Transcript for ECHO Behind the Scenes Webinar

October 13, 2020

Good afternoon and welcome to today's presentation. Before we get started, let's review a few housekeeping items. Audio is available for this presentation through your computer's mic and speakers or by telephone. Your call-in number as well as your access code is in the control panel box on the right-hand side of your screen. All attendees have been muted to minimize background noise. If you have a question during the presentation, please type it in to the questions box on the upper right-hand side of your screen. We will have dedicated time to answer questions during the presentation. If you're experiencing any technical difficulties, please contact us and we'll try to troubleshoot the issues. A recording of the presentation will be available on the ECHO website in the next week. And lastly, a survey will appear on your web browser at the end of the webinar, so please make sure to provide your comments and feedback to us. And with that, I'll pass it on to our first speaker, Amanda Speciale.

Hi everyone, my name is Amanda Speciale and I work for Eastern Research Group supporting EPA's ECHO website. I'd like to welcome you to the Introduction to ECHO series of webinars, which provide basic demonstrations of searching and navigating the ECHO website. 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?
- What are ECHO web services?

Please feel free to submit any additional questions into the question box. We will leave time to answer them at the end of the presentation.

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 (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, the Resource Conservation and Recovery Act Information System, and the Safe Drinking Water Information System. 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 would like to remind you to enter any questions you have into the question box. We will have a question and answer session at the end of the presentation.

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

From which EPA program systems does ECHO retrieve data?

ECHO primarily presents enforcement and compliance data related to four environmental statutes and the data come from various EPA program data systems. This includes Clean Air Act stationary source data from ICIS-Air, Clean Water Act permitted discharger data from ICIS-NPDES, Safe Drinking Water Act public water system data from SDWIS, and Resource Conservation and Recovery Act hazardous waste handler data from RCRAInfo.

ECHO also offers compliance monitoring and formal enforcement action information at facilities regulated by other environmental statutes.

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 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. 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 several 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.

EPA recently created video tutorials that explain these data flows in more detail, which can be found on the ECHO Tutorials webpage.

These tutorials illustrate how ECHO obtains and displays compliance data for Clean Air Act stationary sources, Clean Water Act wastewater dischargers, and RCRA hazardous waste handlers.

There are sometimes errors or inconsistencies in the way data are transferred between systems. Since ECHO pulls data from the 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.

Now, let's start our first example.

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 600 facilities. Let's look specifically for facilities with recently identified violations by checking the Facilities with Current Violations box on the right side of the page.

This narrowed our search down to four facilities. 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 Corporation; 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 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. Looking at the first row, the "System" column indicates the data came from EPA's ICIS-Air system, and the "Lead Agency" column indicates that the EPA reported the inspection. Entries that are italicized indicate that they are not counted in EPA compliance monitoring strategies or annual results.

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 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 (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 in 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 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 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?

Data Updates

On this slide, we provide 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.
- 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.

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

If you want to know how current the data are, 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 very 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; some natural lags may exist as data makes its way from various sources into state and EPA program data systems.

Corrections

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, 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.

Some of you may use the State Dashboards. ECHO has recently modernized two new dashboards: Air and Hazardous Waste. ECHO no longer uses frozen data on these modernized dashboards. For both of

these dashboards, data is refreshed weekly and shows the latest available data from the program system.

For the other dashboards, ECHO uses frozen data. Frozen data provide a static snapshot of the data at the time they were frozen. As a note, fiscal year data used in the State Dashboards are typically frozen at the beginning of the following year (e.g., fiscal year 2018 data were frozen in early 2019). Data for the current fiscal year are updated until the time that they are frozen.

How can I download ECHO data?

ECHO facility searches and reports generally provide the most recent 5 years of compliance and enforcement data. The Data Downloads page offers downloads of data sets by program, which contain enforcement and compliance data since the program inception. These may be of particular use to developers, programmers, academics, and analysts.

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.

If you scroll down further on the page you will see downloads by EPA program. For example,

- Facility Registry Service (FRS) Facilities and Linkages dataset, which includes facility identification data from EPA's FRS service
- 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),
- ICIS-NPDES Biosolids Facility dataset, which includes facility and compliance information for NPDES permits in the ECHO Biosolids facility universe,
- RCRAInfo dataset, which includes hazardous waste handler data
- ICIS FE&C dataset (which includes EPA enforcement action data), and
- SDWA dataset, which includes a national download of Safe Drinking Water Act Data supporting the Drinking Water Dashboard.

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.

Okay, before we move onto our last topic, ECHO web services, I want to remind everyone to submit any questions they have into the question box. We will answer them at the end of the webinar.

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 utilizing 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 "Execute" 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.

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.

ECHO has extensive help documentation. Look out for the help links here on the homepage and on each application page.

Furthermore, if you have a question that is not answered by the help pages, you can reach us using the Contact Us link, which is available at the top right of every ECHO page.

A recording of this webinar, as well as the presentation slides and transcript, will be posted to the ECHO Training page.

As a reminder, we recently published three new tutorials about how ECHO data flows. This includes how data are collected, refreshed, and displayed on the website. Please be sure to view these tutorials if you would like more information.

We have included links to helpful ECHO webpages in today's webinar slides.

Now let's take your questions.

Our first question is, what do the yellow triangles mean on the Detailed Facility report?

We use yellow triangles throughout the site as alert messages. So if you see a yellow triangle, if you hover your mouse over that, there will be an additional message providing some supplementary information or extra information that we want to make sure our users see. As you saw in the demonstration of the error correction, we also use the same yellow triangle icon as a link to the Error Correction report as well.

Our second question is, if I have a list of FRS IDs I am interested in researching, can I download multiple Detailed Facility reports at once?

The answer is yes. Tanvi, do you still have the DFR page open? Ok. So, the URL for the report is made up of the base URL where you have ECHO, Detailed Facility report, and ID number. That ID number is the FRS ID. What you can do is append additional IDs to this URL separated by commas and without spaces.

What that will do is load multiple DFRs on one webpage. And if you need any help with that, you can contact us through the link at the top of the page.

Our next question has to do with error corrections. If NPDES permit limits are incorrect in ICIS, can that be corrected?

So, generally yes. You can use the report an error feature to report any inconsistencies or potential errors that you see through ECHO. That will connect you with a data steward who can fully evaluate the reported error and look at the data and assess the next steps. And that could either be a correction or other clarification. The report an error functionality in ECHO is another way to connect you with an appropriate data steward that can fully evaluate any request.

Our next question: Is the FRS number specific to ECHO? For example, is it different than the ICIS number?

The FRS number, which is the Facility Registry Service ID, is an EPA identifier so it is not specific to ECHO, it is used across EPA. You'll also see it on EPA's EnviroFacts website, among other things, and it's a unique identifier across all programs. So for an ICIS number, or a NPDES permit ID, which might be the unique identifier in a NPDES program, there's also an FRS ID that goes along with that NPDES ID, and that FRS number can be used then to link that NPDES ID to any program. Like if the facility with a NPDES permit is a hazardous waste handler, that way we can identify the set of permits or set of regulated programs that a facility is covered under. You'll see the FRS number is always a 12-digit numeric identifier.

The next question: If I wanted a spreadsheet of all facilities in an area, with latitudes and longitudes, where would I get this?

So there's a few options. If we could go to the ECHO home screen, we could point those out. Thank you. The first thing is you can run a Facility Search. You can do an all data facility search for a particular location, like for a county, city, or state. Or you can any data search, but all data will be the most general. Once you're on the search results page, you can go to the customized columns and select latitude and longitude as table columns in the results and then download that information. That works really well for a specific location, maybe a city, maybe a state. If you are looking for nationwide data, we suggest you go to the Data Downloads and use one of those files that have already been pre-generated, which most of them include lat-long information. But the Help page will clearly identify which files have location information included.

The next question: If we want to do a search for an entire company with many facilities, would a web services query be the best approach?

So, we would suggest if you are looking for information and you are not a web developer, you just want to identify and maybe download information from the ECHO website, use the ECHO Facility Search and use the Search interfaces that we provide for you. This will be the easiest way to obtain information. If you wanted to build a website and use some of the ECHO data to present on your custom website, then the web service query may be a good approach because it allows you to pull in all that information and live data and the ECHO website itself is using these web services that we built to support the searches. So, use the interface first, and if you have custom needs or custom applications, then take a look at our web services and of course if you have any questions about how to use those, please contact us.

Ok. Another question about the web services: Is there a limit to how much data can be queried at once? And if so, how is this limit determined?

Yes, some of the services, especially the search services, have a limit, and if you exceed that limit you will receive an error message to let you know there's kind of an upper limit—I believe it's 100,000 facility records—that can be queried at once. It's the same limit we have in place in the user interface as well. Those limits are just in place because there's a massive amount of data and we want to make sure that those searches are available, and so that one person searching doesn't slow down other users. So you should receive a very clear message if some exceedance is met. And, if that poses a problem, please again contact us and we can help you work through that or work through other approaches to help you get the data you need.

Ok. The next question is specific to population with drinking water systems. The question is, is the population provided as zero or the value five, is that because there is no population data specifically applying to the tribal water system?

Rebecca, are you able to field this question? Hi Eva, yes this is Rebecca. Let me track that one down. Ok, so it would be helpful on this one if we had specific examples just so we could definitely give an accurate answer. So, if you want to contact us on that link with specific examples we'd be happy to answer it. But, generally, if there is no data then the field would be blank or null. If there is actually a zero or a five, then in the area surrounding that system there's not a large known population. But, just to make definite sure that we are giving you an accurate answer, please feel free to contact us from the webpage and we will answer you right away. Thank you.

Thanks, Rebecca. Our next question I believe refers to the Facility Search. Can you show how to select one industry sector for analysis?

Could we go to the Facility Search page? Ok. So if we scroll down below the Geographic Location section, we'll get to Facility Characteristics, and here we have some options for industry sectors identified either by SIC code, which is the Standard Industrial Classification code, and NAICS code. SIC code is used widely for the NPDES permit program. So, if you are looking for a NPDES permit, I'd suggest searching on a two- to four-digit SIC code. But most other programs use the NAICS code as their classification system. So, you could use either of these search criteria to search for a specific industry, and we have lookups as well so you can look up what industry is associated with each numeric code. And as you'll see we have the option of NAICS even searching by keyword versus code. So that's one way to look at industry sector and look at facilities that are classified within it. There's also other criteria in Facility Characteristics that may be of interest.

Ok. Next question also has to do with the Facility Search, but this one would be specific to water, and that's is there a quick way to find the facility design flow in millions of gallons per day for water or wastewater facilities?

For the NPDES program there is information on design flow that is captured in a NPDES permit application that we show on the results page. So, if you specifically run a water search, and then once the search results table loads, again go to customize columns and there'll be a few fields for flow and one of those is facility design flow. That's a first and fastest way to look for that information.

Do we have other questions? Eva, this is Rebecca. We do not have other questions.

Great, thank you Rebecca and Eva. This is Madeline LaPatra and I conduct training and outreach for ECHO at the EPA. On behalf of all of us involved with this training, we thank you for participating in this webinar. If you think of any additional questions about using ECHO, please feel free to contact us using the contact us link in the top right of any ECHO page. I also wanted to remind you that a brief survey will open up as soon as this webinar ends. We would really appreciate your feedback. Thank you again and I hope you have a great week.