supplemental matching (criteria resulting in “possible” matches are rarely defined for supplemental data). Therefore, in most registries, a manual supplemental matching task is only created if the incoming data is a “perfect” match to two or more patient sets. A manual task would be created if the supplemental matching algorithm includes criteria that results in scores greater than 0 but less than 1000; and the only matches to an incoming record are patient sets meeting those criteria.

Manual Consolidate Tasks. The number of Consolidate FUP tasks triggered by data in this import. A manual Consolidate FUP task is created if the incoming data conflict with the patient set’s data and an auto-consolidation “failure” is created. Auto-consolidation rules and failures are documented on the Auto-Cons help page.

Counts of Updated Fields in Patient Set Data. This is a list of Patient Set fields modified via the consolidation of this import’s data; and the number of changes per field. For some fields, the count may exceed the total number of Patient Sets Updated, for example, the count for Survival Time may include changes to the field on 2 or more CTCs within the same Patient Set.

Import Review Task

Requires system permission: *import_resolution*

Once an import is initiated, SEER*DMS will analyze the data files and create a manual Import Review task prior to loading the data into the database. The Import Review Task provides an error summary and a listing showing the frequency of each value for all data fields.

To complete the Import Review tasks:

1. Click the **Import Review** link in the **My Tasks** section of the worklist summary on the home page.
2. Click the Task ID to open the task. At this stage, the files have been checked for errors and analyzed to create frequencies of values in all data fields. The records have not been loaded into the system.
3. Carefully review number of records listed for Records, Errors, Warnings, Duplicates, Ignored, and Valid in the Import Files section of the page. Each of these fields is described in the previous section.
4. Follow steps 5 and 6 for each file with errors.
5. If an incorrect file type was specified for the file:

   a. Select the correct **File Type**. An asterisk is displayed next to the original File Type selected when the import was initiated. The field frequencies shown in Statistics are based on the layout of the File Type that is displayed with the asterisk.

   b. Set **Action** to *Accept.*

6. If the file type is correct but the file has errors:

   a. Click the value displayed in the errors column.

   b. Review the error messages. You must determine whether you want to reject all records in the file or import the records that did not cause errors. (Note: Once an error is encountered, the first record with the error and all subsequent records will be displayed in red.)

   c. Click **Close** to close the Import File Problems window.

   d. Set the appropriate **Action** for this file.

      i. If you would like to contact the reporting facility and obtain a replacement for a file with errors, select *Reject* to reject all records in the file (records in other files will not be affected).

      ii. If all problems are caused by duplicate records, you may use *Re-import valid only* to accept the non-duplicate records. If there are other types of problems, this action is not recommended.

7. To close the Import Review task:

   a. If an Action is specified for each file, click **Import** to load the acceptable data and exit the task. If the data cannot be loaded or a new file type was selected for a file, a second Import Review task will be created and assigned to you. Otherwise, SEER*DMS will begin to load and process the data. An email message will be sent to you when the import step is complete. The data will then be processed in the workflow or in import matching and auto-consolidation routines. The Import Info page and system reports can be used to review the status of imported data.

   b. If you would like to close the Import Review task without importing any of the files, click **Reject All**.

   c. To exit without removing the task from the workflow, click **Cancel**. The task will be assigned to your user account.

Finding Records and Tasks Related to an Import

Each imported record travels through the workflow, triggering the automated and manual "tasks" that must be performed to process the data. There are a variety of tools in SEER*DMS for determining the current location of a record in the workflow and the end result of the workflow processing. Use these tools to determine whether there are open tasks related to the records and if the records have been consolidated into the patient set data.

To view a summary of open tasks related to the import:

1. Select **System > Import**. The most recent imports will be listed on the first page of the manager. The pages of the managers include all data entry sessions, autoloaded data, and data loaded in the system interface. You may use the filters to search for a specific import. The filters correspond to the fields described in the *Import Manager* section of this chapter.
2. Click the **Import's ID**.
3. The **Current Status** on the right side of the screen will be displayed.
4. Click one of the links to view or access the tasks via the worklist.

To search the worklist for open tasks related to the import:

1. Select **View > Worklist**, or click the **Worklist** link on the Home page.
2. Use the **Import** filter to display tasks related to records loaded in a specific import. Enter the Import ID or click the Lookup icon  to search for the Import ID by facility, date, and/or user.
3. Some filters are pre-set when you enter the worklist. To ensure that all tasks related to records in this import are listed:

   a. Do not select a value in the task **Type(s)** filter.

   b. Set the **User(s)** filter to blank.

   c. Check **Show Unassigned**.
4. Click **Apply**.
5. Record-based tasks for records loaded in the import will be listed. When a record moves through the workflow and is linked to a patient set, the focus of the task switches from the record to the patient set. These patient set tasks will not be listed in the workflow for this import; use RPT-061A to obtain a complete list of tasks initiated by records in an import.

To review a sample of records loaded in a specific import:

1. You may initiate your search from the Import Manager or the Patient Lookup:

   a. To search from the Import Manager:

      i. Select **System > Import** Manager.

      ii. Click the **Lookup** link listed for the import of interest.

      iii. A Patient Lookup search will be executed using the import's ID as a search field. All other search fields are ignored when an Identifier Search field is used.

   b. To initiate the search from the Patient Lookup:

i. Select **View > Patients**.

ii. Click **ID Lookup** in the left navigation bar.

iii. Use the lookup to specify an **Import ID.**

iv. Click **Search**

2. The search results will be limited to a sample of 500 records from the import.

Reports Related to Data Imports

Select **View > Reports** to review the current list of system reports provided in SEER*DMS. The reports listed below are the system reports related to data imports that were included in SEER*DMS when this manual was printed. If you require information regarding imported data that is not included in a system report, you may generate an external report. See *Chapter 24: Creating Reports and Extracting Data* and *SEER*DMS Data Dictionary* for more information.

Report ID	Title	Description
RPT-009A	Imported Records	Use this report to generate a listing of records in a specific import or a listing of records imported on a specific date.
RPT-009B	Records Submitted by Facility	A listing intended to be sent to the reporting facility showing the records provided by that facility during a specific time period.
RPT-061A	Workflow Location of Records	For each record imported at a specified time, this report lists the current location of the record in the workflow.
RPT-070A	Import Summary	Number of records loaded per imported data file.
RPT-085A	Mass Change Report (Records)	Changes made to records, contacts, or patient sets via mass change Imports. These reports query the audit log tables of records, contacts, and patient sets to identify changes made via direct updates defined in mass change imports.
RPT-085B	Mass Change Report (Contacts)	
RPT-085C	Mass Change Report (Patient Sets)	