



Environmental Dataset Gateway Search and Discover 101

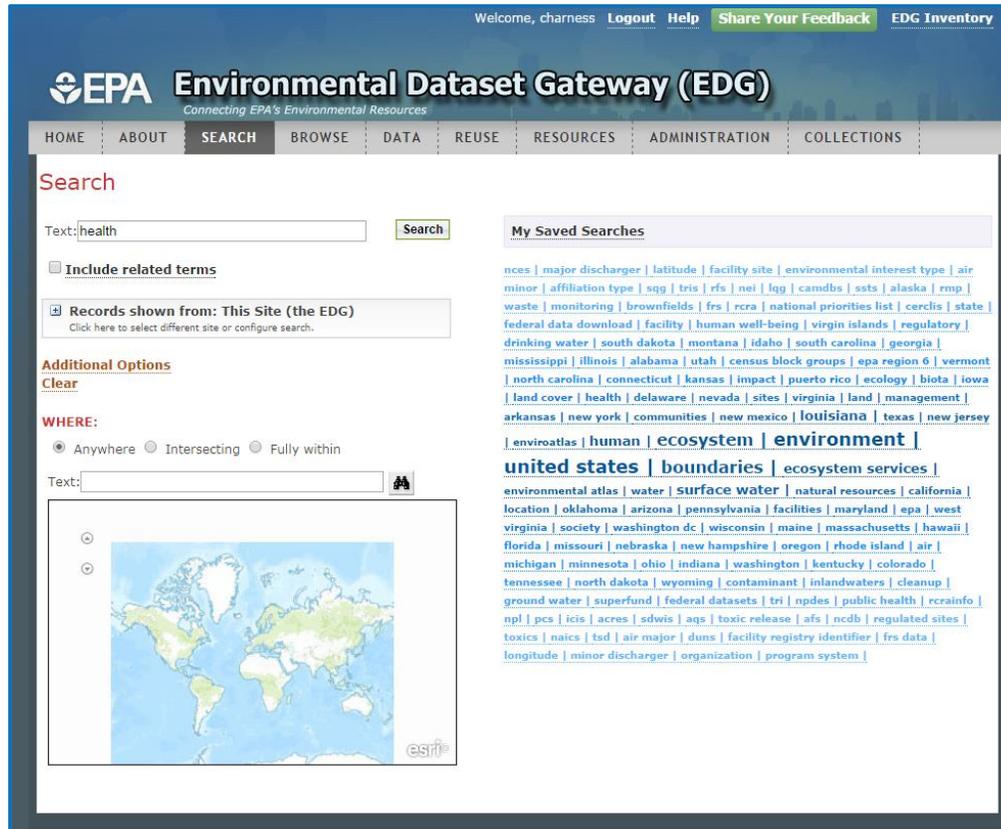
Presented by:

Ana Greene, EDG Program Manager, EPA, OEI

Catherine Harness, Innovate! Inc.

January 2016

Advanced Search Options



Welcome, charness [Logout](#) [Help](#) [Share Your Feedback](#) [EDG Inventory](#)

EPA Environmental Dataset Gateway (EDG)
Connecting EPA's Environmental Resources

HOME ABOUT **SEARCH** BROWSE DATA REUSE RESOURCES ADMINISTRATION COLLECTIONS

Search

Text:

Include related terms

Records shown from: This Site (the EDG)
Click here to select different site or configure search.

Additional Options
Clear

WHERE:

Anywhere Intersecting Fully within

Text:



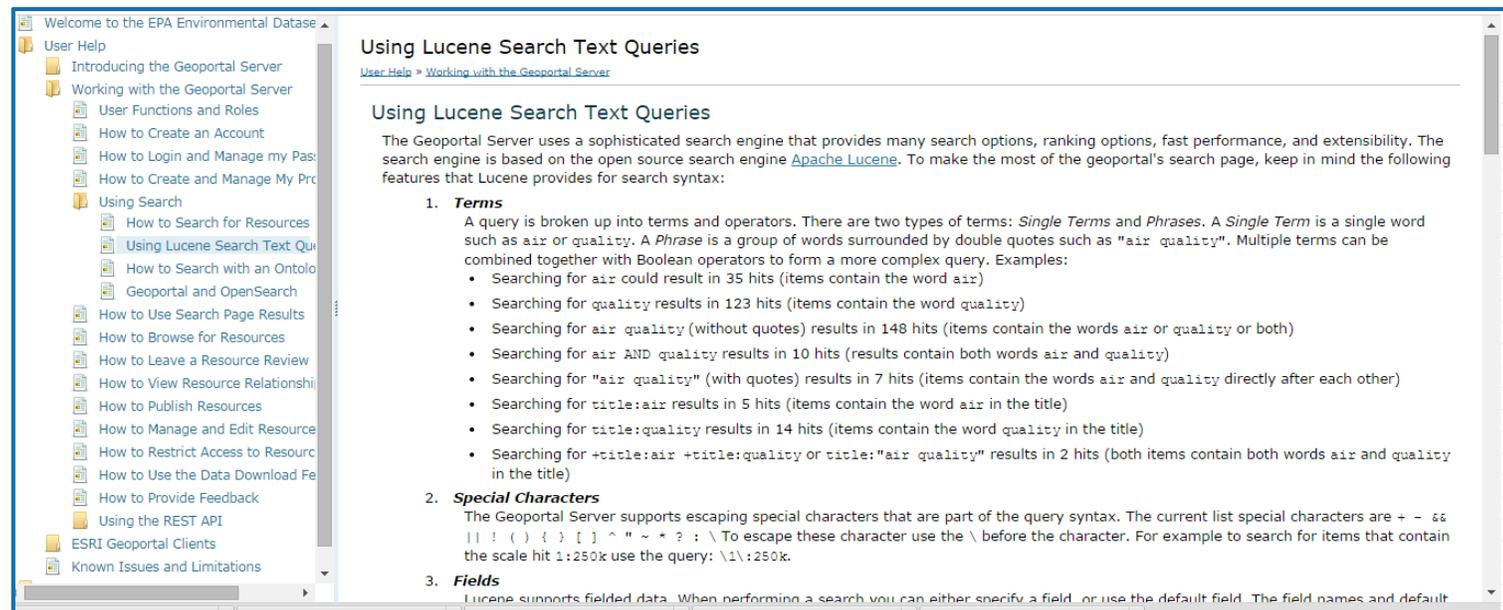
My Saved Searches

[nces](#) | [major discharger](#) | [latitude](#) | [facility site](#) | [environmental interest type](#) | [air minor](#) | [affiliation type](#) | [sgg](#) | [tris](#) | [rfs](#) | [nei](#) | [lqg](#) | [camdbs](#) | [ssts](#) | [alaska](#) | [rmp](#) | [waste](#) | [monitoring](#) | [brownfields](#) | [frs](#) | [rcra](#) | [national priorities list](#) | [cerclis](#) | [state](#) | [federal data download](#) | [facility](#) | [human well-being](#) | [virgin islands](#) | [regulatory](#) | [drinking water](#) | [south dakota](#) | [montana](#) | [idaho](#) | [south carolina](#) | [georgia](#) | [mississippi](#) | [illinois](#) | [alabama](#) | [utah](#) | [census block groups](#) | [epa region 6](#) | [vermont](#) | [north carolina](#) | [connecticut](#) | [kansas](#) | [impact](#) | [puerto rico](#) | [ecology](#) | [biota](#) | [iowa](#) | [land cover](#) | [health](#) | [delaware](#) | [nevada](#) | [sites](#) | [virginia](#) | [land](#) | [management](#) | [arkansas](#) | [new york](#) | [communities](#) | [new mexico](#) | [louisiana](#) | [texas](#) | [new jersey](#)

[enviroatlas](#) | [human](#) | [ecosystem](#) | [environment](#) | [united states](#) | [boundaries](#) | [ecosystem services](#) | [environmental atlas](#) | [water](#) | [surface water](#) | [natural resources](#) | [california](#) | [location](#) | [oklahoma](#) | [arizona](#) | [pennsylvania](#) | [facilities](#) | [maryland](#) | [epa](#) | [west virginia](#) | [society](#) | [washington dc](#) | [wisconsin](#) | [maine](#) | [massachusetts](#) | [hawaii](#) | [florida](#) | [missouri](#) | [nebraska](#) | [new hampshire](#) | [oregon](#) | [rhode island](#) | [air](#) | [michigan](#) | [minnesota](#) | [ohio](#) | [indiana](#) | [washington](#) | [kentucky](#) | [colorado](#) | [tennessee](#) | [north dakota](#) | [wyoming](#) | [contaminant](#) | [inlandwaters](#) | [cleanup](#) | [ground water](#) | [superfund](#) | [federal datasets](#) | [tri](#) | [npdes](#) | [public health](#) | [rcrainfo](#) | [npl](#) | [pcs](#) | [icis](#) | [acres](#) | [sdwis](#) | [ags](#) | [toxic release](#) | [afs](#) | [ncdb](#) | [regulated sites](#) | [toxics](#) | [naics](#) | [tsd](#) | [air major](#) | [duns](#) | [facility registry identifier](#) | [frs data](#) | [longitude](#) | [minor discharger](#) | [organization](#) | [program system](#)

- Using Lucene Search Text Queries
- Include related terms
- Keyword Hints
- Search additional sites
- Additional Options
- Interactive Map
- Tag search

Advanced Search Options: Using Lucene Search Text Queries



The screenshot shows a web browser window displaying the EPA Environmental Data Search help page. The left sidebar contains a navigation menu with categories like 'User Help', 'Working with the Geoportal Server', 'Using Search', and 'Using the REST API'. The main content area is titled 'Using Lucene Search Text Queries' and includes an introduction and three numbered sections: 'Terms', 'Special Characters', and 'Fields'.

Using Lucene Search Text Queries

The Geoportal Server uses a sophisticated search engine that provides many search options, ranking options, fast performance, and extensibility. The search engine is based on the open source search engine [Apache Lucene](#). To make the most of the geoportal's search page, keep in mind the following features that Lucene provides for search syntax:

- Terms**

A query is broken up into terms and operators. There are two types of terms: *Single Terms* and *Phrases*. A *Single Term* is a single word such as `air` or `quality`. A *Phrase* is a group of words surrounded by double quotes such as `"air quality"`. Multiple terms can be combined together with Boolean operators to form a more complex query. Examples:

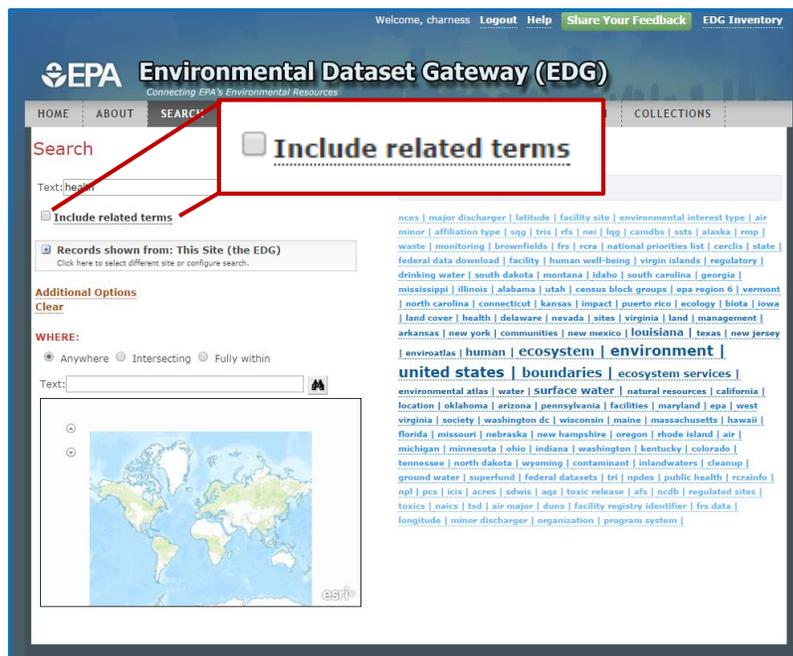
 - Searching for `air` could result in 35 hits (items contain the word `air`)
 - Searching for `quality` results in 123 hits (items contain the word `quality`)
 - Searching for `air quality` (without quotes) results in 148 hits (items contain the words `air` or `quality` or both)
 - Searching for `air AND quality` results in 10 hits (results contain both words `air` and `quality`)
 - Searching for `"air quality"` (with quotes) results in 7 hits (items contain the words `air` and `quality` directly after each other)
 - Searching for `title:air` results in 5 hits (items contain the word `air` in the title)
 - Searching for `title:quality` results in 14 hits (items contain the word `quality` in the title)
 - Searching for `+title:air +title:quality` or `title:"air quality"` results in 2 hits (both items contain both words `air` and `quality` in the title)
- Special Characters**

The Geoportal Server supports escaping special characters that are part of the query syntax. The current list special characters are `+ - & | ! () { } [] ^ " ~ * ? : \`. To escape these character use the `\` before the character. For example to search for items that contain the scale hit `1:250k` use the query: `\1\:250k`.
- Fields**

Lucene supports fielded data. When performing a search you can either specify a field, or use the default field. The field names and default

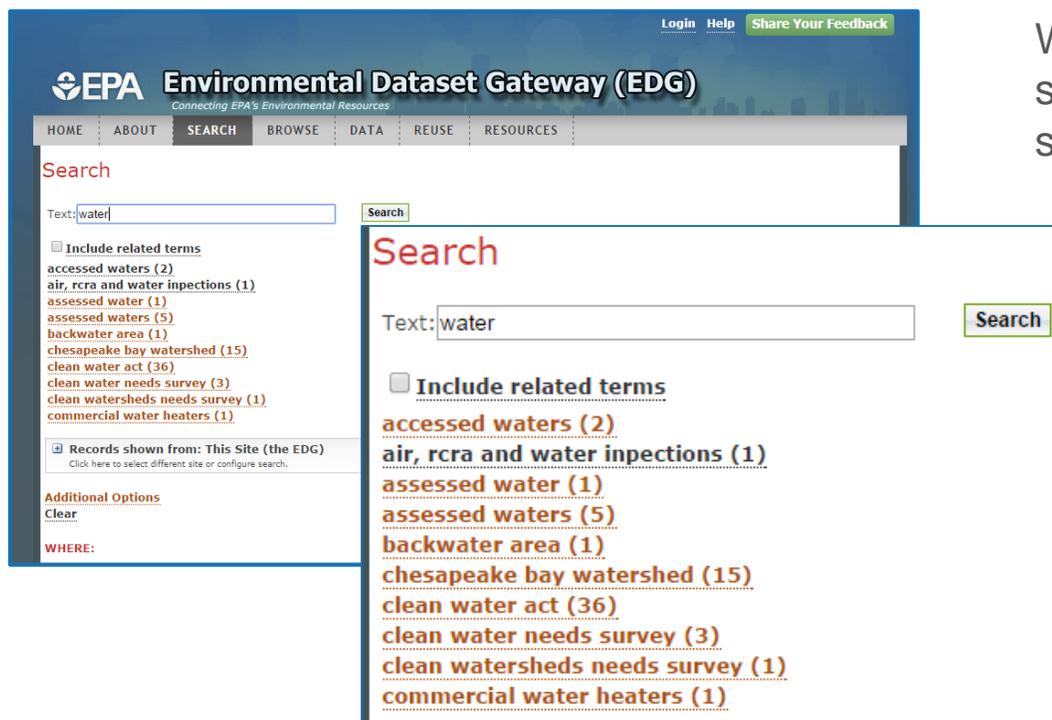
[Link to Detailed Help with Using Lucene Search Text Queries](#)

Advanced Search Options: Include related terms



Search will include acronyms, synonyms and other related terms

Advanced Search Options: Keyword Hints



The screenshot displays the EPA Environmental Dataset Gateway (EDG) search interface. The search term "water" is entered in the search box, and a list of keyword hints is displayed below it. The hints include "accessed waters (2)", "air, rcra and water inspections (1)", "assessed water (1)", "assessed waters (5)", "backwater area (1)", "chESApeake bay watershed (15)", "clean water act (36)", "clean water needs survey (3)", "clean watersheds needs survey (1)", and "commercial water heaters (1)".

Search

Text: water

Include related terms

- [accessed waters \(2\)](#)
- [air, rcra and water inspections \(1\)](#)
- [assessed water \(1\)](#)
- [assessed waters \(5\)](#)
- [backwater area \(1\)](#)
- [chESApeake bay watershed \(15\)](#)
- [clean water act \(36\)](#)
- [clean water needs survey \(3\)](#)
- [clean watersheds needs survey \(1\)](#)
- [commercial water heaters \(1\)](#)

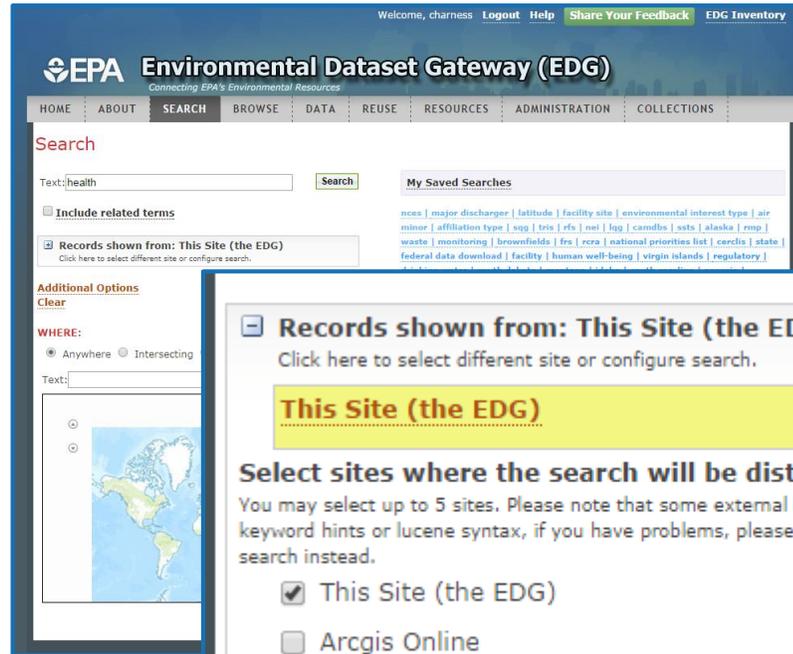
Records shown from: This Site (the EDG)
Click here to select different site or configure search.

Additional Options
[Clear](#)

WHERE:

When a search term is typed in the search text box, keyword hints and suggestions appear below

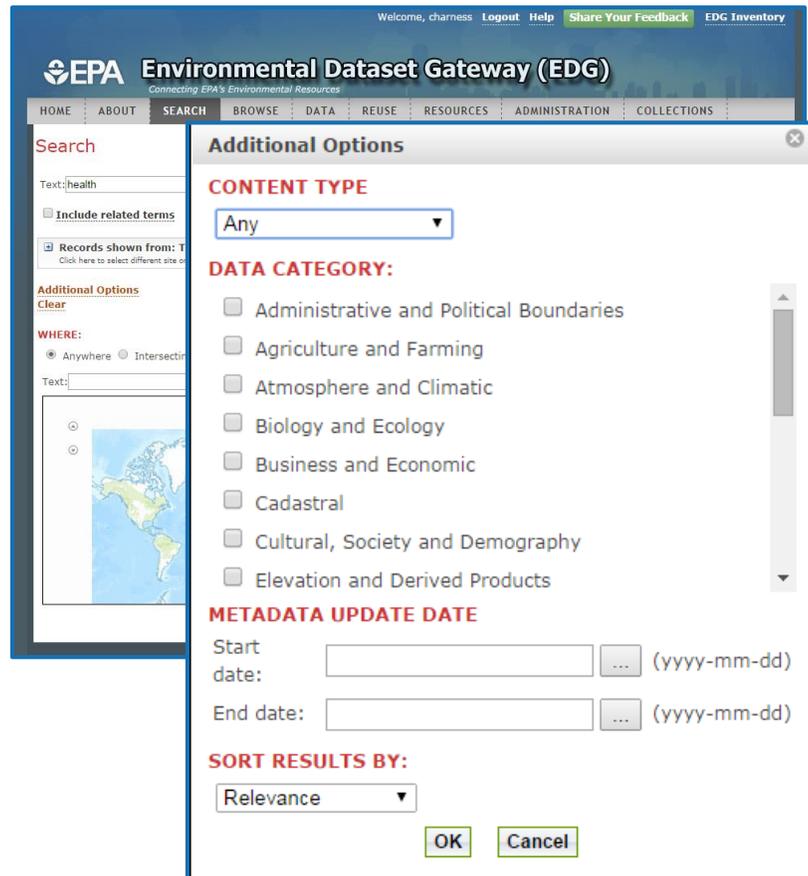
Advanced Search Options: Search Additional Sites



Search more than the EDG

- ArcGIS Online
- Public EPA GeoPlatform Content

Advanced Search Options: Additional Options



The screenshot shows the EPA Environmental Dataset Gateway (EDG) search interface. The main search area is on the left, with a search bar containing the text "health". Below the search bar are options for "Include related terms" and "Records shown from: T". The "Additional Options" dialog box is open, showing the following sections:

- CONTENT TYPE:** A dropdown menu set to "Any".
- DATA CATEGORY:** A list of categories with checkboxes:
 - Administrative and Political Boundaries
 - Agriculture and Farming
 - Atmosphere and Climatic
 - Biology and Ecology
 - Business and Economic
 - Cadastral
 - Cultural, Society and Demography
 - Elevation and Derived Products
- METADATA UPDATE DATE:** Two date input fields with dropdown calendars:
 - Start date: [] (yyyy-mm-dd)
 - End date: [] (yyyy-mm-dd)
- SORT RESULTS BY:** A dropdown menu set to "Relevance".

At the bottom of the dialog box are "OK" and "Cancel" buttons.

Search by Content Type

- 11 content types
- These are properties the EDG uses to identify what kind of resource is being described in a metadata record

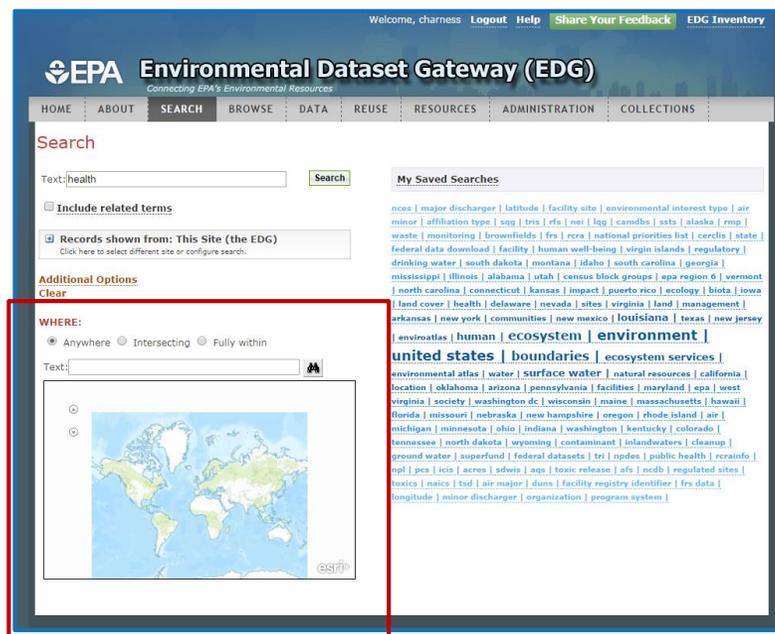
Data Category

- 19 topic categories defined in the ISO 19115 metadata standard
- Can choose to search more than one

Metadata Update Date

- The date a resource's metadata was last modified within the EDG

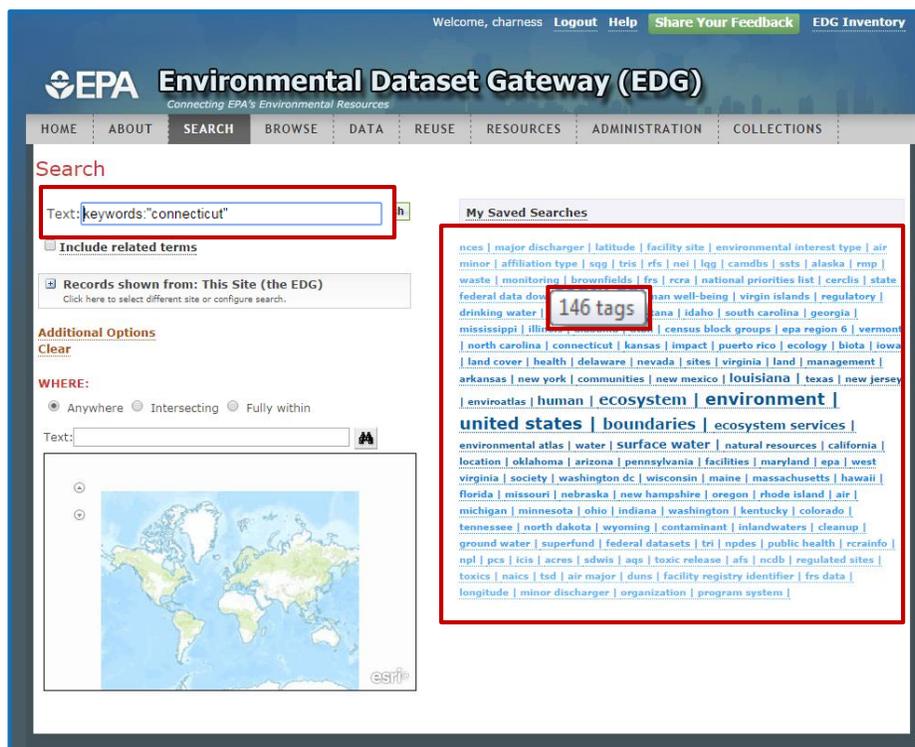
Advanced Search Options: Interactive Map



Define the spatial extent of the records to be returned

- **Locate Tool**
 - Type place name in text box
 - EDG Map zooms in to that area
- **Map Navigation Tools**
 - Pan and Zoom
 - Anywhere – Search results will return resources from any extent
 - Intersecting – Search results will return resources that have some portion of their extent touching or included in the defined area
 - Fully within – search results will return resources that are fully contained in the defined area

Advanced Search Options: Tag Search



Welcome, charness Logout Help Share Your Feedback EDG Inventory

EPA Environmental Dataset Gateway (EDG)
Connecting EPA's Environmental Resources

HOME ABOUT **SEARCH** BROWSE DATA REUSE RESOURCES ADMINISTRATION COLLECTIONS

Search

Text:

Include related terms

Records shown from: This Site (the EDG)
Click here to select different site or configure search.

Additional Options
Clear

WHERE:
 Anywhere Intersecting Fully within

Text:



My Saved Searches

146 tags

nces | major discharger | latitude | facility site | environmental interest type | air
minor | affiliation type | sqg | tris | rfs | nei | lqg | camdbs | sts | alaska | rmp |
waste | monitoring | brownfields | frs | rcra | national priorities list | cerclis | state
federal data download | human well-being | virgin islands | regulatory |
drinking water | hawaii | idaho | south carolina | georgia |
mississippi | illinois | census block groups | epa region 6 | vermont
north carolina | connecticut | kansas | impact | puerto rico | ecology | biota | iowa
land cover | health | delaware | nevada | sites | virginia | land | management |
arkansas | new york | communities | new mexico | louisiana | texas | new jersey
| enviroatlas | human | ecosystem | environment |
united states | boundaries | ecosystem services |
environmental atlas | water | surface water | natural resources | california |
location | oklahoma | arizona | pennsylvania | facilities | maryland | epa | west
virginia | society | washington dc | wisconsin | maine | massachusetts | hawaii |
florida | missouri | nabaska | new hampshire | oregon | rhode island | air |
michigan | minnesota | ohio | indiana | washington | kentucky | colorado |
tennessee | north dakota | wyoming | contaminant | inlandwaters | cleanup |
ground water | superfund | federal datasets | tri | npdes | public health | rcrainfo |
npl | pca | icis | acres | sdwis | aqs | toxic release | afa | ncob | regulated sites |
toxics | naics | tsd | air major | duns | facility registry identifier | frs data |
longitude | minor discharger | organization | program system |

Search metadata keywords/tags

- List of keywords found in metadata records.
- Hovering over the Keyword in the list will give the number of records using the keyword/tag
- Click the keyword and the search is populated in the search box and the search is performed

DEMO



Search and Discover 101:

Important URLs

- Using Lucene Search Text Queries:
https://edg.epa.gov/metadata/webhelp/en/gptlv10/index.html#/Using_Lucene_Search_Text_Queries/00t000000004000000/
- EDG about page:
<https://edg.epa.gov/metadata/catalog/content/about.page#>
- EDG Reuse Document:
https://edg.epa.gov/metadata/webhelp/en/gptlv10/inno/EDG_Reuse.pdf
- Questions/Help, email: edg@epa.gov



Questions?

Ana Greene

EDG Program Manager

Greene.Ana@epa.gov