

# Investigation of Current State, Local, and Tribal Facility Inventory Data Sources and Data Flows

## 1 Overview

The goal of this report is to explore how facility data workflows would have to happen between CAERS and State, Local, and Tribal authorities (SLTs), so that CAERS may be developed to accommodate such SLT workflows with as much flexibility as possible.

Previously, four use cases of SLT workflows had been identified at a conceptual level, depending on whether an SLT would prefer to keep some or all their current SLT interface and backend. These cases are:

Case 1: the SLT interface and backend are retained (CAERS receives data from SLT system) to distribute to the federal programs.

Case 2: SLT interface and backend are retained (CAERS receives data from facilities and pushes data to state interfaces).

Case 3: CAERS replaces SLT interface, but state databases are retained.

Case 4: SLT uses CAERS for reporting instead of their previous reporting system or method.

The current SLTs using CAERS (GA, Washington D.C., Pima County AZ, and Rhode Island) are all considered to fall under Cases 3 or 4.

However, details about how facility data are updated and maintained by SLTs require research because, regardless of the case the SLT falls under, SLTs may choose not to use CAERS as the primary repository of their *facility inventory data*. Furthermore, there may be additional data fields that SLTs require for their facility inventory, that would need to be accommodated in CAERS to enable SLTs to take full advantage of combined air emissions reporting. Overall, while SLTs may focus on the uses of their facility inventory to meet their own needs (permitting compliance, billing, emissions inventory analysis), the ability to share the same facility inventory with the federal programs via CAERS, would keep emissions inventory data reporting at the facility and sub-facility levels consistent across programs. This means that each program would have a similar version of a facility in terms of facility and sub-facility data, instead of slightly different version of the same facility which happens currently. Also, reporters wouldn't have to report their emissions data associated to different versions of the same facility and sub-facility components to each program. Facility and sub-facility components shared amongst SLT and federal programs would ultimately provide data analysts, inventory developers, and modelers, the ability to conduct multipollutant analyses across several different types of pollutants; hazardous air pollutants (HAPS)/toxics, criteria air pollutants (CAPs), and greenhouse gases (GHGs).

To ensure that CAERS can accommodate specific facility data workflows, so that more SLTs may use CAERS, the Facility Research & Development (R&D) Team conducted a research project to find out what

data sources and data flows SLTs use currently for obtaining and updating facility inventory information for their respective emission inventories. The results will help inform how CAERS would need to be enhanced to send facility inventory data to and receive facility inventory data from SLTs.

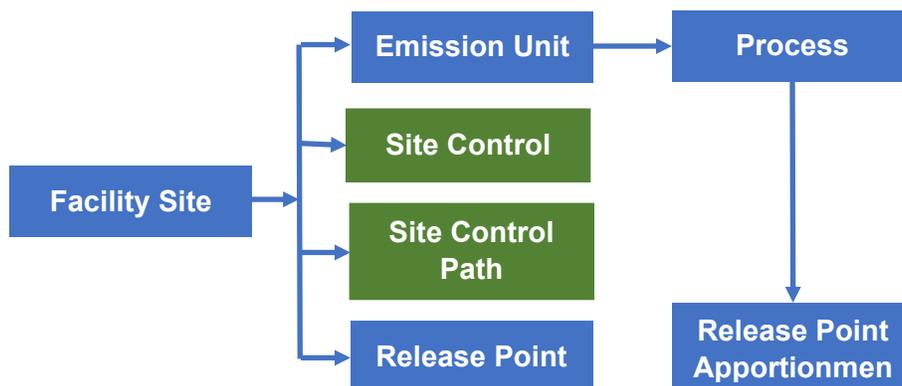
Enhancements in CAERS could include both workflows: the ability to send data to and from a given SLT’s system, additional SLT-specific data fields that might need to be included in CAERS, as well as potential data transformations which could be done by the SLT before sending data to CAERS, or as part of an SLT CAERS module. An SLT CAERS module refers to a set of coding that would be added in CAERS or as part of an SLTs workflow with CAERS, to meet SLT specific requirements (that other SLTs don’t need), but that also does not require a rebuild of CAERS.

## 2 Method

The R&D team gathered input about information needed to understand SLT facility inventory data workflows from team members and the PDT. State members in the R&D team designed and finalized a questionnaire based on that input, and ECOS sent the survey out to SLT contacts who report to the National Emissions Inventory (NEI), in April of 2022. The questionnaire referenced the new Emissions Inventory System (EIS) [Consolidated Emissions Reporting Schema \(CERS\) v2.0](#), to ensure that questionnaire respondents were familiar with the data fields and definitions mentioned in the survey.

CERS is used by the Emissions Inventory System (EIS) to provide a standardized structure for delegated authorities to submit emissions inventory data to the United States Environmental Protection Agency (EPA) to meet the [Air Emissions Reporting Rule \(AERR\)](#) requirements. The questionnaire used simplified data components (Figure 1) to cover all [information specified in CERS](#) for facility inventory that is currently being collected.

Figure 1. Simplified Representation of CERS Facility and Sub-facility Inventory Components Used in The Questionnaire



The online questionnaire used a multiple-choice format (see Appendix A). It was sent to SLT point source contacts via e-mail on 4/5/2022. Two follow-up reminders were sent out on 4/18/2022 and 4/27/2022, respectively. ECOS received 55 jurisdictional responses out of 78 jurisdictions that were contacted, and one incomplete response had to be excluded, leaving 54 valid responses (a 69% response rate). Table 1 summarizes the number of SLTs responses to the questionnaire.

Table 1. Number of Questionnaire Responses by Respondent Type

Respondent Type	Local	State	Tribal	Incomplete	Total	Valid
<b>Individuals Contacted</b>	29	76	2		107	107
<b>Jurisdictions Contacted</b>	23	53	2		78	78
<b>Individuals Responded</b>	16	41	1	1	59	58
<b>Jurisdictions Responded</b>	16	37	1	1	55	<b>54</b>

To clarify the responses further, SLT team members sent follow-up emails to 18 SLTs. Data clean-up was also performed to make responses consistent. For example, one responder did not select “State permitting program” in answering the question regarding information sources for facility sites. It indicated the “other” choice and entered “Permittees submit annual throughput info to the Permitting Section, which calculates the emissions and charges the Permittees an annual emission fee. The Emissions Inventory (EI) group uses the calculated emissions from the Permitting calculation sheets.” This implies the jurisdiction obtains facility site information from a permitting program. Therefore, the choice of “State permitting program” was considered as the actual response. Another example of data clean-up is an initial indication in the “other” choice of the use of “Inspections”, which means that the State compliance program was the right choice. In this case, the choice of State compliance program was considered as the response.

There were 4 states for which multiple people from the state responded to the questionnaire with inconsistent answers. These responders were contacted, and the correct answers were confirmed and adjusted accordingly to represent one complete answer for their jurisdiction.

### 3 Analysis of Questionnaire Responses

#### 3.1 Type of Facilities in Emission Inventories

All 54 valid responses indicated the inclusion of facilities with Title V permits in their emission inventory, while 43 also included facilities with other types of permits (such as minors, synthetic minor, etc.). Of the responses, 14 jurisdictions indicated they have non-permitted facilities in their point sources emission inventory. Those non-permitted facilities include greenhouse gas (GHG) emission facilities that are not subject to reporting criteria air pollutants (CAPs) by EPA’s Air Emission Report Requirements (AERR) and/or SLT requirements, facilities subject to other state requirements, TRI facilities, small facilities with locational coordinates to support reporting for EPA nonpoint source emission inventory requirements, such as landfills, and publicly owned treatment works (POTWs).

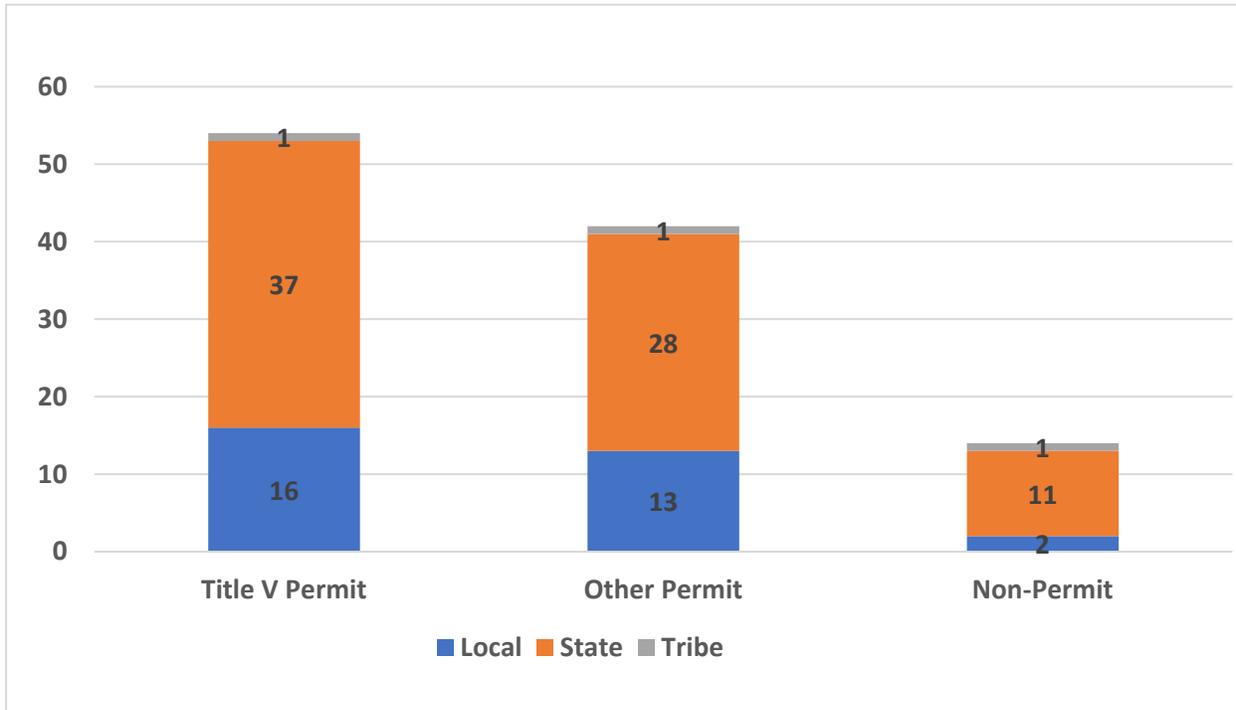
Table 2 and Figure 2 shows a further breakdown of types of facilities handled for emission inventories by different jurisdictions.

*This clearly indicates CAERs could be used by SLTs not only for facilities that are subject to federal-reporting requirements but also for non-federal reporting requirements in use cases 2-4 of SLT workflows. Otherwise, the SLTs would need to have one reporting system for facilities that also report to one or more federal programs, and a separate reporting system for those that don’t, and this may or may not be preferred by SLTs.*

Table 2. Types of Facilities by Jurisdiction

	Title V Permit	Other Permit	Non-Permit
<b>Local</b>	16	13	2
<b>State</b>	37	28	11
<b>Tribal</b>	1	1	1
<b>Total</b>	54	42	14

Figure 2. Number of SLTs with Facilities in Point Source Inventories by Permit Type



### 3.2 Number of Facilities in Emission Inventories

The total number of unique facilities within a given SLT point source emission inventory ranges from tens-of-thousands (for example, 30,000 in Colorado and 28,000 in Wyoming) to less than 10 (5 in Knox County, Tennessee). From the data collected, the median number of facilities in SLT point source inventories is 271.

SLTs may not report to the NEI all the facilities for which they collect emissions data. The total number of facilities reporting to the NEI by SLTs could be as many as 28,000 in Wyoming; and as few as 5 in Knox County, TN. The median number of facilities reported to the NEI by SLTs is 149.

*This information speaks to the need for CAERS to have the capacity to hold large facility inventories for SLTs who may need them and ensure that speed and system performance is not slowed by large facility inventories.*

### 3.3 Data in Tribal Emission Inventories

The U.S. government officially recognizes 574 Indian tribes in the contiguous 48 states and Alaska. Although each tribe is different, most tribes do not have their own emission inventory and do not report to the NEI, as many are not required to report. Tribes with Treatment as a State (TAS) for emissions inventories are subject to AERR reporting, and tribes can voluntarily accept this responsibility in coordination with states and EPA. Eight tribes reported to the 2017 NEI. Table 3 provides details.

Table 3. Number of Facilities Reported by Tribes for the 2017 NEI.

Tribal Name	EIS Code	Program System Code	Number of Facilities
<b>Coeur d Alene Tribe of the Coeur d Alene Reservation, Idaho</b>	88181	TR181	2
<b>Confederated Tribes and Bands of the Yakama Nation, Washington</b>	88124	TR124	28
<b>Nez Perce Tribe of Idaho</b>	88182	TR182	1
<b>Northern Cheyenne Tribe of the Northern Cheyenne Indian Reservation, Montana</b>	88207	TR207	1
<b>Salt River Pima-Maricopa Indian Community of the Salt River Reservation, Arizona</b>	88615	TR615	16
<b>Shoshone-Bannock Tribes of the Fort Hall Reservation of Idaho</b>	88180	TR180	1
<b>Southern Ute Indian Tribe</b>	88750	TR750	293
<b>Ute Mountain Tribe of the Ute Mountain Reservation, Colorado, New Mexico &amp; Utah</b>	88751	TR751	1
<b>Grand Total</b>			343

The Southern Ute Indian Tribe reported the most point source data to the 2017 NEI with 293 facilities. This tribe responded to the questionnaire, and the results of this report reflect that response.

*It is expected that the CAERS could facilitate emission inventory development by tribes by providing tribes a way to collect point source data, because in most cases, there are similarities between tribal emission inventories and State and Local (SLs) emission inventories. By enhancing how CAERS handles facility inventory data, tribes might find the use of CAERS more advantageous and could be encouraged to report more so than they do now.*

### 3.4 Facility Inventory Information in Emission Inventories

Twenty-eight SLTs do not collect emission inventory (EI) information beyond what is identified in the CERS 2.0, while 26 SLTs collect EI information beyond that found in CERS 2.0. A variety of additional information collected depending on special SLT needs. Examples are:

- Idaho collects extra geographic data for the facility and release points, design capacity for all emission units (versus the subset required to do so for NEI purposes), some extra control fields, and a Federal Employer ID to support emission fee billing.
- Massachusetts collects information on potential-to-emit and permitted limits for throughput and emissions, detailed information on emission units (especially emergency generators), monitors, and controls.
- Minnesota collects information on permit type, source type, comment fields for each data component, and NAICS for emission units, whereas the federal program requires only facility level NAICS. 'Also, the CERS requires information on one field for the percent reduction achieved for a pollutant when all control measures are operating as designed for a site control measure. Minnesota collects information on two fields for a pollutant controlled by a site control measure: control efficiency and capture efficiency.
- Wyoming collects permit limits at the unit-process level, U.S. Well Numbers for wellhead facilities, and an Oil & Gas Facility Category for facilities in the oil and gas sector.
- Illinois collects additional information that pertains to air emissions rules specific to the state.
- Colorado collects data related to permit processing and analysis.
- Wisconsin collects number of employees, area from fence line to fence line, whether an Environmental Management System (EMS) exists and if the EMS is reviewed by a third party, and whether the facility is a small business (less than 100 employee and annual receipts not more than \$750,000).

Several respondents also indicated that the information they collect for a facility inventory depends on the type of a facility. For example, North Carolina has facility data for both permitted and permit-exempt facilities, but only permitted facilities report emissions.

Some SLs also indicated more pollutants were covered in their EIs than in the NEI, such as Colorado; Maine; City of Louisville, Kentucky; Washington; and Wyoming, although pollutant emissions are not included in the facility inventory data. Therefore, the analysis of pollutant differences is excluded from this report.

*CAERS workflows with SLTs would have to be designed with these SLT-specific data fields in mind. This means CAERS would have to accommodate additional codes to existing fields, additional data fields that SLTs require, as well as QA checks, for example, to enforce reporting of required fields where the EPA does not require them.*

### 3.5 Techniques to Obtain and Update Facility Inventory Data in EIs

The techniques used to obtain and update facility data existing within EIs vary by jurisdiction. Within any one jurisdiction, they may also be dependent on facility characteristics, such as new or existing facilities; permitted or not permitted; and permit type. One jurisdiction may use more than one technique to obtain facility data for emission inventories. For example:

- Oklahoma maintains an Air Quality facility database (called TEAM) used for their permitting, compliance and enforcement, and emissions inventory programs. Oklahoma collects facility emissions data using a version of SLEIS from Windsor Solutions. SLEIS and TEAM are not dynamically connected but data do flow in both directions between the systems in an ad hoc

way. Data in TEAM relevant for EIs are only for facility level data, not sub-facility level data. Facility reporters are responsible for reporting all data for emission units, release points, processes, etc., that is, data considered to be facility data only within the context of EI reporting. Oklahoma EI staff, within operational bounds, also make updates to data during the QC process.

- Contrasting with the above, EI Staff at SLTs that have a small number of facilities may only create and update facility inventory data manually, for example: Hawaii, Lane County, Oregon, and Chattanooga City, TN.

About 30 (56%) SLT EI programs obtain facility inventory data “with data flows”, which means they do so directly from other within-SLT program systems and databases (e.g., permitting, compliance, state master database), using: system integration with other within-SLT systems, linking with other SLT databases, sharing tables, and/or snapshot synchronization from other SLT programs and/or databases. Table 4 shows the number of SLTs that use each technique noting that some SLTs may use more than one “with data flows” technique.

Table 4. Number of SLTs Using Techniques to Obtain and Update Facility Inventory Data in EIs

Technique to Obtain and Update Facility Inventory Data in EIs*	Number of SLTs
EI system is part of an integrated/within state database used by other programs (e.g., permitting, compliance, state master DB)	19
EI system is a unique application/database linked to other state database(s)	12
EI system shares certain tables with state database(s) for other programs	8
Snapshot sync** with state database for other programs	3
Manually created and updated by EI staff	32
Reported by facility operator	34

\*Note: Georgia relies on CAERS for facility inventory information, which was not listed as a technique. Before using CAERS in 2019 their EI system was a unique application/database linked to other state database(s).

\*\*Note: The sync function can be performed automatically or manually, and timing of the sync depends on needs. Here the emphasis is "snapshot" not live link. For example, Minnesota can sync all facilities or individual facilities from a state master database automatically at any desired time.

Conversely, it is notable that 24 SLTs do not have data flows such as those described above. Table 5 shows the distribution of the 24 jurisdictions “without data flows”. This means EI staff and/or facility operators manually create and update facility inventory data to report relevant facility configuration changes.

Table 5. Number of SLTs that Create and Update Facility Inventory Data Manually

Jurisdiction	No Data Flow	Total Responded to the Questionnaire	% No Data Flow from Total Number of Respondents
Local	12	16	75%
State	11	37	30%
Tribal	1	1	100%

Table 5 illustrates higher percentages of local and tribal jurisdictions without data flows as compared to states. This result is consistent with the smaller number of facilities needing to be included in emissions inventories for local and tribal jurisdictions. For example:

- Chattanooga City, TN obtains information for 170 facility sites from permitting staff during annual inspections. Data are manually entered.
- Clark County, Nevada obtains the facility site information during the permitting process and then retains it in the EI database and reports 70 facilities to NEI.
- At Southwest Clean Air Agency, Washington, facility site data are entered into the EI system at the time of permitting or in response to an enforcement action. The EI of Southwest Clean Air Agency only contains data for 19 facilities.

Table 6 shows the cases where one jurisdiction uses more than one technique to obtain facility data for facility inventories. Note that values in Table 6 include automated and manually conducted techniques. For example, the manual (without data flow) Tribal case in Table 5 clearly appears in Table 6; because the tribal respondent indicated they do manual updates, as well as obtain data from facility operators, both of which are manual.

Table 6. Number of Techniques Used by SLTs to Obtain Facility Inventory Data

Number of Techniques Used	Local	State	Tribal	Total SLTs
1	10	12		22
2	4	12	1	17
3		9		9
4	2	3		5
5		1		1
<b>Total</b>				54

Although 22 SLTs only use one technique in creating and updating facility inventory data, Connecticut uses 5 techniques except the snapshot sync with state database for other programs to obtain a complete set of facility inventory information due to the specific structure of their database and EI system.

*The nature of these updates indicates that SLTs with more than one method might need more time to update their facility inventory than other SLTs before reporting begins in CAERS every year. Therefore, the time that building and updating the facility inventory takes, should be factored in the annual workflow process between CAERS and the SLTs in question. Similarly, if the SLTs would like to get updates on the facility inventory from CAERS itself, time would have to be factored into the process so that SLTs have time for their relevant systems to absorb the changes. Finally, while in the case of EI reporting for the NEI the SLT manages the facility inventory data, it is industry reporters themselves who edit and report facility data directly to other EPA programs (for example, CEDRI). SLTs who prefer their own facility inventory databases, rather than direct industry input, would need to work with EPA to design their workflow in CAERS, so that they may provide input on the correct facility data inventory to be shared between CAERS and the other federal programs.*

### 3.6 Information Sources for Facility Site, Emissions Unit, and Process

In survey responses, SLTs commonly indicated using multiple information sources to compile their facility inventory. Some information sources may be popular for one data component while not for others. Table 7 shows the number of SLTs using the different information sources for facility site, emission units, and processes. Responses do not add up to the total number of respondents because multiple sources can be used by an SLT.

Table 7. Number of SLTs Using Different Sources of Information for Facility Site, Emission Unit, and Process

Information Sources	Facility Site	Emission Unit	Process
State permitting program	37	37	30
State compliance program	23	18	17
State master database	18	13	8
Combination SLT permitting program, compliance program, and/or master database	40	39	33
Facility operator	43	45	46
NEI facility inventory	1	1	NA
Previous EI	32	31	31
Auto created by EI system	NA	NA	10
Manually created by EI staff	NA	NA	24

Note: “NA” means the choice was not in the survey questionnaire where only the most common information sources for the components were listed.

Depending on the SLT, multiple programs can be supported by a central database, which supports interrelated handling of facility data. The state master database may include databases for a permitting program and/or a compliance program in many SLTs. Some respondents selected all three sources (State permitting program, State compliance program, and State master database) when they were asked what sources they used for creating and updating facility inventory data, while others selected only having a single master database. The entry in Table 7 listed as “Combination SLT permitting program, compliance program, and/or master program database,” avoids double counting with other rows in the table. For example, Alaska responded obtaining data from a combination of state permitting program, the state compliance program, and the state master databases. Alaska is counted as 1 state in the row of “SLT permitting program, compliance program, and/or master database”.

About 74% (40 out of 54) of SLTs are using information directly from combined SLT databases for facility sites. Although the number of SLTs using the combined information source decreases for emission units and decreases further for processes, about 61% (33 out of 54) of SLTs are using the combined information source for processes. This is because SLT permitting, compliance, and/or master databases contain facility site information but less unit (or process) level information. For example, Oklahoma maintains an air quality facility database that only contains data at the facility site level, not at the sub-facility level.

About 80% of SLTs also obtain information on the facility site level from facility operator reporting. The number of SLTs obtaining information from industry reporting increases for emission units (83%) and increases further (85%) for processes.

Only few SLTs use information from the NEI facility inventory to initiate some facility and emission unit data update. More than 57% of SLTs carry the previous EI as the starting point for the current EI. For example:

- Washington, Maricopa County, and Arizona use the previous EI report as the starting point for a new report and require facilities to review and update all information.
- Minnesota copies the previous EI to the current EI, then, only synchronizes new information in the master database with the current EI.

*As described at the end of the previous section, for SLTs wanting to use their previous year reported facility inventory as a starting point, their workflow would have to include steps to ensure data to CAERS come from SLTs with updated information necessary from all their systems or databases. Also, their workflow would have to include steps to ensure data from CAERS, once received by the SLT, would be distributed to any systems or databases the SLT would want to update.*

While a jurisdiction may use multiple information sources for facility sites, emission units, and processes, the information sources used could be dependent on the facility permitting types. For example, MN obtains information from its master database for permitted facilities that are required to report to the state EI program. However, the master database contains information at different levels of detail for facilities with different types of permits due to the state EI rule requirements. For small facilities with state general permits, the master database only has information for facility-level attributes. For large facilities with Title V and state permits other than general permits, available information includes facility sites and emission units but not processes. For non-permitted facilities that do not report to the MN EI program, such as TRI sources, landfills, POTWs, dry cleaners, and human crematories, MN uses the Toxics Release Inventory (TRI) and other information sources. In all cases, MN uses the previous EI as the starting point. For a large facility, the facility needs to add new processes or shut down the processes. For small facilities, the system auto-generates emission units and processes based on the process source classification codes (SCCs) that are selected by facilities.

*These examples indicates that in some cases, it might be beneficial for CAERS to obtain the updated facility inventory data from the SLT permitting system or database directly, instead of from the SLT EI system that is updated from the SLT permitting system or database. The detailed SLT dataflows should be considered in the development of the SLT CAERS modules. Where SLTs may not need sub-facility data for some of their facility types we require additional thinking as to how the CAERS module for each could accommodate the different types of facilities and thus, potentially different levels of detail in their inventory.*

### 3.7 Information Sources for Site Control and Control Path

EIS contains information about the types of emissions controls that are present at the facility. Site control paths provide information on how the site controls relate to each other, and between processes, and release points at the facility. Recently, the EPA changed the way it collects data on control devices to improve the configuration of control characterizations in the NEI. CERS v1.0 was designed to collect that data using a control “approach”. However, the new way of reporting controls is by including information on “paths”, and this change is reflected in CERS v2.0. While SLTs may have implemented their EI programs based on the AERR with EI systems based on the CERS v1.0, those SLT EI programs may

not collect the information for control path configurations and need time to catch up to the new way of representing control information.

Furthermore, while all SLTs are required to submit data for facility site, emission unit, and process to NEI. However, SLTs are not currently required to collect or submit data for site control and site control path because the requirements are not included in the AERR. They are also not included as an automated quality check by EIS or listed in the NEI submission guidance “cheat sheet” for facility inventory data. The EPA is considering whether new control information included in CERS v2.0 should be more explicitly required.

Therefore, additional questions were included to understand the variation of SLT data practices with respect to controls data. Table 8 shows responses about control and control path reporting practices.

*Table 8. Number of SLTs Using Site Control Device and Path Reporting Practices*

<b>SLT Practice</b>	<b>Site Control Data Total SLTs</b>	<b>Site Control Path Total SLTs</b>
<b>Collect and Report</b>	30	22
<b>Collect Only</b>	12	5
<b>Report Data Only</b>	3	3
<b>Neither</b>	8	24
<b>Blank</b>	1	0
<b>Total</b>	54	54

It is interesting to note the number of SLTs that collect and report data to NEI for controls is larger than those collecting path information. Also, some SLTs may be collecting data but not report it to NEI. Furthermore, 3 SLTs do not collect data but report data for the control devices and control path component that they have created. Eight SLTs neither collect nor report control information, and 24 SLTs do not collect or report control path information at all.

Table 9 shows the information sources for control and control paths, including the same ‘SLT permitting program, compliance program, and/or master database’ information source as described in Section 3.4.

*Table 9. Number of SLTs Using Different Sources of Information for Control Devices and Paths*

<b>Information Sources</b>	<b>Number of SLTs Using Source for Control Data</b>	<b>% Surveyed*</b>	<b>Number of SLTs Using Source for Control Path Data</b>	<b>% Surveyed**</b>
<b>State permitting program</b>	29	63%	17	55%
<b>State compliance program</b>	13	28%	5	16%
<b>State master database</b>	8	17%	4	13%

Information Sources	Number of SLTs Using Source for Control Data	% Surveyed*	Number of SLTs Using Source for Control Path Data	% Surveyed**
SLT permitting, compliance, and/or master database	31	67%	20	65%
Facility reporting	42	91%	29	94%
NEI Facility inventory	3	7%	2	6%
Previous EI	23	50%	17	55%
Based on emission units/SCCs	11	24%	7	23%

\* Forty-six SLTs collect data for site control component.

\*\* Thirty-one SLTs collect data for site control path component.

A high percentage of SLTs, about 67% (36 out of 54), incorporate information from SLT permitting, compliance, and/or master databases. About 91% (49 out of 54) of SLTs also use information from facility reporting. About a half of SLTs use previous EI reported data as the starting point of the facility inventory for the next EI report. A few SLTs use information from NEI facility inventory. About 24% of SLTs generate information based on emission units and processes (SCCs).

There are many different requirements and business practices for handling controls data. SLTs collect data based on SLTs regulations and needs that may not match those of the federal program. To meet the CERS v2.0 specifications for items required by the AERR, some SLTs may, thus, need to transform their data either during facility reporting or afterwards when the SLTs elect to report control components to the NEI. SLTs may make assumptions or gap-fill information with their best knowledge to transform the data. The following list provides some examples to illustrate the range of SLT practices in data collection and transformation.

- In ID, controls are reported by facilities. The names for the required control paths needed for the CERS v2.0 control tables are generated by the Department of Environmental Quality after receiving data from the facilities. Facilities do not see them. Controls are then manually added to the CERS staging tables. ID reported control path names to EPA for the first time for the 2020 NEI.
- In OK, facilities report control device sequence, percent capture, percent uptime, controlled pollutants by the control device, and control percent reduction efficiency. These data are carried forward into the next inventory cycle and can be amended by reporters as equipment configurations and/or processes change. OK has not implemented the site control path as a reporting requirement or as an option in the front end of its SLEIS (OK's custom reporting system). OK transforms its data into the new control path schema during the EIS submission and aligns with CERS v2.0.
- MA has built simple linear path data into their CAP emissions reporting system. More complex paths are rarely observed at facilities in MA, and MA is considering the best approach for future implementation.

- The CT emission reporting system has a sequence order included in the reporting of controls but is based on a simple unit-level control that does not model controls at a facility level, or sophisticated control configurations, such as controls in parallel.
- In Michigan, emission units are associated or linked with controls in the state system. Facilities can select which controls are installed on each emissions unit when filling out their annual emissions report. Additionally, MI creates an Excel template to collect supplemental control information from facilities to populate the CERS v2.0 data fields. This Excel template supports facilities creating control paths to the associated processes (SCCs).
- Puerto Rico does not collect control path data but creates control paths using the control data available from facilities and reports it to the NEI.
- WA manually creates new tables using previous data to meet AERR reporting requirements, and CERS v2.0 format definitions. WA does not collect “Sequence Number”, so the state artificially creates those automatically in its software to convert its data into the XML format for EIS.
- The MN EI system is designed based CERS v1.0. It only collects control information for large facilities from its master database where site controls are connected to emission units. One emission unit can connect to multiple site controls either in parallel or in series, but there is no information on connections between site controls. The code that exports data from this database to be submitted to the NEI has built-in assumptions to allow upload using CERS v2.0. MN does not collect and report control information for small facilities and non-permitted facilities.

CAERS reporters are currently submitting data to EIS directly and data are formatted in accordance with the CERS v2.0 specification. The CAERS schema allows facilities to report data under both federal requirements by the AERR and SLT-specific requests. The SLTs can request data fields in CAERS as a part of their overall SLT inventory. *While CAERS may be able to collect SLT-specific data, special care must be taken to ensure that technological artifacts (artificial requirements) are not produced when designing technical solutions, and do not create unnecessary difficulty for industry reporters or SLTs.*

### 3.8 Information Sources for Release Point and Release Point Apportionment

Although SLTs must identify release points and release point apportionments when submitting data to the NEI, SLTs have diverse requirements, business practices, and approaches in data handling. There are two types of release points: stacks and fugitives. Table 9 shows the results of the survey question asking about the data source for release point apportionment data.

Table 10. Number of SLTs Using Different Sources of Information for Release Points and Release Point Apportionment

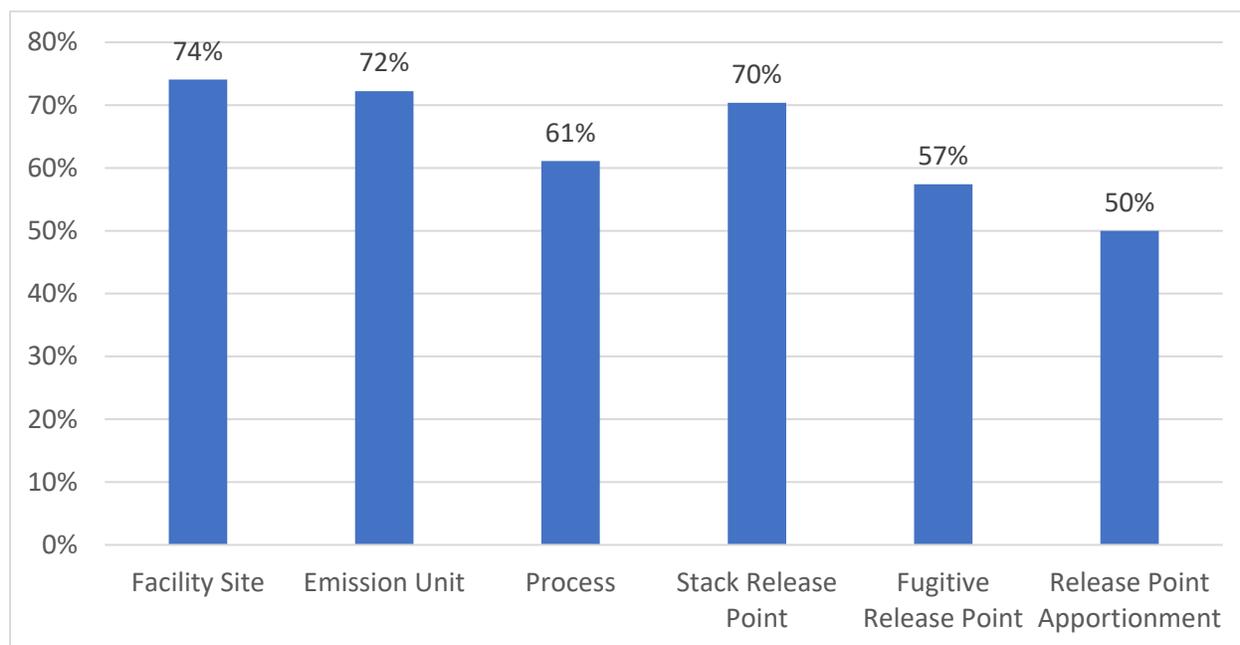
Information Sources	Stack Release Point	Fugitive Release Point	Release Point Apportionment
State permitting program	37	29	23
State compliance program	17	13	11
State master database	10	8	7
SLT permitting program, compliance program, and/or master database	38	31	27
Facility reporting	45	38	37
NEI Facility inventory	4	6	3

Information Sources	Stack Release Point	Fugitive Release Point	Release Point Apportionment
Previous EI	28	23	20
Auto created by EI system	9	10	8
Manually created by EI staff	18	16	20

The observations again show a strong dependence (at least 50%, 27 out of 54 SLTs) on SLT permitting program, compliance program, and/or master database as sources of information for release point and release point apportionments data. This reliance on SLT non-EI programs and databases for release points and release point apportionment information is less than for data on facilities, emission units, and processes (at least 61%, 33 out of 54 SLTs). This result shows that SLTs are less likely to have information readily available for release points and release point apportionments than they are to have facility-, unit-, or process- level data. Therefore, they may be having to take additional steps to create and report release point data to NEI.

Figure 3 illustrates the reliance of data sourced from non-EI datasets for facility and sub-facility information. It compares each of the data types previously described by showing the percentage of SLTs that rely on other databases within the SLT to get that information.

Figure 3. Percent of SLTs Relying on SLT Permit Program, Compliance Program and/or Master Database for Facility and Sub-facility Data



Because of the lack of release point information, it is possible SLTs will rely on their NEI facility inventory data more or exclusively to maintain and update their release point information. Specifically, 4 SLTs use NEI facility inventory as the source for their stack parameters and 6 SLTs use it for fugitive release parameters. For the release point apportionment, 3 SLTs use the NEI facility inventory, which is more than those that use it for facility (2 SLTs) and emission units (1 SLT). Fewer SLTs use previous EI information stored at the SLT, particularly for release point apportionments (36%), than for facility sites, emission units, and processes (for which at least 53% of SLTs use the previous EI). Many SLTs need their

EI information from the NEI on release point apportionments and paths, as that information is recently required and SLTs may not have had time to adjust their other facility inventory sources to intake and update the new information if they desire to have it stored there.

Facility reporting is the most popular information source of release point and release point apportionment information. Finally, about 15% of SLTs use information created manually by EI staff and 10% use data created automatically by the EI system.

A single jurisdiction may use a variety of approaches to obtain data on release points and release point apportionments. The data collected from facilities may be different from the data a SLT transforms and submits to the EIS. SLTs must use assumptions and gap-filling in data transformation to the CERS v2.0 format. In doing this, some SLTs collect total emissions and assume 100% of emissions go through stacks, and don't consider fugitive release points at all, even though doing so is not expected by the AERR. Examples are provided below.

- MA does not collect release point apportionment data. The release point apportionment information is auto created by their EI system based on some very basic assumptions.
- Pima County, AZ, has not collected or reported release point data for fugitive emissions. Fugitive emissions have been reported but not the associated release point data. Therefore, all emissions (100%), including fugitive emissions, are reported as being released through stacks. Because Pima is a CAERS user, and because CAERS is set up for facility reporters to apportion data to fugitives, in future Pima could refine its data on fugitive emissions.
- In WI, the EI system assumes all emissions from emissions units are fugitive if the emissions are not partitioned to any stacks.
- Indiana's database does not support apportionment. IN gathers the information through an online system used to prepare the emissions statements. IN saves the data in an Oracle database based on a very early version of the NEI, then, submits it through the node. IN assumes 100% of emissions are released through stacks.
- MI created an Excel template in addition to its EI system to fulfill the requirements of the new CERS v2.0. This includes the creation of release point apportionments for any associated stacks.
- MN's master database contains information for about 428 large facilities on stacks but only identifies fugitive emission units, not fugitive release points. Fugitive release points are auto generated by the EI system. MN does not report any parameters associated with fugitive release points, such as width and length. For about 1,700 small permitted and non-permitted facilities, the EI system auto generates one dummy release point (could be a stack or a fugitive release point) for an emission unit based on the first process SCCs. The dummy release point is used by all processes under that emission unit.
- Lincoln/Lancaster County, Nebraska assumes 100% emissions are released through stacks.
- In CO, process-pollutant level emissions are 100% assigned to a single stack (AIRS/AFS model).
- HI uses the default SLEIS input format for 2020 emissions. This means collecting data meeting the CERS 1.0 format and then letting the Bridge Tool convert to the CERS 2.0.

*Because CAERS is already set up to obtain detailed data for release points, SLTs taking advantage of CAERS for reporting would benefit by getting this data from industry reporters. In addition, having data for release point apportionments include fugitives would mean a better alignment with TRI reported emissions, as TRI requires fugitives to be reported, and there can be a mismatch in reporting for both NEI and TRI solely due to the fact that a facility may be reporting fugitives to TRI but not NEI.*

## 4 Discussion and Recommendations for CAERS development

The Facility PDT R&D team discussed the survey responses and their implications for CAERS. The following recommendations were issued from those discussions.

1. **CAERS capacity:** CAERS needs to be able to handle a large volume of facilities, about 30,000 facilities in one jurisdiction. This need applies to data management within CAERS and data transfer between CAERS and SLTs. Additionally, many SLTs contain non-NEI reporting facilities in their inventories and some SLTs keep such facilities in EIS already. CAERS would be used by SLTs not only for facilities that are subject to federal-reporting requirements but also for non-federal reporting requirements in use cases 2-4 of SLT workflows. Otherwise, the SLTs would need to have one reporting system for facilities that also report to one or more federal programs, and a separate reporting system for those that don't. CAERS system performance should not be slowed down for SLTs with a large volume of data.
2. **SLT-specific data fields beyond NEI requirements:** These include:
  - **data fields,**
  - **codes, and**
  - **types of facilities** including differing levels of detail for facility and sub-facility level data for the facilities depending on their type (for example, size, type of permit). Some implications of these differences are:
    - a. CAERS workflows with SLTs would have to be designed with SLT-specific data fields in mind. This means CAERS would have to accommodate additional codes to existing fields, additional data fields that SLTs require, as well as QA checks, for example, to enforce reporting of required fields where the EPA does not require them. Considerations include.
      - b. The facility inventory information beyond CAERS is not needed by more than 50% of SLTs. The CAERS development team should use the information collected here to help decide which data elements should be included in the CAERS core information model, which should not be included, and which could optionally be included depending on further evaluation and specific SLT needs.
      - c. The need for CAERS to support facility inventory information beyond CERS v2.0 depends on SLTs regulations and requirements. CAERS will support SLT-specific facility data fields already included in CERS v2.0 when webservices between CAERS and SLTs are established. In addition, at the time this report was being finalized, the PDT had discussed additional SLT-specific data fields that SLTs may want to see in CAERS in future. The need for one jurisdiction may not be a need for others, and CAERS should account for that.
      - d. The CAERS codes and functions (including QA checks) for using those data fields that meet SLT-specific needs should be considered in future CAERS versions as SLTs onboard so their specific needs may be met. However, where an onboarding SLT would like to take advantage of the fact that CAERS is already set up to take information in using a certain format, the SLT may want its module to take advantage of this set up. For example, through CAERS, the facility reporter may add fugitive release points, and associate controls and control paths to stack release points. This data could also be fed back to the SLT.

3. **Differences in SLT data collection and updating methods:** Many SLTs, from 31% to 59% of respondents depending on the components, use their previous EI as a starting point for the current EI. CAERS already copies the previous year EI to the current EI as a starting point for facility reports as a default. However, many SLT's EI systems have inherent relationship with other SLT's programs such as permitting and compliance. The systems for those other programs are also SLT-specific. CAERS could interreact with SLT emission inventory systems and/or directly interact with databases for other SLT programs (for example, permitting) depending on the SLT's specific needs. Or the SLT could create, as part of its workflow with CAERS, the compilation of the facility inventory data that the SLT wants to have in CAERS for reporting; SLTs could develop more specialized modules with guidance and help from EPA for this purpose. The SLT-specialized CAERS module could have specialized interfaces and perhaps special web services that use the SLT data fields between the CAERS facility data and the multiple SLT data systems.

There are multiple scenarios that CAERS should handle:

- a. **Differences in the SLT starting facility inventory for a given reporting year.**
- i. An SLT may rely on CAERS exclusively to obtain facility inventory updates as reported by industry.
  - ii. An SLT may rely on some of the data in CAERS but may want to supplement the facility inventory with updates from some of its other databases or systems, and/or manually.

In either of these cases, if the SLT would like to get updates on the facility inventory from CAERS, this would require a workflow that would allow them to get that data when and how they'd like to get it. In addition, if the SLT would like to parse out the facility inventory data to update different SLT databases and systems this step would also have to be factored into its overall workflow with CAERS.

- iii. The SLT may not want to rely on the data in CAERS as a starting point for the facility inventory to start reporting for a given inventory year and thus, may want to pull its facility inventory into CAERS every year.

SLTs that would like to have their EI in CAERS before annual reporting begins, and who use multiple sources of data would need to factor in more time to update their facility inventory in CAERS and their workflow with CAERS could involve coordination amongst multiple databases. If those SLTs would also want to take advantage of CAERS to allow facility reporters to update the facility inventory, they could do so with the help of QA checks in CAERS that would ensure data is entered correctly. Then, a workflow that brings those updates back to the SLT database(s) would have to be designed and implemented.

SLTs who prefer their own databases for facility inventory updates, rather than direct industry input, would also need to think about when and how to update CAERS with the goal of ensuring their most updated facility inventory is shared with other EPA programs such as CEDRI, for example. This is because for other EPA programs, industry reporters directly report and edit the facility data themselves, and thus, the SLT would have to weigh in to determine the correct version of the facility's inventory data to be shared from CAERS. Further research and discussions are needed to this effect.

4. **Control device data:** In terms of controls and control paths, there are three broad use cases for SLTs:
- a. **SLTs require detailed control and path information:** CAERS is already equipped to allow facility reporters to build paths from processes to release points and “reuse” those paths where applicable to more process-release point combinations. While the process for collection control data is straightforward for facilities with single controls between processes and release points, complex control configuration sometimes require assistance from SLTs and EPA to set up appropriately in CAERS. And this assistance has been available to facility reporters since CAERS MVP. However, in the dataflow between CAERS and the SLT, the SLT may require additional functionality to be able to update the information in CAERS as needed, if it’s facility inventory will not be based on the data in CAERS.
  - b. **SLTs do not have rules to require facilities to report but encourage facilities to report:** Ideally, CAERS could include a mechanism to auto-generates control path configurations for SLTs that do not require that information. While the data would not be as robust, this could be done based on existing information for controls, their connections and/or SCCs. The need for this feature in CAERS is highly dependent on the need to support different SLT regulations and practices, which could change over time depending on any future changes to the AERR. Because CAERS already has the capability to obtain control path data from facility reporters, SLTs may want to take advantage of CAERS reporting. Where SLTs wish to send their data to CAERS (Case 1 SLT) their use cases would have to be reviewed to see how this automated functionality might be added to their module. This means that the workflow between CAERS and SLT would have to include additional code for this transformation.
  - c. **SLTs do not have rules to require facilities to report and do not see the value of collecting detailed information on site control component and site control path component in the SLT:** The workflow would not require CAERS to return the updated control and path data to the SLT. In this case, the SLT-specific CAERS module could auto populate CAERS site control component and site control path component with the control information at a higher level (such as 80% of control for PM10-FIL) and basic assumption provided by the SLT.
5. **Release point data:** Handling fugitive release points and release point apportionment is another challenging task. Some SLTs may not have regulations to support it and facility reporters may not report what is needed in a correct way. Although some SLTs allow facilities to report the information, facilities may not do a diligent work to obtain the information without SLT requirements. Assumptions have been made by SLTs to be able to include this information in the NEI submissions. SLTs adopting CAERS, may benefit from allowing their facilities to report the release point apportionments, and incorporating that information into their systems, or apply assumptions to their data in CAERS so it is set up with fugitive release information before sending it to EIS.

The same applies to release point parameters, such as stack heights, stack diameters. Currently, some SLT EI systems auto fill the missing values by using SCC-based default stack parameters. If those kinds of needs arise, CAERS or most likely, SLT-specific CAERS modules should have this function. The current CAERS has QA checks to make sure no data are missing.

6. **Time series:** While not explicitly mentioned in the survey analysis, the team discussed that it is part of a good facility inventory to be able to track components through time. Date stamps would be helpful to include for many facility inventory data fields. This time-stamp-ability is already being explored in CAERS so that SLT-specific start and end date type fields may be included in future for SLTs that would like to have them. As new data is added these time stamps may also be generated in CAERS.
7. **Data editors:** Another aspect the team derived from the above discussion that it deemed important, and has already been mentioned, has to do with who should edit which parts of a facility inventory. SLTs may want to prevent facility reporters from editing certain data fields. SLTs need to be able to manually create and update certain facility inventory data before submitting data to NEI by EI staff. Different SLTs may require different data field restrictions. CAERS already allows SLTs to determine which data they enter versus which data they want facilities to enter. To alleviate burden to SLTs, an option that some onboarded SLTs have opted for is to issue warnings when a facility reporter has edited certain fields. This allows the SLT to be aware and accept the change if desired, without having to spend SLT staff time entering the data themselves.
8. **Editing functionality:** Given the variety of needs and that some SLTs have certain functionalities to assist in the creation and editing of their inventories, the team considered that it would be helpful for CAERS to have future enhancements or new capabilities to allow copy/paste records and auto fill of sequential numbers to make manual data editing easier and less time consuming.
9. **Reporting by tribes:** Noteworthy was the fact that only one tribal jurisdiction responded to the survey. Tribes are not currently compelled to report to NEI, and in may not have the resources to do so. CAERS could facilitate emission inventory development by tribes by providing tribes a way to collect point source data, because in most cases, there are similarities between tribal emission inventories and SLs emission inventories. By enhancing how CAERS handles facility inventory data, tribes might find the use of CAERS more advantageous and could be encouraged to report more so than they do now.
10. **Future research.**
  - a. **Facility management:** EPA has the Facility Registry Service (FRS) integrates facility information from several different national and state information systems (<https://www.epa.gov/frs>), including compliance and permit data for stationary sources of air pollution regulated by the EPA, state, and local air pollution agencies. The R&D facility sub team explored the possibility to use FRS as one information source for CAERS. It was found that while initial work had already been conducted to explore the possibility of using FRS as a data source for facility information in CAERS, including workflows with the SLTs, FRS is not designed to meet future CAERS needs in terms of facility inventory data sharing with other federal programs or SLTs without major overhaul of the system, so at this time FRS is not feasible to be used with CAERS.
  - b. **QA/QC checks:** A subsequent report from this team will explore the types of quality assurance (QA) and quality control (QC) checks that CAERS should have to allow SLTs to customize their use of CAERS.

- c. SLT facility inventory updates and how they should be shared with the other federal programs via CAERS is also a topic for future research. Especially in situations where SLTs do not wish to rely on facility reporters to obtain facility inventory information.

Integration between CAERS and SLT systems will be crucial to keep facility inventory data updated. Also, EI systems are linked with other SLT systems. This means that there are different cases for CAERS facility workflows ranging from SLTs who would use CAERS as their starting point for reporting the following year, to workflows between CAERS and SLT facility inventory databases, and potentially allowing CAERS to receive certain data from different SLT databases to mimic that SLTs current practices. Differences in the facility inventory content (types of facilities, data fields and codes) beyond the NEI also exist. CAERS will require much flexibility with respect to facility inventories to accommodate future SLTs and their workflows.

## Appendix A – Original Letter and Questionnaire

See Appendix A. Original Letter and Questionnaire.docx

## Appendix B. EIS facility staging table requirements

See Appendix B. Facility Staging Requirements.xlsx

## Appendix C. Responses to the Questionnaire

See Appendix C. All Original Responses.pdf

## Appendix D. Data analysis

See Appendix D. SLT Facility Inventory Research.xlsx