Increasing Automation in SEER*DMS

SEER*DMS Face-to-Face Meeting
July 13, 2017

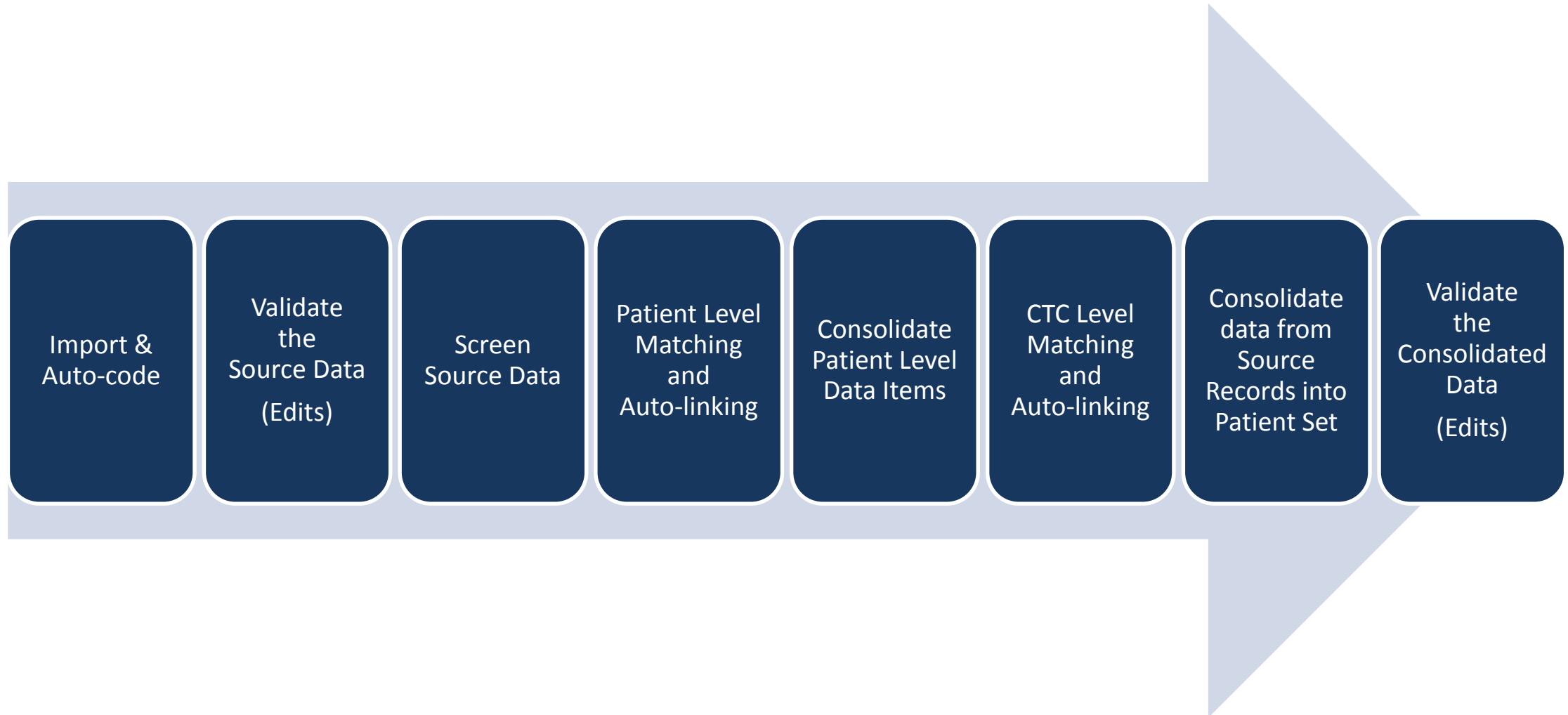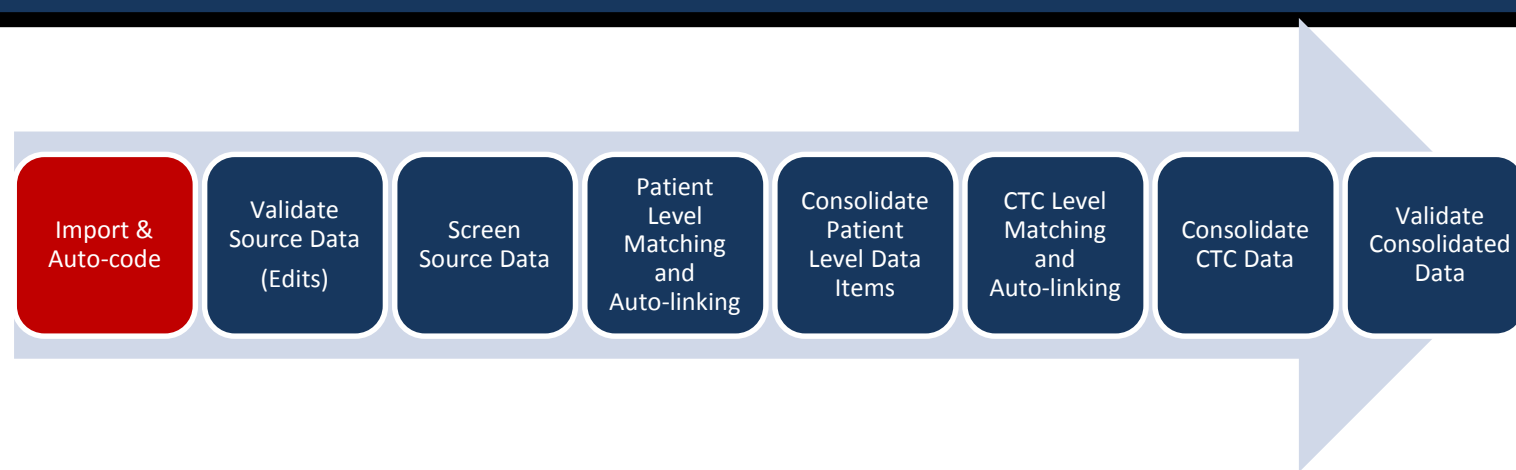Linda Coyle

Information Management Services, Inc.

Increasing Automation in SEER*DMS

- In this afternoon's session, speakers will discuss automating the processing of these data submitted to the registry:

    - Source NAACCR Abstracts

    - Pathology reports

    - Insurance claims

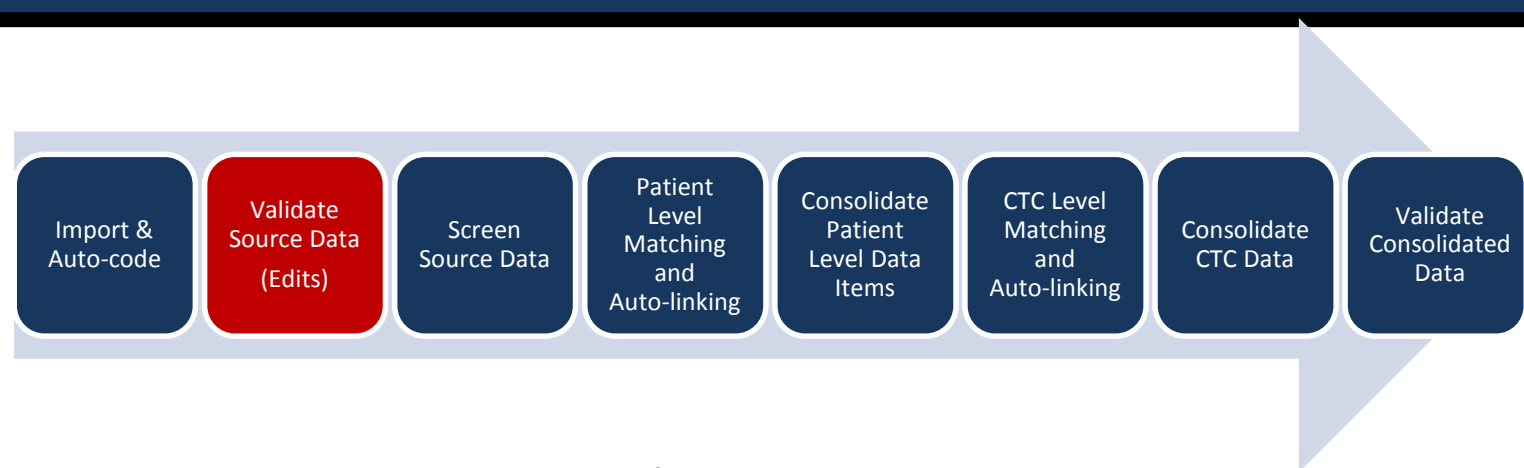   - MU2 Electronic Health Records (EHR)

Increasing Automation in SEER*DMS
SEER*DMS Workflow

Import & Auto-code | Validate Source Data (Edits) | Screen Source Data | Patient Level Matching and Auto-linking | Consolidate Patient Level Data Items | CTC Level Matching and Auto-linking | Consolidate CTC Data | Validate Consolidated Data
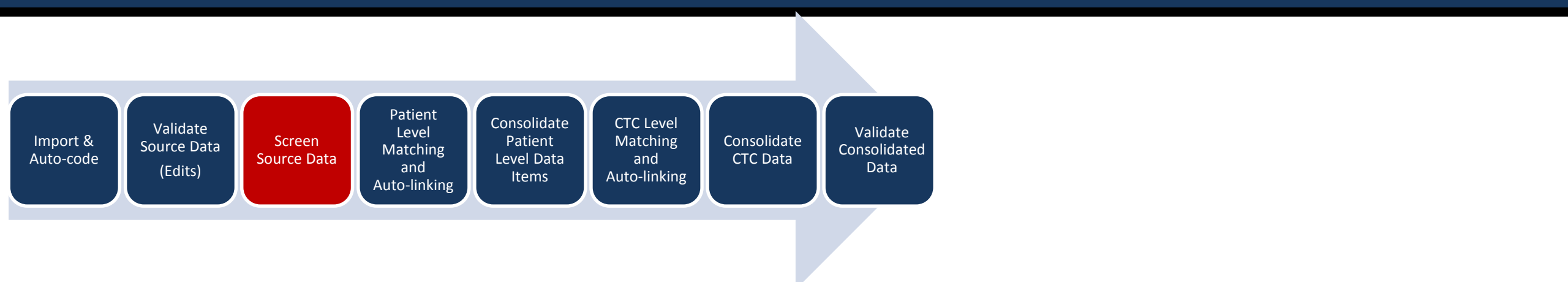
- Importing is automated process – except for small % imports when a file fails validation of the file structure.   For example, expect 22,824 characters for a NAACCR Abstract file.

- Auto-coding:
  - Fully automated (obviously)
  - Affects automation of later steps.
    - Example: Converting claims disease codes from ICD-10 to ICD-O-3 facilitates claim to CTC matching.
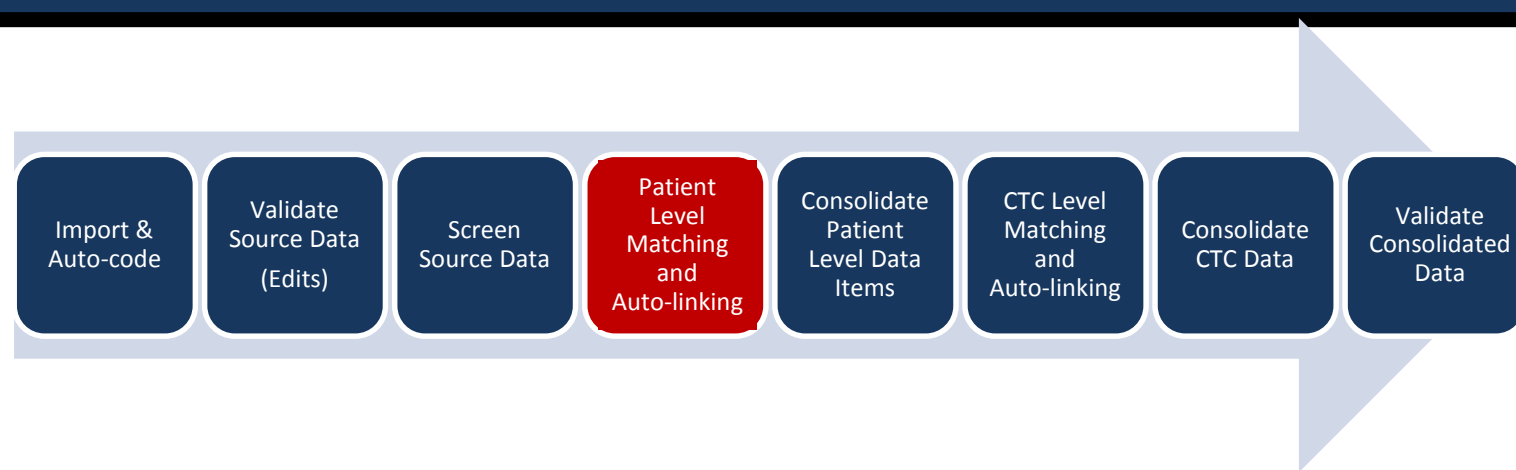
Import & Auto-code → **Validate Source Data (Edits)** → Screen Source Data → Patient Level Matching and Auto-linking → Consolidate Patient Level Data Items → CTC Level Matching and Auto-linking → Consolidate CTC Data → Validate Consolidated Data

- Current SEER*DMS workflow:
  - Minimal changes are made to the source abstract records
  - CTRs edit the consolidated data (CTC); records rarely changed

- Prerequisite for auto-consolidation:
  - Edits/visual review – source records should be complete and accurate
  - Auto-consolidation workgroup will be making recommendations for changing the validation of source data.
  - The processes to validate source data may happen in the case upload tool and/or the standard SEER*DMS workflow

Changes Needed ??

| Import & Auto-code | Validate Source Data (Edits) | Screen Source Data | Patient Level Matching and Auto-linking | Consolidate Patient Level Data Items | CTC Level Matching and Auto-linking | Consolidate CTC Data | Validate Consolidated Data |

- Screening is a process to determine if a record is of value to the registry.

- Screening Source Abstracts
  - 99% of abstracts are auto-screened
  - Algorithms are based on data items that are well coded on an abstract:
    - Site, Histology, Behavior, Year of Diagnosis, Address at Diagnosis

- Screening Path Reports
  - A large number of path reports are manually screened by SEER*DMS registries. 312 thousand path screening tasks were completed by SEER*DMS registries in 2016.
  - NLP algorithms are being developed to increase the automation of path screening.

- Screening of Claims and MU2 Electronic Health Records (EHR)
  - Not being screened in initial release. All claims and EHR data are retained. Screening algorithms may be considered later, if needed.

- Deterministic matching algorithms are used for patient level matching in SEER*DMS

- Approximately 60 criteria are evaluated

- Registries share a base set of criteria; registry-specific criteria are also supported

- Scores above a defined threshold trigger auto-linkage – the record is linked to the Patient Set (at the "patient" level)

- Threshold varies by data type (eg, lower threshold is used for claims linkage)

| Import & Auto-code | Validate Source Data (Edits) | Screen Source Data | Patient Level Matching and Auto-linking | Consolidate Patient Level Data Items | CTC Level Matching and Auto-linking | Consolidate CTC Data | Validate Consolidated Data |

Examples of patient level matching criteria in SEER*DMS

Record Linkage: Primary Algorithm for Automatic Match Task   linkage

Algorithm to match incoming record against database for linkage.

Matched Against

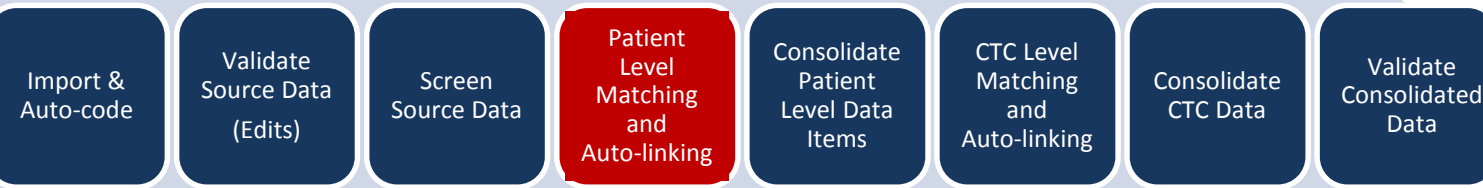All Patient Sets that are not flagged as deleted.

Unlinked Records. Supplemental records are matched against Patient Sets only.

All claims (if supported by the registry).

Matches

Score	ID	Description
2000	MB001	SSN, Date of Birth, Last Name (Patient or Maiden or Alias), First Name (Patient or Alias or Nickname)
1990	MB002	SSN, Date of Birth, Last Name (Patient or Maiden or Alias)
1980	MB003	SSN, Date of Birth, First Name (Patient or Alias or Nickname)
1970	MB004	SSN, Date of Birth, First Name Initial (Patient or Alias), Sex
1960	MB005	SSN, Year and Month of Birth, Sex
1950	MB006	SSN, Year and Day of Birth, Sex
1940	MB007	SSN, Year of Birth, First Name Initial (Patient or Alias), Sex
1930	MB008	SSN, Date of Birth (Two Parts), Last Name (Patient or Maiden or Alias), First Name (Patient or Alias or Nickname)
1910	MB009	SSN, five first characters of Last Name (Patient or Maiden or Alias), three first characters of First Name (Patient or Alias), Sex
1900	MB010	SSN, five first characters of Last Name (Patient or Maiden or Alias), First Name Initial (Patient or Alias), Middle Name Initial, Sex
1880	MB011	SSN, Maiden Name, First Name (Patient or Alias or Nickname)
1870	MB012	SSN, two first and last characters of Last Name (Patient or Maiden or Alias), two first and last characters of First Name (Patient or Alias)
1850	MB014	SSN (8 digits match), Date of Birth, three first characters of Last Name (Patient or Maiden or Alias), First Name Initial (Patient or Alias), Sex
1840	MB063	Date of Birth, Last Name (Patient or Maiden or Alias), First Name (Patient or Alias or Nickname), Middle Name Initial, Sex
999	MB050	SSN, Last Name (Patient or Maiden or Alias)
998	MB051	SSN, First Name (Patient or Alias or Nickname)
997	MB052	SSN, Date of Birth
996	MB066	Date of Birth, Last Name (Patient or Maiden or Alias), First Name (Patient or Alias), Street Number (Current or DX), Street Name (Current or DX)
991	MB064	Date of Birth, Last Name (Patient or Maiden or Alias), First Name (Trigram) (Patient or Alias), Street Number (Current or DX), Street Name (Current or DX)
990	MB065	Date of Birth, Last Name (Trigram) (Patient or Maiden or Alias), First Name (Patient or Alias), Street Number (Current or DX), Street Name (Current or DX)
989	MB060	Date of Birth, Last Name (Patient or Maiden or Alias), three first characters of First Name (Patient or Alias), Middle Name Initial, Sex
987	MB025	Date of Birth, five first characters of Last Name (Patient or Maiden or Alias), three first characters of First Name (Patient or Alias), Middle Name Initial, Sex

| Import & Auto-code | Validate Source Data (Edits) | Screen Source Data | Patient Level Matching and Auto-linking | Consolidate Patient Level Data Items | CTC Level Matching and Auto-linking | Consolidate CTC Data | Validate Consolidated Data |

Opportunity

- % of match tasks that are completed automatically:
  - Abstracts: 87% (four registries combined). Percentages for individual registries: 82, 87, 90, 93.
  - Path Reports: 84%
  - Claims: 100%
  - EHR: 100%
- Validating Patient Level Matching
  - IMS staff execute a system task to execute matching algorithms across database.
  - Temporary table is created that can be used to evaluate matching criteria
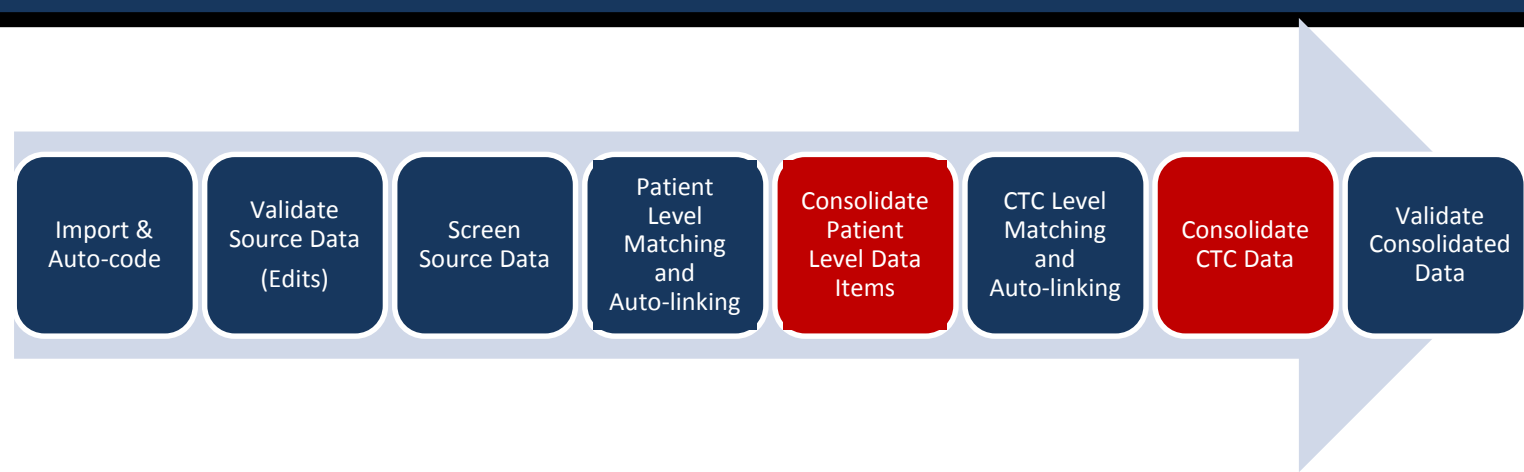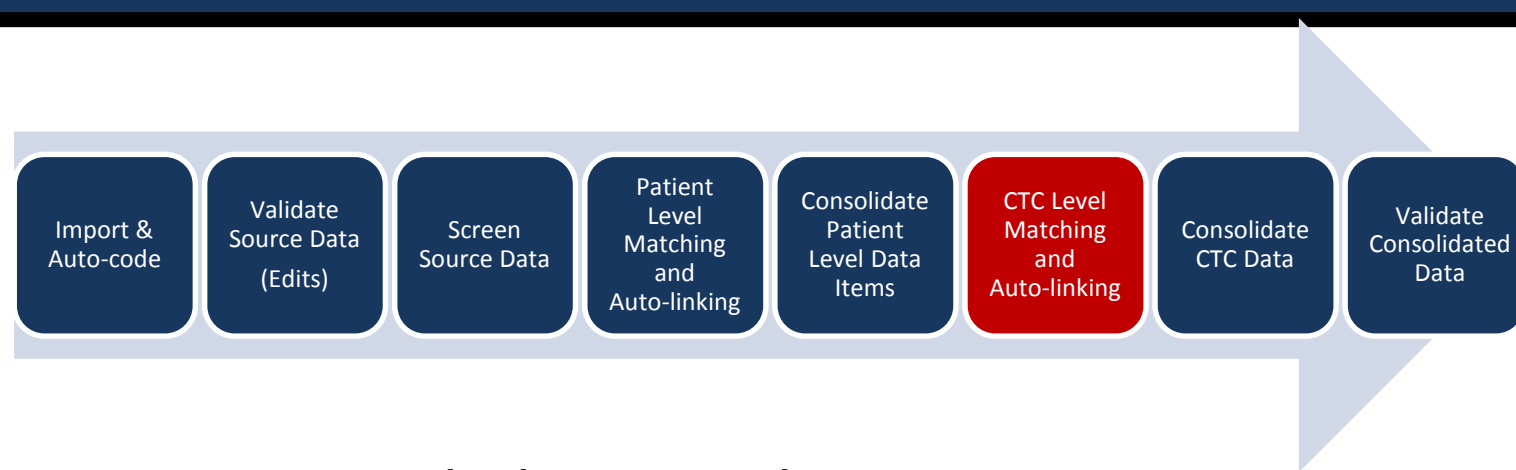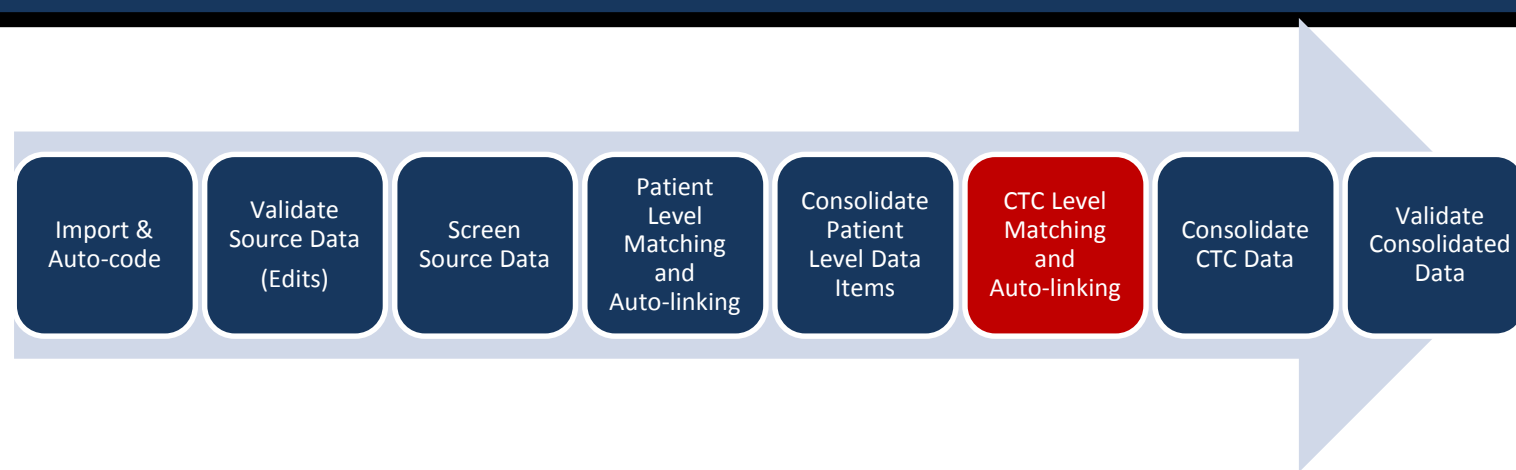
- SEER*DMS supports auto-consolidation, but additional auto-consolidation rules are needed

- SEER*DMS registries auto-consolidate demographic level data items; and treatment data items

- The Auto-consolidation Workgroup will be defining rules for diagnostic information; staging data items; additional demographic data items
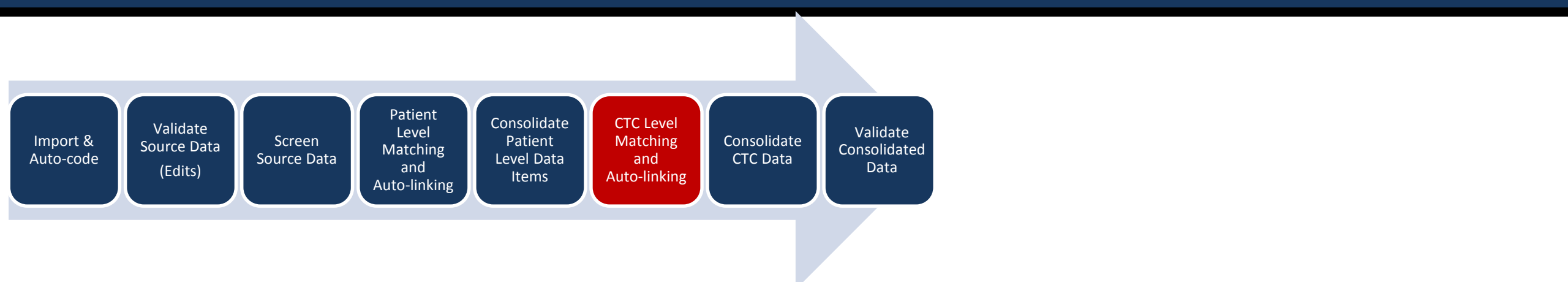
- SEER Multiple Primary Rules are used for tumor or "CTC" level matching of Abstracts, Path Reports, Claims, EHR

- "Strict" and "Lenient" implementation of MP rules:
  - Site and histology from claims are set by converting ICD-10 to ICD-O-3
  - Claims data are less likely to have a specific histology code; but the CTC is likely to have a specific histology
  - Lenient implementation is used for claims data – NOS histology codes are considered a match to more specific values

| Import & Auto-code | Validate Source Data (Edits) | Screen Source Data | Patient Level Matching and Auto-linking | Consolidate Patient Level Data Items | CTC Level Matching and Auto-linking | Consolidate CTC Data | Validate Consolidated Data |

- Auto-linking workgroup:
  - IMS staff have been working with registry data to test auto-linking abstracts at the CTC level
  - IMS testing identified and corrected differences in auto-linking vs manual linking:
    - V17.50 update – rules for auto-building a TX page are applied when a record is auto-consolidated. This allows registries to auto-link abstracts to CTCs.
    - V17.60 – Changes were made to use the same processes to de-duplicate TX pages.

Import & Auto-code → Validate Source Data (Edits) → Screen Source Data → Patient Level Matching and Auto-linking → Consolidate Patient Level Data Items → **CTC Level Matching and Auto-linking** → Consolidate CTC Data → Validate Consolidated Data

- Auto-linking workgroup – next steps:
  - Review data processed on Detroit, New Jersey, Utah test servers
  - Test with additional registries
  - Meet as a group to discuss:
    - Possible extensions to MP rules (eg, consider a record to be a match if a similar record was manually matched to a CTC)
    - Matches that require manual review (eg, hematopoietic diseases)
    - Auto-consolidation requirements for records that match CTCs with year dx < 2014

Import & Auto-code → Validate Source Data (Edits) → Screen Source Data → Patient Level Matching and Auto-linking → Consolidate Patient Level Data Items → CTC Level Matching and Auto-linking → Consolidate CTC Data → Validate Consolidated Data
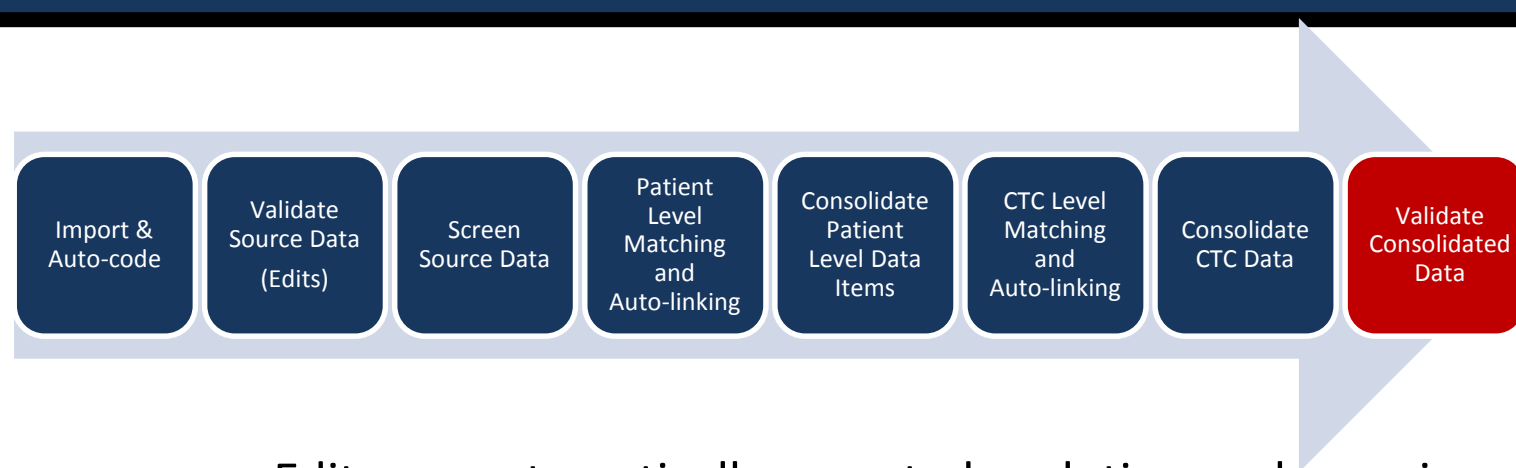
- Edits are automatically executed each time a change is made to patient data. CTRs resolve most edits during manual consolidation or visual editing tasks.

- SEER Edits
  - Same version of edits are maintained in SEER*Edits and SEER*DMS
  - Shared source code

- EDITS metafiles
  - Timely deployment of edits from all standard setters
  - Automated process to translate the metafile edits into a language that is compatible with Java
  - Translated edits are part of the SEER*Utils Java library and used by SEER*DMS and SEER*Abs

- SEER*DMS edit manager allows registries to:
  - Set edits as active or inactive
  - Define conditions for edits (eg, Dx Year = 2016 or later)
  - Write, test, and maintain registry edits