Analysis plan

Project name: Requiring electronic manifests of hazardous waste generated by EPA-led cleanup sites

Project code: 2410

Date finalized: 6/12/2025

Project description

This descriptive study will use EPA data to address multiple questions about the EPA's e-Manifest program and hazardous waste shipments. The objective of answering these descriptive questions is to better understand the use of electronic manifests for EPA-led clean-ups, the network of waste handlers that engage with EPA-led cleanups, and corrections to manifests. Answering these questions will provide insights to the EPA about current conditions on the ground and support the need to improve situational awareness as the electronic manifest requirements are implemented. The descriptive study will also provide a clearer picture of the potential for, and design of, an impact evaluation that could answer the causal questions that the EPA may be interested in.

Research questions

This descriptive study seeks to answer the following research questions (RQs), which also appear below in our table of analyses. RQs labeled as "Primary" are ones that are a priority for EPA to have answers to; other RQs should be considered exploratory.

RQ1: What does the universe of waste receivers look like?

a. **Primary:** How many waste receivers are there, and how many do business with EPA cleanup sites?
b. **Primary:** How is the distribution of EPA business allocated across receivers?
c. How stable is the universe of receivers and business distribution over time?

RQ2: What does the network of waste handler relationships look like?

a. **Primary**: How long in duration are relationships between generators and receivers, and how frequent are shipments?
b. **Primary**: How many receivers does a generator typically send to?
c. What factors (e.g., size, distance, etc.) predict which receiver will get which shipments, particularly EPA shipments?

RQ3: What do manifest corrections look like?

a. **Primary:** How frequently are manifests of different types (i.e. electronic vs. non electronic) updated with at least one substantive correction, and how has the rate of corrections changed over time?

b. **Primary:** Are particular fields/sections of a manifest more prone to corrections than others?
c. **Primary:** How much of substantive correction reduction is due to the fact that some fields are, by construction, correct and unalterable in electronic manifests and what percentage is due to a reduction in corrections in fields (waste information) that people enter information into on all types of manifests?[1]
d. What is the relationship between substantive manifest corrections and characteristics of a transport (excluding generator/receiver characteristics)?
e. What is the relationship between substantive manifest corrections and generator and receiver characteristics?

RQ4: What does the new policy rollout look like on the ground?

a. **Primary:** How has the distribution of manifest types (grouped by EPA and non-EPA led cleanups sites) changed over the study period (especially around the fall of 2024)?
b. **Primary:** How many receiving sites use any electronic manifests (grouped by whether they do business with EPA-led sites), and how has the number changed over the study period?
c. What proportion of the manifests filed by receivers who use any electronic manifests are electronic?

RQ5: Who has adopted electronic manifests?

a. **Primary:** What are the characteristics of receiving sites that have adopted electronic manifests since the start of FY2025, compared to those that have not?
b. **Primary:** For receivers working with EPA cleanup sites, what proportion of their non-EPA manifests are electronic?

Preregistration details

This Analysis Plan will be posted on the OES website at oes.gsa.gov before analyzing outcome data.

Data and data structure

This section describes variables that will be analyzed and changes that will be made to the raw data concerning data structure and variables.

Data source(s):
Data will be sourced from RCRA's Metabase tables MMANIFEST and HHANDLER5 and the EPA Superfund NPL. Appendix A describes the specific fields of interest in these tables.

1) Data on each manifest status (shipment record):

[1] To elaborate, in electronic manifests some key fields (generator, transporter, and designated facility IDs) must be correct before initiating the shipment and they cannot be changed en route. We are trying to understand how much of overall error reduction is due to this type of mechanical error elimination and how much is due to error reduction on the parts of the manifest (waste information) that transporters and receivers still fill in.

a) The data includes information like the date, the generator of the waste, the receiver, the transporter, load information, and the manifest type.

b) We will identify manifests at the manifest level using the most recent version of each manifest in the RCRA MMANIFEST table.

c) The partners have provided a table from a query indicating whether a manifest has had a correction in each of the correctable fields it considers substantive. The table captures the first correction only in cases where there are ultimately more than one. A significant correction is defined as whether there was a change to any of the following fields:

i) Generator ID

ii) Any of the Transporter IDs (concatenated)

iii) Designated Facility ID

iv) Hazardous Materials flag

v) Hazardous Waste flag

vi) Non-Haz (or state-regulated hazardous) Waste Description

vii) Hazardous Waste Description

viii) Management Method Code

2) Data on each handler:

a) The HHANDLERS table includes the locations of each handler (address and/or longitude and latitude) and the characteristics of the handlers that will be merged onto the manifest data.

b) The Superfund NPL data includes the list of EPA cleanup sites, the EPA ID of the generators serving these sites, and the site's location. We are using Superfund NPL data dated 6/24/24 that was shared by our EPA partners. They have told us that only one site has been added subsequently. We have manually appended it.

c) The MREJECTIONS table includes alternate designated facilities locations for manifest rejections, in case the original designated facility is not the final designated facility (receiver) of the waste. This will need to be accounted for to identify relationships between generators and receivers.

3) Inclusion criteria for handler relationships (RQ1-2) will include the following:

a) Data for manifests with received date between `Jan 1, 2023` and `Apr 30, 2025`.

4) Inclusion criteria for manifest correction and manifest type analyses (RQ3-5) will include the following:

a) Data for manifests will be collected for manifests with received date of Jan 1, 2023 to Apr 30, 2025. This should account for the lag between when shipments occur and when some manifests appear in the data (handlers are given 30 days). We plan on analyzing manifests with received dates through Apr 30, 2025. However we will push this date earlier (i.e. to March 31) if we see a drop off relative to base rates in the April data which we would interpret as evidence of the being more than 30 days in practice.[2]

b) Data from the manifest's received date before Sep 30, 2024 will be considered the "pre-EPA requirement of electronic manifests" period, and data starting Oct 1, 2024 will be considered the "during-EPA electronic requirement" period.

5) Exclusion criteria

a) We will limit the scope of the data to manifests where the generator origin is within U.S. states and territories.

Outcomes to be analyzed:

Variables	**Description**
Receiver	The receiving facility for each shipment/manifest. We will summarize, e.g. ● Total number of receivers ● Number of receivers using electronic manifests
Generator-receiver pair	Dyad variable capturing the originator and receiver of each shipment. We will summarize, e.g., ● Duration of pair (time between the first and last time the pair appears in the study period) ● Number of pairs conditional on the generator
Substantive correction	Captures whether at least one substantively important piece of information (e.g., shipment weight) was corrected on a manifest after submission. We will summarize, e.g. ● Correction rates by manifest type over time
Manifest type	Type of manifest used (i.e., electronic or non-electronic). We will summarize, e.g., ● Distribution of manifest types over time for EPA-led and other cleanups ● Percent of manifests that are electronic by receiver
Electronic adopter	Whether a receiving facility uses any electronic manifests. We will summarize, e.g. ● Number of receivers using electronic manifests over

[2] Since we are pulling the data in June, using data through the end of April will provide a cushion on top of the 30 day expected lag. However, we will be conservative here (and push back to March if needed) out of concern that lag times before appearing in the data are correlated with manifest type.

	time • Correlates of being or becoming an electronic manifest using facility
Receivers No Longer Doing EPA Business	We will summarize, • Proportion of recent EPA receivers who stopped receiving waste from EPA-led sites

Imported variables:

All variables will come from the EPA partners (substantive correction encoding and the Superfund NPL) and the RCRA Info via the Metabase reporting tool.

Transformations of variables:

Transformed variable name	**Description**	**How calculated**	**Format**
generator_receiver_pair	combination of waste generator and receiver	Concatenation of generator and receiver IDs	Factor
electronic_manifest	Whether the manifest type is either fully electronic or hybrid (two categories the EPA considers electronic)	1 if manifest type is fully electronic or hybrid. 0 if type is paper/image or paper/image + data	0,1 indicator
correction	Whether at least one substantive correction was made to the manifest before finalization	1 if any of the corrections fields from corrections query data is “Y”(or 1 in binary encoding)	0,1 indicator
electronic_adopter	Whether the receiver has submitted any electronic manifests	1 if the receiver has submitted electronic manifests during the study period. Zero otherwise. Note - it is possible some receivers could show up with a very small number of electronic manifests. Rather than precommit to an exact cutoff (e.g. more than zero) we will make a determination based on the data and conduct sensitivity analysis to this decision as necessary.	0,1 indicator
receiver_nshipment	number of shipments received by a receiver in a month	Count the number of manifests showing that a shipment was received by the receiver in a specific month	Numeric

receiver_pctEPA	share of shipments received by a receiver in a month that are from EPA-led cleanup sites	Count the number of manifests showing that a shipment was received by the receiver in a specific month from generators serving cleanup sites/receiver_nshipment	Percentage
receiver_ngenerator	number of generators a receiver receives shipment from in a given month	Count the number of unique generator IDs from the shipments received by a receiver in a given month	Numeric
Last_EPA_manifest	when the receiver last did business with EPA generated waste	Quarter of most recent shipment of waste from EPA led cleanup	Quarter/date
is_EPA_led_cleanup	Whether the generator is on the list of EPA-led cleanup sites	Merging from the superfund NPL list[3]	0, 1 indicator
adoption_period	Whether the manifest was generated during the EPA requirement period or before the EPA requirement period	1 - The first manifest generation date was between Oct 1, 2024 and May 31, 2025 0 - The first manifest generation date was on or before Sep 30, 2024	0, 1, indicator
is_waste_hazardous	Was there any hazardous waste on the shipment	1 if the sum of hazardous waste in the final manifest is greater than 0, 0 if not	1, 0
weight of manifest	Total weight of manifest as a proxy of size of the shipment (in kg)	Sum of total acute waste and non-acute waste (kg)	Numeric

Transformations of data structure:

The raw data will be merged and collapsed in two stages - one within the RCRAinfo Metabase reporting tool for export and one within a local environment with external EPA data because the Superfund NPL is sourced from outside the RCRAin Metabase reporting tool. The plan for the final data analyzing manifests will be one manifest ID per row. The plan for the final data analyzing generator-receiver relationships will be to use exported collapsed data from the RCRAinfo Metabase reporting tool due to the ability to query data from the database.

Data exclusion:

NA

[3] Note that even if a generator ID is found in the list of NPL sites, it may not mean that all manifests generated by this generator are from the cleanup work. Identifying which manifests result from the cleanup work may require additional data fields and assumptions.

Treatment of missing data:
We do not anticipate missing data in the set.

Descriptive statistics, tables, and graphs

Analyses denoted as ***primary*** are planned for the abstract and reanalysis. Other analyses are exploratory.

Research question	Descriptive statistics to conduct	Plots
Questions about waste handlers RQ1: What does the universe of waste receivers look like?		
Primary: How many waste receivers are there, and how many do business with EPA cleanup sites?	1. Count of receivers with at least one manifest in the data, by time period (likely either quarter or half-year, but possibly month or year, depending on data). 2. Count of receivers with at least one manifest from an EPA-led cleanup in the data, by time period (likely either quarter or half-year, but possibly month or year, depending on data).	3. Bar plot or table 4. Bar plot or table
Primary: How is the distribution of EPA business allocated across receivers?	1. For each receiver, percentage of FY 2024 business (by number of manifests/shipments and/or shipment weight) from EPA cleanup site generators. 2. Percentage of total EPA-generated shipments (by number of manifests/shipments and weight) going to each receiver.	1. Histograms 2. Density plot
How stable is the universe of receivers and business distribution over time? - How often do new receivers appear, and if so, are they close to Superfund sites? - Did the number of sites receiving EPA shipments change around the new policy rollout?	1. Count (and percentage) of new receivers appearing for the first time in the data, by quarter 2. For new receivers only, average distance (including range and median) to nearest Superfund site, by year 3. Count of receivers who stopped receiving EPA waste, by quarter ("EPA_business_stopped")	1. Table showing counts (and percentages) of new receivers, with bar for each year 2. N/A 3. N/A
Questions about waste handlers RQ2: What does the network of waste handler relationships look like?		

Primary: How long in duration are relationships between generators and receivers, and how frequent are shipments?	1. Distribution of time (in months) between the first and last observed shipment for each generation site* receiver pair observable in the data (up to study period max) 2. Distribution of shipment frequency for each pair (total shipments/relationship duration in months)	1. Histogram or scatterplot 2. Histogram or scatterplot
Primary: How many receivers does a generator typically send to?	Count (and median) number of receivers getting at least one shipment from each generation site.	N/A
What factors (e.g., size, distance, etc.) correlate with which receiver will get which shipments, particularly EPA shipments?	1. Percentage of shipments (EPA and not) by binned distance and that go to one of the three closest receivers 2. Other exploratory correlations related to which receivers get shipments	1. Histogram 2. TBD
Questions about corrections **RQ3:** What do <u>manifest corrections</u> look like?		
Primary: How frequently are manifests of different types updated with at least one substantive correction?[4] How has the rate of corrections evolved?	1. Percentage of all manifests with at least one substantive correction overall and by type 2. Percentage of manifests by type (and pooled "electronic" vs "not electronic" (see above)) that have at least one substantive correction by quarter and month.[5]	1. Summary table of manifests includes number of manifests by type and percentage with at least one substantive correction by type and overall. 2. Line plot of percentage of manifest corrections by manifest type by quarter/month.
Primary: Are particular fields/sections of a manifest more prone to corrections than others? **Primary:** What percentage of substantive corrections are due to the inherent	1. Percentage of manifests that have a correction by field/section (i.e., Rejection, Waste Information, Special Handling Instructions, Additional Information, Port of Entry, Discrepancy, Residue, Rejection Information, and Attachment).[6] 2. Among all manifests that experienced at least one substantive correction, for each manifest type, percentage that experienced	1. Bar graph showing percentage of field/section corrected by manifest type. 2. Bar graph showing which percentage of all manifest corrections are due to corrections to

[4] The definition of a "substantive correction" is to be provided by our project partners.
[5] The appropriate timespan for analysis and visualization (i.e. month, quarter, year) will depend on the number of manifests submitted/completed per month. We leave this decision up to the primary analyst.
[6] Note that the fields/sections specified here can be modified in fully electronic, hybrid, and paper + image manifest types.

features of electronic manifests?	a correction ONLY in the following field/sections: Rejection, Waste Information, Special Handling Instructions, Additional Information, Port of Entry, Discrepancy, Residue, and Rejection Information, and Attachment. 3. For paper + image manifest types which had at least one substantive correction, a percentage that experienced a correction in the following fields/sections: Manifest tracking number, Import, and Transporter Information.[7]	Manifest tracking number, Import, and Transporter fields/sections by manifest type.
What is the relationship between substantive manifest corrections and characteristics of a transport (excluding generator/receiver characteristics)?	1. Percentage of manifests with at least one substantive correction by journey traits such as binned distance travelled (difference between generator and receiver geographical location) or binned time in transport.[8] 2. Percentage of manifests with at least one substantive correction by geographical location (i.e., by state or national region). 3. Correlation between whether a manifest had at least one substantive correction and variables (e.g. weight) capturing amount of waste.[9]	1. Maps showing percent of manifests with corrections by state and by weight and state. 2. Bar charts showing percentage of manifests with at least one substantive correction by distance travelled (bins) and time in transit (bins).
Secondary: What is the relationship between substantive manifest corrections and generator and receiver characteristics?	1. Distribution of substantive manifest corrections across generators/receivers. For example, "5% of the generators are responsible for 95% of all substantive manifest corrections". 2. Correlation between generator and receiver traits and percentage of manifests with at least one substantive correction. 3. Correlation between the age of generator-receiver relationship and the percentage of manifests with at least one substantive correction.	1. Histograms of percentage of manifests with substantive corrections by individual generators and receivers and dyads.[10] 2. Scatter plot (error rate by dyad age and/or total number of shipments.
Questions about adoption RQ 4: What does the new policy rollout look like on the ground?		

[7] Note that the fields/sections specified here cannot be modified when the manifest type is fully electronic or hybrid.
[8] The number of distance travelled bins and the definition of thresholds for each bin is to be determined by the analyst.
[9] While we may do additional exploratory analysis the motivation for (2) and (3) is that some states rely on these data for assessing things like weight based fees.
[10] Conditional on having a small set of generators which can be visualized on a bar chart.

Primary How has the distribution of manifest types (grouped by EPA and non-EPAled cleanup sites) changed over the study period (and especially around the fall of 2024)?	The proportion of manifests by manifest type in each month. - Manifest types are: electronic (including both full electronic and hybrid) and nonelectronic (including image + data and image only) - Also split by EPA and non-EPA-led cleanups.	1. One line plot by month for each manifest type as a share of all manifests. 2. A similar line plot, but for manifests from EPA cleanup sites only. 3. A similar line plot, but for manifest from non-EPA cleanup sites only.
Primary How many receiving sites use any electronic manifests, and how has the number evolved? Total, then split by EPA-led sites and non-EPA sites.	Proportion and raw count of all active receivers using electronic manifests ("electronic adopters" (those filing at least one manifest in a month).	1. Line plot by month showing proportion of all receivers that use electronic manifest; bar graph showing raw count of the number of receivers, by electronic users vs. nonusers. 2. Similar line plot and bar graph, the EPA site only. 3. Similar line plot and bar graph, non-EPA site only.
What proportion of the manifests filed by receivers who use any electronic manifests are electronic?	For the subsample of receivers who file any electronic manifests, the proportion of manifests that are electronic in each month.	1. Line plot by month showing the proportion of electronic manifests among adopting receivers. 2. Similar line plot, receivers working with the EPA site. 3. Similar line plot, receivers working with the non-EPA site only.
RQ5: Who has adopted electronic manifests?		
Primary What are the characteristics of receiving sites that have adopted electronic manifests since the start	Select the subsample of receivers who never filed an electronic manifest before October 1, 2024. Then split this subsample by whether a receiver has become an electronic adopter since October 1, 2024, to examine the following:	A table with three columns that list the median of each variable representing receiver characteristics: column 1 covers all

of FY2025 compared to those that have not?	- number of shipments received per month, - percentage of received shipments that are from EPA cleanup sites, - EPA region (indicator variable for each), - monthly average number of generators from which waste has been received.	nonadopter before FY2025, column 2 covers nonadoptors who have since filed electronically, and column 3 reflects nonadoptors who still have not adopted.
Primary For receivers working with EPA cleanup sites, what proportion of their non-EPA manifests are electronic?	Among manifests filed by receivers working with EPA cleanup sites in FY2025 but are not from these sites (i.e., noncleanup waste), the proportion that are electronic in each month.	A line graph showing the proportion of electronic manifests among non-cleanup manifests of receivers working with EPA cleanup sites.

Limitations:

A limitation of the study is that we primarily observe information about waste handling and shipments after the shipment has been completed; this observation limits our ability to understand which actors in the chain of custody may be taking different actions. In particular, we are unlikely to know which handlers make decisions about when to adopt and use electronic manifests, or the origin of errors in manifests that require corrections. It is also possible that factors like the rate at which actual errors are caught and corrected could be correlated in unobserved ways to other attributes of interest.

A second limitation is that the duration of the study only covers a limited amount of time since the implementation of the new electronic manifest policy for EPA-led cleanup work. Implementation and adoption is ongoing, and an early snapshot of manifest use after October 1, 2024 may not be representative of electronic manifest usage over the longer term or in the months following the study period. One potential concern here is that the laggards (those who have yet to adopt electronic manifests) may be different in other important ways. For instance they may have more to lose from switching or have less capacity and fewer resources in general. To the extent the laggards are different it is possible that what we observe in the first few months might be different (potentially more muted) as more receivers transition.

Similarly, the observations of waste handling and manifests are associated with a specific set of clean up sites and waste generation. Especially for EPA-led cleanups of Superfund sites, the location and types of sites in the future may vary compared with the current set of sites.

Appendices

Appendix A. Sample raw data fields of interest

Field	Table	Type	Example	Purpose
Manifest Tracking Number	MMANIFEST	String of Letters and Integers		ID for mapping Unique identification number associated with the manifest.
Creation Date	MMANIFEST	Date		Creation date of the manifest
Updated Date	MMANIFEST	Date		The date of the manifest's last update. We can track the manifest's updates through its journey from formation to final signature
Manifest Status	MMANIFEST	String	Corrected or re-signed or submitted	The last status of the manifest
Submission Type	MMANIFEST	String	Hybrid, Data5Image	What type of manifest was submitted
Origin Type	MMANIFEST	String	Service, mail, web	How the manifest was provided
Generator EPA ID	MMANIFEST	12-digit Integer		ID for mapping EPA identification for the generator
Generator Location ZIP	MMANIFEST	ZIP		The zip code where the generator is physically located. This will aid in calculating "crow flies" distances between the generator and receiver.
Designated Facility EPA ID	MMANIFEST	12-digit Integer		ID for mapping EPA identification for the receiver (designated facility)
Designated Facility	MMANIFEST	ZIP		The zip code where the receiver is physically located.

Location ZIP				This will aid in calculating “crow flies” distances between the generator and receiver.
Original Manifest Tracking Numbers	MMANIFEST	Integer	Integer	For mapping in case one of the substantive changes to the eManifest data was a tracking number (could be attributed to typing or writing errors)
Broker EPA ID	MMANIFEST	12-digit Integer		Brokers coordinate with the generator on treatment, storage, and disposal. Network analyses between brokers and generators can give insights into adoption and errors.
Manifest Tracking Number	MREJECTION	Numeric		ID for mapping
Rejection Type Indicator	MREJECTION	String	Partial, Full	Categorization of the rejection. A potential indicator for the cost of adoption/non-adoption of electronic manifests (time/resources)
ALternate Designated Facility EPA ID	MREJECTION	12-digit Integer		ID for mapping If a manifest switches receiving facilities, then we use the alternate designated facility EPA ID and create a flag for switching to an alternate receiver
Alternate Designated Facility Location ZIP	MREJECTION	ZIP		To determine the distance between a generator and an alternate receiver in case of switching receivers
EPA Handler ID	HHANDLER5	12-digit Integer		ID for mapping to other columns of IDs
Location Latitude Measure	HHANDLER5	Numeric		Latitude of the handler. To be joined with the mapping ID and to calculate distances.
Location Longitude Measure	HHANDLER5	Numeric		Longitude of the handler. To be joined with the mapping ID and to calculate distances.

EPA ID	NPL spreadsheet	String of letters and numbers - first two characters are the state of the site	CT0121 343459	Generator ID for mapping to other columns
FF Ind	NPL spreadsheet	Character	Y/N	Is the generator site a federal facility? This variable indicates if the generator is an EPA-led cleanup site
NPL Status Date	NPL spreadsheet	Date		Indicates when the site became an NPL
Manifest Tracking Number	Partner-provided table	String of letters and numbers		To merge with MMANIFEST and MREJECTION
Submission TYpe	Partner-provided table	String	DataIma ge5 Copy	The type of manifest
Gen_ID_SC_In dicator	Partner-provided table	Character	Y/N	There was a correction on the generator ID
TRP_ID_SC_I NDICATOR	Partner-provided table	Character	Y/N	There was a change on the transporter ID
DF_ID_SC_IN DICATOR	Partner-provided table	Character	Y/N	There was a correction on the designated facility (receiver) ID
EPA_FLG_SC_ INDICATOR	Partner-provided table	Character	Y/N	There was a change to the hazardous materials flag
PCB_SC_INDI CATOR	Partner-provided table	Character	Y/N	There was a change to the hazardous waste flag
NH_WST_SC_ INDICATOR	Partner-provided table	Character	Y/N	There was a change to the non-hazardous waste description
HZ_WST_SC_I NDICATOR	Partner-provided table	Character	Y/N	There was a change to the hazardous waste description
MMC_SC_IN DICATOR	Partner-provided table	Character	Y/N	There was a change to the management method code.