

Transcript for Enforcement and Compliance History Online (ECHO) Advanced Webinar – Behind the Scenes

April 9, 2019

Hello, and welcome to EPA's Advanced ECHO webinar Behind the Scenes, our first speaker is Melinda Miller.

Hi everyone, my name is Melinda Miller and I work for Eastern Research Group supporting EPA's ECHO website. I'd like to welcome you to the Advanced ECHO series of webinars. Today we're going to explore the Behind the Scenes of ECHO, addressing common questions such as:

- From which EPA program systems does ECHO retrieve data?
- How frequently are the data updated, and how do I make corrections?
- How can I download ECHO data? and,
- What are ECHO web services?

I'll just mention that this is a re-recording of the live April 9th webinar due to some technical difficulties that we had, and we will have a question and answer session at the end of the presentation using questions that we had submitted during the live webinar.

Now, I would like to take a few moments to introduce everyone to ECHO to make sure we are all beginning from the same starting point. The U.S. Environmental Protection Agency or EPA provides public access to its regulatory compliance and enforcement data through the Enforcement and Compliance History Online website, which we call ECHO.

Data included in ECHO indicate how a facility is regulated, when an inspection occurred, whether violations were found, and whether any enforcement actions were taken.

ECHO presents the compliance history for more than one million EPA-regulated facilities. This includes three-year compliance status history and five-year inspection and enforcement history for Clean Air Act stationary sources, Clean Water Act permitted dischargers, Safe Drinking Water Act public water systems, and Resource Conservation and Recovery Act hazardous waste handlers.

Data are reported by EPA and state and local environmental agencies. ECHO also includes EPA enforcement action data under other statutes.

ECHO itself is not a data system of record. ECHO pulls data weekly from several EPA program data systems, such as the Integrated Compliance Information System (ICIS), the Resource Conservation and Recovery Act Information System (RCRAInfo), and the Safe Drinking Water Information System (SDWIS). It also utilizes EPA's Facility Registry Service and pulls select data from EPA's Envirofacts, including Toxics Release Inventory and Greenhouse Gas pollutant release data, which helps provide a more complete picture of facility activity. We'll discuss these data systems in more detail later in this presentation.

ECHO provides a number of features to help you access and understand environmental data, such as the Facility Search and Enforcement Case Search. You can use these tools to search for facilities that match specific characteristics of interest and then choose specific reports to view detailed environmental information.

We will have a question and answers session at the end of the presentation using questions that were submitted during the live webinar

Okay, let's get into the Behind the Scenes of ECHO.

From which EPA program systems does ECHO retrieve data?

The first question we will look in to is, from which EPA program systems does retrieve data?

As we mentioned in the introduction, ECHO primarily presents enforcement and compliance data related to four environmental statutes. These data come from various EPA program data systems. Clean Air Act stationary source data are from ICIS-Air, Clean Water Act permitted discharger data are from ICIS-NPDES, Safe Drinking Water Act public water system data are from SDWIS, and Resource Conservation and Recovery Act hazardous waste handler data are from RCRAInfo. Each of these environmental statutes and their associated regulations has certain data reporting requirements for the various types of data, such as inspection or violation information, and these reported data are entered or submitted into their respective EPA program data system. ECHO does not present violation data related to other statutes; however, it does offer compliance monitoring and formal enforcement action information at facilities regulated by other environmental statutes.

ECHO uses EPA's Facility Registry Service (FRS) ID numbers, which link EPA-regulated facilities and permits across programs under one unique ID number. ECHO also pulls select pollutant release data from EPA's Envirofacts data warehouse, including information from the Toxics Release Inventory and Greenhouse Gas reporting program, which helps provide a more complete picture of facility activity. You can find a full list of ECHO data sources on the About the Data page, which we will look at later.

This is a diagram that generalizes the data flows to ECHO. A fundamental take away from this webinar is that ECHO itself is not a data system of record. You can see that data about facilities regulated under the four main environmental statutes come from different sources and are collected in separate EPA Program Data Systems, and the data are then pulled from these systems into a database that we call the ECHO Datamart. In the ECHO Datamart, EPA processes and integrates the data from these separate systems. ECHO presents the data in an understandable and searchable format for the public, regulated entities, and government agencies who use and analyze compliance information. One of ECHO's goals is to present a comprehensive look at compliance and enforcement history at regulated facilities.

The program data systems that store environmental data are populated in a number of ways. For example, some regulated facilities electronically report data in state or EPA data systems, while others submit physical reports to these agencies. Some agencies that collect this data bulk transfer the data from their systems to EPA national systems. There are sometimes errors or inconsistencies in the way data are transferred between systems.

Since ECHO pulls data from the two nationwide program systems, EPA lists general data migration issues on the Known Data Problems page. We can access the Known Data Problems on the Help tab. Note that

broad-scale issues that impact the completeness, timeliness, or accuracy of data are labeled as Primary Data Alerts. We recommend that you review this page to see if your data of interest are affected.

EPA is currently producing some video tutorials that explain these data flows in more detail. We expect these to be released over the next year. Keep an eye out for new video tutorials on the ECHO Tutorials webpage.

Before we do an example search, we want to point out some training resources on this page. If you are unfamiliar with ECHO searches, we recommend viewing the 5-minute Facility Search Tutorial available on the Video Tutorials page or the Intro to ECHO webinar available on the Webinar Trainings page.

We're going to do a simple example ECHO search to see facility data. We will do a Quick Search using a city and state. The Quick Search looks for facilities across all data or all four environmental statutes, which is the broadest type of facility search. Notice that our search has returned over 500 facilities. Let's look specifically for facilities with recently identified violations by checking the Facilities with Current Violations box on the Search Results page. This narrowed our search down to four facilities. You can click on a hyperlinked facility name to open a Detailed Facility Report, or DFR for that facility. We'll open the DFR for Titanium Metals; note that this is just an example for demonstration purposes.

The DFR presents detailed enforcement and compliance information for a facility. It is organized into six sections, with the Facility Summary providing summary information for the main statutes available in ECHO. Generally, when reading the report from top to bottom, information will flow from overall summaries to more detailed information, split out by each environmental program. The report automatically defaults to show data for all media programs, but you can customize it to show data for just one.

The Facility Summary presents an overview of the facility. It lists basic information about the facility, such as location and industry type. You can also see the Facility Registry Service, or FRS, ID. Most facilities in ECHO have a unique FRS ID; this is how EPA links together data from multiple environmental data systems. The Enforcement and Compliance Summary table presents a brief overview of the facility's most recent compliance status for different statutes. Below this table, we can see permit numbers and ID numbers associated with various data systems.

Let's use the navigation menu to jump to the Enforcement and Compliance section. This section is the bulk of the DFR. The first table lists information about local, state, and EPA compliance monitoring activities. Let's look at the System and Lead Agency columns in this table. This is a key to understanding the sources of ECHO data. For example, "RCR" indicates the data came from EPA's RCRAInfo system, and "State" indicates that the state environmental agency reported the inspection.

Also, in this section is the Three Year Compliance History by Quarter table, which provides a detailed history for each statute. You can see compliance information for the previous 12 quarters, as well as specific data for reported violations. Note that some of these violations are reported by environmental agencies, such as violations found during inspections, which like the compliance monitoring data will indicate a lead agency. Others, such as Clean Water Act effluent violations, are automatically generated in the program system by comparing electronic Discharge Monitoring Report or DMR data with permit limit data.

You can download the data for the Three Year Compliance History Table by clicking the Download Data button. If you have questions about the data or terms used on the DFR, we recommend visiting the DFR Data Dictionary by clicking on the book icon. Farther down, we can see both informal enforcement actions and formal enforcement actions for the facility. Additional details, such as environmental conditions and TRI data are also available farther down in the report.

While we are looking at the example DFR, I will mention that we receive a lot of questions about violations at specific facilities. We are happy to answer questions about the ECHO website and how it uses environmental data, but the ECHO team doesn't have information that is any more specific than what is displayed on the website and in the downloads. Remember that ECHO pulls all data from EPA program systems and facilities do not report directly to ECHO. If you ever have a question about a specific facility or violation, we recommend contacting the permitting agency, which is usually a state agency or EPA Region, or the Lead Agency that is inspecting and enforcing at the facility.

How frequently are the data updated, and how do I make corrections?

The next question we will get into is how frequently are the data updated, and how do I make corrections?

On this slide, we have provided some links to resources that give more information about ECHO data, including definitions, data updates or "refreshes", and error reporting. Information about how often and how recently data were refreshed can be found on the About the Data page.

The About the Data page provides links to webpages and documentation of data sources outside ECHO, where available, as well as specific information on when data are refreshed. In general,

- ECHO presents the past five years of facility inspection and enforcement action data.
- It presents the past three years of facility compliance data are presented as months or quarters of compliance based on the federal fiscal year. A quarter is a three-month period.
- All years of EPA formal enforcement action data from EPA's Integrated Compliance Information System are available via the EPA Enforcement Case Search.
- Additional years of data are available in the ECHO Downloads, which we will talk about later in the webinar.
- The table on the About the Data page lists the different source databases, a short description, and the data extraction dates.
- Most data are refreshed on a weekly schedule from EPA source databases. Drinking water data are an exception and are refreshed quarterly.
- Some of you may use the Comparative Maps or the State Dashboards. These use frozen data, which provide a static snapshot of the data at the time they were frozen. As a note, federal fiscal year data used in the Comparative Maps and State Dashboards are typically frozen at the beginning of the following year. For example, fiscal year 2017 data were frozen in early 2018. Data for the current fiscal year are updated until the time that they are frozen, consistent with the dates on this page. Based on common requests, it is likely that ECHO will stop using frozen data on dashboards over the next year.

To put this information into better context, let's look at the DFR from our previous example.

If you want to know how current the data are on the DFR, you can scroll to the Compliance Summary Data table. This table has a column called Data Last Refreshed, which contains dates that show when data from each program were refreshed. Additionally, there are two pieces of information on the bottom of the page. The Last Updated date tells you when the webpage itself was most recently updated. However, EPA refreshes the underlying data more often than the user interface. The Data Refresh Information link takes you to the About the Data page. Note that these dates indicate when ECHO refreshed its data; but some natural lags may exist as data makes its way from various sources into state and EPA program data systems.

Since we have the DFR open, this is a good time to point out how to report a potential data error.

If you suspect that there is an error in the data shown on the DFR, whether it's an incorrect date, old facility address, or some other issue, this button allows you to report it to EPA. You can use this button to fill out and submit a report, which will be reviewed by EPA and routed to an official who can correct program system data as appropriate. Notice that small yellow icons have appeared next to specific rows in the report. You can click the icon on the table row containing the suspected error to fill out a corresponding Error Report. If there isn't an icon corresponding to the specific table row with the suspected error, you can report a general error by clicking Report a General Error at the top of the form. It will take you to a new page where you can enter contact information and describe the error. Once you submit the report, it is entered into EPA's Integrated Error Correction Process and routed to the appropriate EPA or state agency data steward for evaluation.

The data steward will receive a snapshot of the DFR and your description, so it's important that you are specific and concise. Additionally, note that this is a vehicle for reporting data errors, and not for asking questions about the data. It's likely that there are some data stewards are listening to this broadcast too, so I'll just add that if you mistakenly receive a report for data you aren't responsible for, please reply to the report and let us know.

Lastly, while this example is on the DFR, there are analogous Report Data Error buttons available on other ECHO reports, such as the Effluent Charts and Enforcement Case Reports.

If you forget the steps we just showed you, you can always click on the Help tab on the ECHO homepage. Here, there is a link to a page with a video tutorial that has step by step instructions on how to report a data error. We also provide this link in the presentation slides.

How can I download ECHO data?

The next topic of the webinar is, how can I download ECHO data?

ECHO facility searches and reports generally provide the most recent 5 years of compliance and enforcement data. For users who want larger datasets that cannot be easily queried through a facility search, or for those who need data that is over 5 years old, ECHO makes data sets available on the Data Downloads page. The Downloads page contains links to download nationwide data sets that may be of particular use to developers, programmers, academics, and analysts. This slide has links to both the ECHO Exporter and the National Data Downloads, which we will talk about in a moment.

You can access the downloads page by going to the ECHO home page and clicking on the Data Services tab. Here, there is a link to the National Data sets.

The first download on this page is the ECHO Exporter. The exporter includes summary data for more than 1.5 million regulated facilities. Each facility, distinguished by FRS ID, has a single line of data. Over 130 data fields are available for each facility including pollutant release data, industry codes, permit types, etc. For more information about what is included in the ECHO Exporter, open the ECHO Exporter Summary file. The ECHO Exporter is updated weekly as a part of the ECHO data refresh.

One thing to note is that this is a very large file, and Microsoft Excel isn't capable of loading the whole thing. We recommend using database software like Microsoft Access or MySQL to handle the data.

If you scroll down further on the page you will see downloads for specific EPA programs. For example,

- ICIS Air Dataset, which includes the Clean Air Act stationary source data
- ICIS-NPDES dataset, which has the Clean Water Act dischargers data (split into multiple files by the type of data, including effluent violation datasets and DMR datasets for different years),
- RCRAInfo dataset, which includes hazardous waste handler data
- ICIS FE&C dataset (which includes information about EPA enforcement actions).

All of these are also updated weekly as a part of the ECHO refresh.

Let's open up one of these files as an example. We downloaded the Part 1 of the ICIS-NPDES Dataset before the webinar, so we can show you that. Notice that there are many CSV files included in this download. There are separate files for permits, violations, enforcement actions, and other information. Let's open the Formal Enforcement Action file. In this table, you can see fields such as NPDES ID, Enforcement Action Types, and Settlement entered date. Usually these files contain some overlapping key data fields, so it is possible to join them together to perform analyses.

Help documentation is available which provides descriptions of the CSV files, defines the data fields, and identifies the key fields to appropriately join the tables in a relational database. For example, the ICIS-NPDES Data Summary page has information about the files we just downloaded. The help pages will also tell you which fields are included in all files, which can help you join datasets together. In this example, NPDES ID is one of the fields that is included in each file. If you have any questions about the data that are not answered in this documentation, you can use the Contact Us link to send us a question

Like the Exporter, these are large files, so you may need database software to handle the complete files. However, we have broken a few of the popular data sets down by EPA region and state. You can look for the links that say download by jurisdiction.

If you have a need to link together datasets from multiple EPA program data systems, then you will need to use the FRS Facilities and Linkages file. This file provides facility records by FRS ID number with cross-references to data contained in the various program data systems. The help page also has documentation that explains what is found in the file.

What are ECHO Web Services?

The next topic of the webinar is, what are ECHO web services?

Web services are functions that provide business logic, data, and processes and can be accessed over the internet. Generally, you won't need to use web services unless you are a web developer or a researcher. Web services support the ECHO website and are available to the public. The services allow

web developers to design custom applications using a live feed of data from ECHO. For example, researchers who find that ECHO doesn't provide the specific set of results that they want can use these web services build their own queries. EPA provides a collection of GET or query-only RESTlike services, available through a simple URL HTTP link. Outputs are available in XML, JSON, or JSONP formats.

We have a link to the Web Services on this slide, but we will show you how to get there from the ECHO homepage. To access the ECHO Web Services page, click on the Data Services tab, which is the same place we went to access the National Datasets link. Here, there are links to both the Web Services and the Map Services, which provide geospatial data. Today, we will look at the ECHO Web Services.

This page lists the available ECHO web services. They are grouped into Search services (such as the Air Facility Search and Drinking Water Search) and Report Services (which includes the Air Pollutant Report, Detailed Facility Report, and Effluent Charts Services).

We won't spend a lot of time going over the services, but we will do a brief demonstration of how to create a web service query using the Detailed Facility Report service. This link brings you to a page with documentation and instructions for using the service.

On this page, we can click on the Detailed Facility Report text to view service documentation. These buttons show you more information about how to use different kinds of service requests. We will click the first GET button, for the Detailed Facility Report Service. This type of request is used to populate all of the data on the Detailed Facility Report. We looked at an example DFR earlier in the webinar, and we can use that same one for the services demonstration. Let's copy the FRS ID from the DFR.

This documentation lets you enter different input values to generate example URLs and responses. I'll specify the output as JSON and paste the registry ID from the DFR. Then, we just click on the Try it Out button. This returns an example Request URL that could be used to call the service, and also shows the example response that would be returned. We can also copy and paste the Request URL directly into our browser to see the response. In the response body, you can see that the service returned the same data that are presented in our example DFR. For example, the Facility Name is Titanium Metals Corporation, which matches our DFR. This is useful because a web developer could use this service to retrieve and display data within their own application, without having to store the data in their own database. Every time they call the web service, they retrieve the most recent ECHO data.

If you have questions about ECHO web services, you can submit them through the Contact Us link at the top of any ECHO page.

That's the end of the web services demonstration.

To recap, during this presentation we:

- Learned about system data sources and how the data flow into ECHO;
- Learned about ECHO data updates, refresh schedules, and how to make corrections;
- Learned about ECHO data downloads for those who need older data or larger datasets; and
- Learned about ECHO web services and web service documentation.

On this slide we have links to additional resources for ECHO. If you are new to ECHO, we recommend starting out by watching the short video tutorials or the recorded Intro to ECHO webinars. The FAQ page

also presents concise background information and answers to common questions about ECHO. If you want to receive email updates about the ECHO website, we recommend joining the ECHO Listserv.

Alright, now we will answer some questions that were submitted during the live webinar.

Hi, this is Rebecca Kane at U.S. EPA Headquarters and I will be reading the questions we received. As mentioned before, these are questions we received during the live webinar.

The first one is, can I view hazardous waste permits for a specific facility?

No, permit documents are not available currently on the ECHO website, but by using the Facility Search you can get a lot of information that might be captured on a permit or other related facility information. And I will just add that EPA has a separate sight called RCRAInfo web. It's rcrapublic.epa.gov/rcrainfoweb/ that has some additional data about RCRA permitting.

The next question is, could you please elaborate on what are considered formal and informal enforcement actions in ECHO?

Yes, if a violation is identified at a facility, facilities can return to compliance in several ways and enforcement actions are a tool that Environmental Agencies can use to help return a facility to compliance. Some facilities will correct the issues that lead to the violation without the need for any enforcement action. Informal actions are typically notices to the facility that an enforcement proceeding could occur if the violation is not resolved quickly. Formal enforcement actions which could include administrative orders, administrative penalty orders, or civil judicial enforcement cases filed in Federal court are used for more serious types of violations or continuing violations that are not corrected quickly by the facility. We have more information on the help tab on the homepage, if you look at the frequently asked questions, and also a lot of good information on the detailed facility report data dictionary with links to additional resources if you want more detail about these enforcement actions.

If I have questions about data download content, who should I address?

If you have questions about the fields that in the data download, the first thing we recommend is looking at the help documentation. Download summaries are on the National downloads page, then if you have questions after that, please use the Contact Us page, Contact Us link at the top of the page.

The next question is about Clean Water Act data; there is so much NPDES data that I am having trouble homing in on simple information, for example, permit dates by cities. Do you have any tips?

We start with using a Water Facility Search to look for NPDES data. When you search on a facility, it returns one row of data per facility and provides some summary level information. Also, on the facility summary search results, you have links to more detailed information through report links. You can look, you can identify what facility you are interested in and then one by one look at more detail. If you need to summarize some information across a city or state and want a larger set of information, you can the data download. You could start with a facility search and move to more detailed information from there. We have also had some past webinars that have focused more specifically on water discharges and NPDES data if you are interested in watching those.

The next question is specific to drinking water data; it applies to the data on ECHO. The question is, when ECHO updates SDWIS data, it is also updating the full three years of compliance history data or only the data from the current quarter?

As we mentioned, we can look at About the Data for the last refresh and the next expected extracts. When ECHO is refreshed, we refresh all years of data, so that would be the full three year of history shown on the Detailed Facilities Report.

The next question is about error reporting from a state agency. Are primacy agencies supposed to use the same data error reporting process that the public uses or is there a different process for primacy agencies to use?

Primary agencies may use the error reporting feature if it is something maybe outside their particular program, or a staff in a privacy agency may be able to identify the data steward within that agency and contact them directly. I agree that the first step would be to contact a data steward in the particular state agency for that program. If that doesn't apply, then talk to your EPA regional contacts to get help from them and then if all else fails, to contact us and we can help direct you to the right person.

The next question is, is there a way to track corporate ownership across facilities, like all stores that are owned by a particular company?

There is not a simple way to track corporate ownership across the facilities. We do not have an identifier equivalent to an FRS id for corporate ownership, however you can use a number of strategies to identify facilities owned by the same entity. That could be searching by location, searching on partial facility names and other identifying information to develop a list of facilities that associated with a particular identity. We offer a particular search to help facilitate that known as the Corporate Compliance Screener and encourage you to take a look that tool or go through the facilities search to try to identify and maintain a list of facilities.

The next question is, how reliable is the State Agency data?

EPA works with States to continuously maintain the information shown in ECHO and ensure the data are accurate and up to date. We would encourage you to read more about the data quality on the About the Data page which provided more information on how environmental agencies work together to maintain information. It also includes a link to known data problems which we discussed in the webinar which identifies issues usually with data transfers between program databases and EPA databases. ECHO also offers a reporting feature that allows others when they identify potential errors to correct them in the program data system.

Can we connect this data to our state-level open data portal and present defined queries based on local knowledge?

Yes, this sounds like a good use case for the Web Services. You could develop custom queries using the web services to pull ECHO data into your particular application or data portal.

The next question is, sometimes facilities appear more than once in a search, why is this?

Depending on which search type you use, sometimes we see facility names that are very similar or identical. The facility searches identify facilities either by the FRS ID or another program identifier. If the

rows in the table look identical by name, we encourage you to look at the other columns of data, either FSR ID, could be NPDES ID, etc. or add some additional columns and you will see unique program identifiers in those fields of data. If there still happen to be similar duplicates, you could also contact us with that particular example, and we can look more into it.

The next question is related to the specific data systems that update in ECHO. Are any free text ICIS data fields queryable in ECHO?

To remind everyone, the ICIS database provides information on water, air and also brings the enforcement cases. The only queryable fields, open text fields, that we can search on is in the enforcement case search and that is for the case summary where you can search for terms contained in the case summary. I can't think of any other fields in the facility searches that are free text and queryable. We would encourage you to look at, use the facilities searches, see if that information is provided that you are looking for or look at the National Data downloads to see if those fields exist. I would just add that this applies to anybody, if there is information that you would like to be able to query in ECHO, you are welcome to submit suggestions to us in the Contact Us link. I can't promise that we will implement it, but we put it on our list of possibilities.

This is the final question for our re-recording today. Can I download RCRA data for a specific state?

Yes, you can; you can do it two ways. The first is to run a Hazardous Waste Facility Search on your State and download your table of results. The second method would be to use the National Data Sets for RCRA and then filter out the information for your particular state.

That was our last question, thank you Melinda and Eva and Rebecca for a great presentation. And a big thank you to everyone that joined us on this webinar. We hope you found it very helpful. This concludes the presentation, thank you again and enjoy the rest of your day.