

Chapter 1: Introduction to SEER*DMS

The Surveillance, Epidemiology, and End Results (SEER) Program is an initiative sponsored by the National Cancer Institute (NCI) to collect cancer incidence and related information from population-based cancer registries. Participating registries are required to obtain information on all reportable cancers occurring in their area from a variety of hospitals, pathology laboratories, radiology facilities, physician offices, and other facilities. Data collected by the registries are submitted to the NCI and other organizations. This information contributes to national surveillance data on cancer incidence and forms the sole source of population-based estimates of U.S. cancer survival and prevalence.

The SEER Data Management System (SEER*DMS) supports all core functions of a central cancer registry – importing data, editing, linkages, consolidation, and reporting. The NCI, the individual SEER Registries, and Information Management Services (IMS) collaborated on the requirements analysis and design of SEER*DMS. It was determined that despite the variety and complexity of registries' processes, a centrally designed data management system could be employed by central cancer registries. SEER*DMS improves cost efficiency and reduces duplication of effort in terms of system maintenance and administration. Furthermore, the centralized system design and development improves data quality and consistency, increases efficiency, and increases the sharing of knowledge and experience among cancer registries.

This chapter provides an introduction to SEER*DMS and an overview of the system's support of registry operations. It includes a summary of the system's core components with references to relevant chapters in this manual and in other technical documents.

In this chapter, you'll learn about

- SEER*DMS Reference Materials
- System Architecture
- Disaster Recovery for SEER*DMS Databases
- SEER*DMS Database
- Processing Data in SEER*DMS:
 - Loading Data into SEER*DMS
 - Workflow and Worklist
 - Record Reportability
 - Matching and Consolidating Data
 - Visual Editing and Resolving Patient Set Errors
- Geocoding
- SEER*DMS Edits
- Special Studies
- Contact Information for Physicians and Facilities
- Death Clearance
- Tools for Registry Managers:
 - Tools for Case Finding, Active Follow-up, and Follow-back
 - SEER*DMS Dashboard
- Using the Data:
 - Reports and Analysis
 - Data Submissions and Extracts
- Managing System Accounts and Roles
- System Management

SEER*DMS Reference Materials

The SEER*DMS User Manual, several tutorials, and database diagrams are available on the SEER website: <http://seer.cancer.gov/seerdms>. Anyone with Internet access can view these materials. The User Manual provides an in-depth description of the SEER*DMS features and provides step-by-step instructions for using the system. Each of the SEER*DMS tutorials presents a single topic in a more interactive approach using screen shots and examples. The tutorials can be viewed as a screencast with audio; or registry staff can use them as presentations to train new employees.

In addition to the SEER*DMS User Manual, the Help menu within SEER*DMS provides access to:

- Keyboard Shortcuts – the keyboard shortcuts for the current page will be displayed. To see shortcuts that can be used when editing a Patient Set, open a Patient Set and then select **Shortcuts** from the menu.
- Documentation for registry-specific algorithms organized by module (AFL, Auto-build, etc).
- Version History – a detailed list of changes made for each release of SEER*DMS.
- Tech Support (Squish) – a link to the SEER*DMS Technical Support Squish project. This is a web-based issue tracking system.
- SEER*DMS Portal – a link to the SEER*DMS Web Portal, a secure website that enables NCI-SEER, IMS, and registry staff to share information related to the SEER*DMS project. The portal contains registry-specific reference materials; Change Control Board (CCB) meeting agendas, minutes, and related documents; and other information. Access to registry-specific documents is restricted to the single registry and IMS staff. Access to the SEER*DMS portal is restricted to NCI, IMS, and registry staff who are actively involved in the SEER*DMS project.

System Architecture

SEER*DMS allows the user to interact with the registry's database through an interface displayed in a web browser. This design reduces maintenance of individual workstations by providing a mechanism for simultaneously delivering system upgrades to all registry desktops. System requirements and instructions for getting started with SEER*DMS are described in *Chapter 3: Using SEER*DMS*.

The SEER*DMS application is completely comprised of open-source software components and libraries. These include Java/J2EE, Linux, PostgreSQL, Hibernate, Groovy, and JasperReports. The registry's instance of the SEER*DMS application and database are hosted at an IMS computer center. The benefits of this distribution model include:

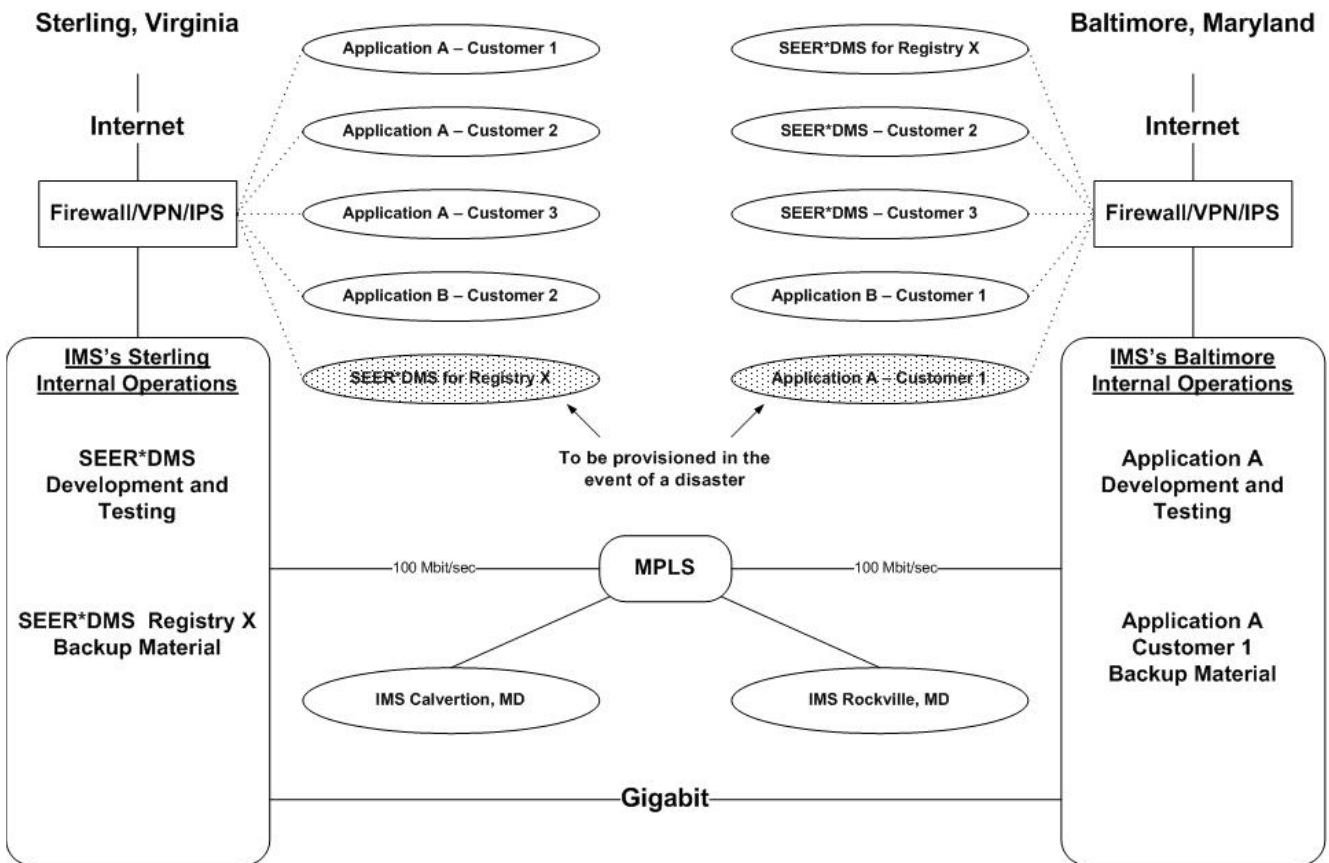
- All operating system and SEER*DMS application updates are managed by the IMS technical support team.
- Database is maintained by IMS administrators.
- No registry responsibility for database backup, system maintenance, or physical security of the database and application servers.
- Designed to provide high up-time and avoid unscheduled interruptions. The IMS technical team is notified in real time of system errors and warnings; enterprise class servers with redundant hardware components eliminate single points of failure.
- Each SEER*DMS hosting customer has a separate network isolated from other registries and isolated from IMS's computer operations.
- Ability to provide robust data restoration capabilities. Redundant data centers are located 60 miles from each other. Production data are copied to a secondary data center multiple times per day to minimize data loss in the event of a disaster.

- Built on open source products such as Linux and PostgreSQL to eliminate licensing fees.

IMS system engineers work with local IT staff to implement and maintain a secure network connection from the registry to the SEER*DMS application and database. This connection is typically implemented using a site-to-site or client-to-site VPN. Figure 1 provides an overview of the system infrastructure for a sample registry's instance of SEER*DMS. The redundant data centers are located in Sterling, VA and Baltimore, MD. This illustration is provided as an example and may not be an exact representation of your registry's configuration.

Figure 1

Overview of Infrastructure for Registry X's Instance of SEER*DMS Hosted at IMS, Inc.



IMS, Inc. – 4/3/2015

Disaster Recovery for SEER*DMS Databases

- Redundant data centers are located 60 miles from each other.
 - Production data are copied to a secondary data center multiple times per day to minimize data loss in the event of a disaster.
 - The full database is copied to the secondary data center once per day. The database transaction logs are copied hourly. The combination of these would be used to restore the database in the event of a disaster that affects the primary data center. This ensures that no more than one hour's worth of data changes would be lost.

- In a typical configuration, the SEER*DMS database snapshots are taken hourly and 128 snapshots are saved. These snapshots are saved in the primary data center and can be used to restore data in the event of data corruption due to human error.

SEER*DMS Database

The SEER*DMS database platform uses PostgreSQL (or Postgres). Postgres is an object-relational database management system (ORDMS). IMS staff have extensive experience and success with large-scale projects using Postgres databases. The benefits of using Postgres include:

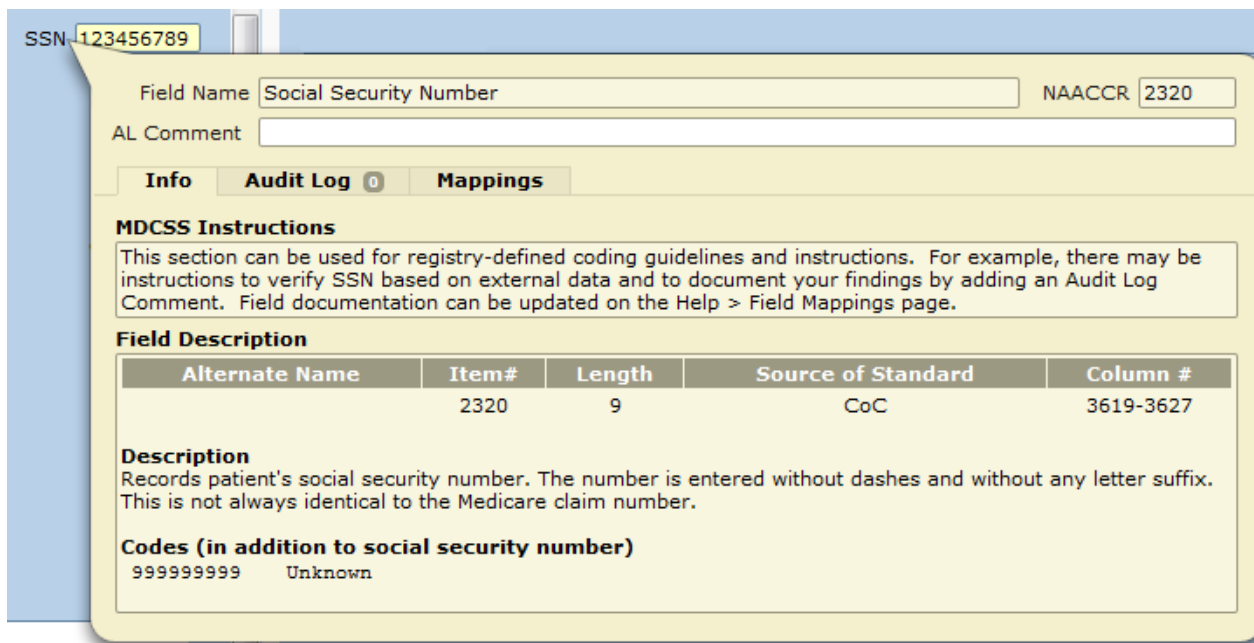
- Postgres is an open-source product; this reduces costs by eliminating licensing fees.
- Postgres is an enterprise class database with advanced features such as point in time recovery, a sophisticated query planner/optimizer, and robust fault tolerance.
- Postgres is also highly compliant to SQL standards and very well-documented, making it readily accessible to registry users. PostgreSQL documentation is available at <http://www.postgresql.org/docs>.

SEER*DMS users are strongly encouraged to use the integrated Data Search to execute queries against the database. Execution of queries via external software is discouraged; external software should only be used if the SEER*DMS Data Search does not provide the functionality required for the project and registry security protocols permit database connections via external software.

Database documentation is included in the Data Search help system. This documentation includes:

- Entity-Relationship (ER) Diagrams (also available at http://seer.cancer.gov/seerdms/db_models/). This is also Appendix A of the Users Manual.
- A brief description of each database table. The table descriptions are available in the database itself, they are stored in the *database_doc* table.
- Sample SQL to illustrate the relationship between commonly used tables.
- A searchable Database Explorer that returns detailed field documentation. This includes:
 - Field documentation from the NAACCR codebook for NAACCR data items.
 - Registry-specific coding instructions for each field.
 - Database fieldname and type.
 - The name of the lookup used by the field, if applicable.
 - Java path for the field (used in edit and mass change syntax).

The field information that is available in the Data Search is also available in the Patient Set and Record Editors. Click the field label to view field documentation, as shown below. This example shows that that NAACCR field documentation and registry-specific coding instructions can be made available for each field in the editor.



The Field Mappings section of the SEER*DMS Help menu is a searchable tool that provides:

- Mapping of database fields to the labels used on SEER*DMS screens.
- Mapping of database fields to NAACCR items. You can search for a field by NAACCR item number, screen label, database field name, or database type.
- The name of the Patient Set screen on which the field is displayed. For record fields, Field Mappings provides a list of record types on which the field is displayed.

A basic overview of the SEER*DMS database structures related to patient data is provided in *Chapter 2: Records and Patient Sets*. Chapter 2 is designed for a non-technical audience and should be considered required reading for editing staff, registry operation managers, and all staff who use registry data (researchers, special study staff, etc). Appendix A of the user's manual contains a more technical view of the database including ER Diagrams and a brief description of each database table. The table descriptions are also stored in the database itself in the *database_doc* table.

Loading Data into SEER*DMS

Data files may be uploaded using the SEER*DMS interface or loaded automatically from a registry network location. The autoloader is designed for "lights out" transfers, that is, fully automated transfers of files. The registry uses an automated script to securely move files from the registry's network to the autoloader on the SEER*DMS server. SEER*DMS automatically loads data files from the main autoloader folder and its sub-folders. The IMS technical team can assist registry staff in implementing the transfer script and configuring the autoloader folders.

In addition, SEER*DMS includes data entry screens that allow a user to enter data using the keyboard. Any record type can be entered via data entry. If the data are available in an image file (PDF, TIF, etc), an Image Data Entry task can be created. Image Data Entry tasks allow the user to view the image in one panel of the screen while they enter values into record fields.

A "record" is created in the SEER*DMS database from the data loaded from files or keyed in a data entry session. Each record represents an individual abstract, case finding report, pathology report, or other type of report containing patient health or follow-up data.

Instructions and technical information are provided in:

- *Chapter 5: Importing Data Files*
- *Chapter 6: Data Entry*

Workflow and Worklist

The workflow controls the flow of data through the system and identifies tasks that must be completed by registry staff. A record entering the system is processed in a series of tasks, and the route of that record through the workflow is controlled by a series of routing scripts. You can view the routing scripts with the proper permissions by clicking **System > Workflow > Task Manager** and selecting the Routing tab. The Help tab displays documentation and possible outcomes of each routing script. The routing scripts are configurable for each registry. Contact IMS to configure your registry's scripts.

SEER*DMS attempts to send the record through the workflow in a series of automatic tasks. Manual tasks may be triggered depending on conditions specified in the routing scripts and edits. For example, a Match task may be triggered if more than one perfect matching Patient Set is found. A Screen for Reportability task may be triggered if the reportability for the incoming record is unknown. Once a person completes a manual task, the record continues to the next automatic task. The record moves from automatic to manual tasks as necessary, until all processes are complete.

The SEER*DMS worklist provides a view of automatic tasks that are currently running and manual tasks that require attention. The manual tasks displayed in the worklist are the staff's "to do list". Users can select tasks or managers can assign tasks to users, depending on registry policy. Registry scripts can be used to automatically assign tasks and to assign priority levels to tasks.

A more detailed description of the SEER*DMS workflow and step-by-step instructions for using the Worklist are provided in *Chapter 4: SEER*DMS Workflow*.

Record Reportability

SEER*DMS uses a combination of automated and manual screening tasks to determine whether a record contains data related to a reportable case. But first, registry-defined record edits are applied to the record to ensure that the record has core data items that are required for screening and other workflow processes. A manual Resolve Record Errors task is created if required data items are missing. The purpose of the Resolve Record Errors task is to allow a person to code missing data items that are essential. The fields required for processing typically include site and histology. The edits that trigger Resolve Record Errors tasks vary by record type and registry. For example, an edit may fail if a coded value for site is not available for an abstract record but the same edit would not fail for a pathology record. In most registries, a very small percentage of records stop in Resolve Record Errors tasks.

The record then moves to the screening portion of the workflow. If possible, an automatic screening task will set the reportability flag. Otherwise, either the standard Screening task or the Pathology Screening task will be created so that a user can set the flag manually. The usage of these tasks varies by registry.

A manual screening task is rarely required for abstracts and other records that contain coded values for site, histology, and behavior. The auto-screener is typically able to set the reportability flag based on the registry's screening algorithm. A manual screening task is only created for an abstract record if the auto-screener could not determine reportability.

In some registries, a manual Pathology Screening task is created for most records containing pathology report data. The Pathology Screening task has a specialized interface that highlights keywords in the text fields in order to facilitate the coding of site, histology, behavior, and grade. The user completing the task also sets the reportability flag for the record.

A description of edits and step-by-step instructions for screening are provided in:

- *Chapter 7: Edit Errors*
- *Chapter 8: Resolving Record Errors*
- *Chapter 9: Screening for Reportability*

Matching and Consolidating Data

A "Patient Set" is a packet of data in the SEER*DMS database which includes all data associated with a particular patient. These data include patient demographics, information on all reportable tumors, admissions information, diagnostic procedures, treatment information, and text documentation. An incoming record for a reportable cancer is matched against the database to determine if it needs to become a new Patient Set or become part of an existing Patient Set. Once the matching process is complete, the record is either consolidated with an existing Patient Set or used to create a new Patient Set.

Information and instructions related to matching and consolidation are provided in:

- *Chapter 10: Matching Records to Existing Data*
- *Chapter 11: The Patient Set Editor*
- *Chapter 12: Consolidating Data*

Visual Editing and Resolving Patient Set Errors

Quality control of data involves the review of edit errors identified by computerized edits; and a visual comparison of coded fields to text (visual editing). Registry managers determine the level of visual editing required in their registry. Some registries require 100% visual editing of newly created cases; other registries only require visual editing if a standard edit is failing; and other registries implement a level of visual editing that falls somewhere in between.

In SEER*DMS, the Resolve Patient Set Errors task is used for visual editing and for resolving edits. A worklist flag can be automatically set to indicate whether the task was created exclusively for visual editing.

Consolidation tasks enable registrars to visually review incoming data as they consolidate it with other data in the system. Failing edits may be resolved in the task, or the Patient Set may be saved with edit errors and forwarded to a Resolve Patient Set Errors task.

Information and instructions related to editing and resolving Patient Set errors are provided in:

- *Chapter 12: Consolidating Data*
- *Chapter 14: Resolving Patient Set Errors*

Geocoding

Geocoding and address standardization are automatically executed whenever an address field is changed. SEER*DMS submits secure API calls to the Automated Geospatial Geocoding Interface Environment (AGGIE) geocoding system. This system was developed through the collaborative efforts of NAACCR, Texas A&M University, and the NCI. Census tract fields, latitude, longitude, and other geospatial variables are set for each valid address. Real time geocoding using the AGGIE system provides high quality data at no cost to NAACCR Full Member Registries.

There are special tools in the SEER*DMS Patient Set editor for consolidating and editing addresses, as described in *Chapter 11: The Patient Set Editor*. These include a tool to compare addresses received on source records; and a popup that displays results from AGGIE. Detailed results from AGGIE are available in XML or CSV files.

SEER*DMS Edits

Patient data are validated against standard edit sets and edits written by registry staff. The SEER*DMS Edits Manager is a robust tool for writing, testing, and managing edits.

- SEER, NPCR, NCDB, and NAACCR Call for Data edits are integrated and updated in a timely manner.
- The SEER*DMS Edits manager allows the registry to deactivate edits that are not needed by the registry.
- The coder can review a list of failing edits as they view and code a case. The edits are executed each time a user opens, validates, or saves data.
- The Patient Set Edits System task can be used to execute edits on all Patient Sets or a cohort of Patient Sets. The primary purpose of this task is to apply new or modified edits to all Patient Sets in the database.

For more information regarding system edits, see *Chapter 7: Edit Errors*.

Special Studies

In SEER*DMS, "Special Studies" consist of user-defined criteria to determine the eligibility of incoming records for the purpose of rapid case ascertainment. The criteria specified for a Special Study are applied to the record as it moves through the workflow's auto-screening task. The criteria may define a record as "eligible for the study", "ineligible", or "possible data for the study". A manual Review Special Studies task is created if the record is identified as a "possible" study case that requires review. In addition, you may assign records or Patient Sets to a special study via mass change imports or manually assign data to a study using the SEER*DMS editor.

Information and instructions related to Special Studies are provided in *Chapter 28: Special Studies*.

Contact Information for Physicians and Facilities

SEER*DMS maintains information about organizations and facilities that collaborate with the cancer registry and contact information for physicians and representatives of these organizations. The organization and contact lists are referenced throughout SEER*DMS when selecting values for data fields, tracking data and follow-back sources, and managing abstracting assignments.

Information and instructions related to these topics are provided in:

- *Chapter 18: The Facility List*
- *Chapter 19: The Contact List*

Death Clearance

Death certificate records provide a valuable source of follow-up data for cases that are known to the registry and they provide a case finding source for new potential cases. Death certificates with non-cancer causes of death provide valuable data for cases in the registry database (cause of death, date of death, place of death, etc.) The majority of these “non-reportable” death certificate records are processed in automatic tasks; a small percentage require manual review.

A death certificate is considered “reportable” if a cause of death is a disease of interest to the registry. A reportable death certificate is automatically linked at the cancer-level (CTC) if it meets the registry’s death certificate matching and consolidation rules. A standard set of rules, based on the NAACCR Death Clearance Committee guidelines, was developed by the SEER*DMS Change Control Board and adopted by all SEER*DMS registries. Minor modifications to the rules were implemented for some registries, as requested.

When a reportable death certificate does not match CTC data, case finding protocols are implemented by the registry. The SEER*DMS Death Clearance Manager provides tools to support the registry’s death clearance processes. Information and instructions related to Death Clearance are provided in *Chapter 17: Death Clearance*.

Tools for Case Finding, Active Follow-up, and Follow-back

SEER*DMS includes manager pages specifically designed to manage active follow-up activities, monitor and assign case finding leads, and process requests for follow-back information.

Information and instructions related to these topics are provided in:

- *Chapter 16: Follow-up*
- *Chapter 21: Managing Abstracting Assignments*
- *Chapter 22: Follow-back*

SEER*DMS Dashboard

The SEER*DMS dashboard provides an at-a-glance view of the state of the registry. The data are refreshed each night or can be refreshed manually. Six components are currently included in the dashboard: the SEER Data Quality Profile (DQP), the Number of Reportable CTCs by Year of Diagnosis, Edits, Completeness of the Data Submitted by Facilities, Productivity of Editing Staff, and Records Loaded by Year.

The numerator, denominator, result, and goal is shown for each marker on the DQP page. The label in the Data Quality Marker column is a link to the Data Search. The registry manager can use that link to find a list of cases failing that marker. For example, the manager could click the Follow-up < 20 link to get a list of pediatric patients who are lost to follow-up.

SEER Data Quality Profile						
November 2015 Submission						
(Reporting Year: 2013)						
updated 9h ago refresh						
#	Data Quality Marker	Criteria	Num	Denom	Goal	Result Met Goal
1	Death Certificate only	2013 cases only			< 1.5% and > 0.0%	0.35%
2	COD missing for patients known to be dead	Deceased patients coded as 7797 or 7777 or blank; 1995-2013 cases			< 2.5%	1.37%
3	Unknown or ill-defined primary site	2013 cases coded C76 or C80; DCOs excluded			< 2.5%	0.95%
4	Non-specific histology	2013 cases coded 8000 or 8001; DCOs excluded			< 2.5%	0.89%
5	Unknown laterality	2013 cases coded as 3 or 9; DCOs excluded			< 6.0%	2.55%
6	Unknown historic stage	2013 cases mapped to 9 in historic stage. Pleura, brain, lymphoma, KS, misc, benign/borderline, and DCOs excluded			< 10.0%	2.05%
7	Invalid or missing census tract	Invalid census tract in designated county or unknown county			< 2.0%	0.08%
8	Completeness Estimate (2013 Cases)	From Data Quality Report prepared by IMS Last 10 Dx Years used if available			≥ 98.0%	95.23%
9	Completeness Estimate (2014 Cases)	From Data Quality Report prepared by IMS Last 10 Dx Years used if available			≥ 95.0% in Feb 2016	49.31%
10	1 year reporting delay	Difference between number of cases reported for 2012 in Nov 2014 and Nov 2015 submissions			< 2.5%	
11	Follow-up age <20	Invasive cases, 1995-2012 followed into 2013			≥ 90%*; ≥ 80%**	92.12%
12	Follow-up age 20-64	Invasive cases, 1995-2012 followed into 2013			≥ 90%*; ≥ 80%**	95.23%
13	Follow-up age 65+	Invasive cases, 1995-2012 followed into 2013			≥ 95%*; ≥ 90%**	97.42%
14	Follow-up all in situs	In situ cases, 1995-2012 followed into 2013			≥ 90%*; ≥ 80%**	92.04%
15	Urban Geocoding Accuracy	2013 cases coded 1 or 2 for census tract certainty in county with > 75% of population in urban area; DCOs excluded			≥ 90%	
16	Failing SEER Edits	Cases diagnosed 1973-2013 must be edited prior to submission using the SEER edits		0		4
						Total 11

* Contractual standard
** Minimum acceptable

A dashboard summary is displayed on the home page if the user has the appropriate system permissions. Current statistics related to SEER completeness are displayed if the user has the system permission for the DQP dashboard. Summarized counts of failing edits are shown on the home page if the user has the system permission for the Edits dashboard.

Reports and Analysis

SEER*DMS includes a variety of integrated reports which summarize registry activities, track data through the system, and provide quality control and completeness metrics. Research and technical staff can write ad hoc queries using the SEER*DMS Data Search. The Data Search provides an interface to define complex search criteria based on Boolean expressions and to execute SQL statements. Data Search results can be written to extract files for analysis in external software.

Information and instructions related to integrated reports and extracts are provided in:

- *Chapter 20: Searching for Records and Patients*
- *Chapter 24: Creating Reports and Extracting Data*

Data Submissions and Extracts

SEER*DMS provides a push button mechanism for creating data submission files required by standard setters such as NAACCR, NPCR, and SEER. Data requirements are reviewed and updated each year. Ad hoc extracts can also be created in the NAACCR file format for any subset of cases. The confidential, full case, and incidence record layouts are supported. In addition, SEER*DMS includes system extracts for standard linkages such as NDI and SSA.

The extract module includes a powerful interface to create custom extracts. Custom extracts are typically used to make modified versions of standard extracts, for example, a custom NAACCR extract could be created that uses coded data items from consolidated (CTC) fields and text from the original source records. But the custom extract module is much more powerful than that. It allows the registry to create fixed column data files containing any data in the database. Multiple data files can be created at a time; SQL and the Groovy scripting language can be used to extend the functionality. In SEER*DMS version 17, the custom extract module will support parameters.

Information and instructions related to integrated reports and extracts are provided in *Chapter 24: Creating Reports and Extracting Data*.

Managing System Accounts and Roles

Registry managers and/or system administrators are responsible for creating SEER*DMS user accounts and enabling system permissions by assigning roles to those accounts.

Information and instructions related to these topics are provided in:

- *Chapter 25: Managing User Accounts*
- *Chapter 26: System Roles and Permissions*

System Management

SEER*DMS uses an HTML-based interface displayed via a web browser. The registry's information technology (IT) staff are responsible for providing on-site technical support for workstations and 3rd party software used with SEER*DMS. A current web browser is the only software required of all users (specifications are provided in *Chapter 3: Using SEER*DMS*). VPN client software may be required in some registries and, in those registries, VPN software may not be required for all staff.

Registry staff may be required to provide ad hoc programming support. In some registries, this support is provided by research staff and in other registries it is provided by IT staff. The extent to which ad hoc programming is required varies by registry; the tasks may include writing SQL, writing scripts or programs to manage data files, and generating registry-specific reports. The IMS support team will provide assistance, but cannot provide large-scale support for these activities. An overview of the duties of registry staff is provided in *Chapter 27: System Management*.