

User Help for the USGS Science Data Catalog (BETA 1.0)

The index for the USGS Science Data Catalog is built on standardized *metadata* written by the research project teams to describe their data. Successful search and retrieval on data in this Catalog will be influenced by the completeness and comprehensiveness of these metadata, as well as the degree to which terms chosen by metadata authors to describe their data match terms chosen by end users when searching the Catalog.

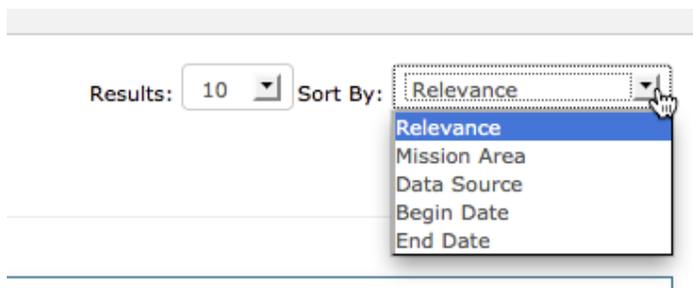
Datasets described in this Catalog are available online, via a Web page interface, direct download, or online service. The Catalog does not include USGS data for which there are currently no online access points.

The default Home Page view of the Catalog displays ALL records in the index; users can then hone in on specific records of the interest by one or more techniques:

- searching by term or concept using the text box search
- restricting by geographic polygons using the “limit search by location” option
- filtering by Theme Keyword, Mission Area, Data Source, or Scientist using the facets in the left column

ORDERING, SORTING, & DISPLAY OF RESULTS

By default, results are sorted by Relevance ranking. You can re-sort results on other parameters using the Sort By pulldown menu:



In the same area of the screen you can change the number of dataset results per page from the default of 10 to 50 or 100 per page.

TYPES OF DATA AVAILABLE

All items in the Science Data Catalog contain a link to data or data service, accessible via the colored button(s) beneath the brief description of the dataset. Sometimes the link will take you to a downloadable dataset, which may take the form of a zip file, an individual spreadsheet, a topographic map, or a USGS Data Series Report. Other links may take you to an online data system containing hundreds or even thousands of individual datasets, where you will be prompted to execute a more specific search.

Finally, a number of data sources in the USGS provide access to data via APIs and services, including map services.

The screenshot shows a data record for "U.S. Geological Survey Gap Analysis Program Species Ranges". The record is marked as a "Featured Dataset". Key information includes:

- Data Source:** Core Science Analytics and Synthesis
- Mission Area:** Core Science Systems
- Begin Date:** Unknown
- End Date:** Present

The description states: "GAP species range data show a coarse representation of the total areal extent of a species or the geographic limits within which a species can be found (Morrison and Hall 2002). To represent these [more](#)". Below the description are three buttons: "html", "html", and "csv", followed by a "View Metadata" button. Three callout boxes provide additional context:

- The first callout points to the "View Metadata" button, stating: "Get preview and full original metadata that describes this dataset".
- The second callout points to the first "html" button, stating: "Link to a guided data download site".
- The third callout points to the second "html" button, stating: "Link to page providing ancillary data".
- The fourth callout points to the "csv" button, stating: "Prompts FTP download of CSV-formatted data file".

FEATURED DATASETS

You may notice that a resource in your results set has a blue flag noting that it is a "Featured Dataset." These data constitute some of the major, well-known, national datasets published by the USGS. Frequently these are large, searchable systems containing many individual datasets. Because the number of individual datasets contained within these systems could easily overwhelm the Science Data Catalog index, these systems are represented in this Catalog through single records. A system record for a database flagged as a "Featured Dataset" will link to a specialized, guided interface that will take you to very specific datasets of interest.

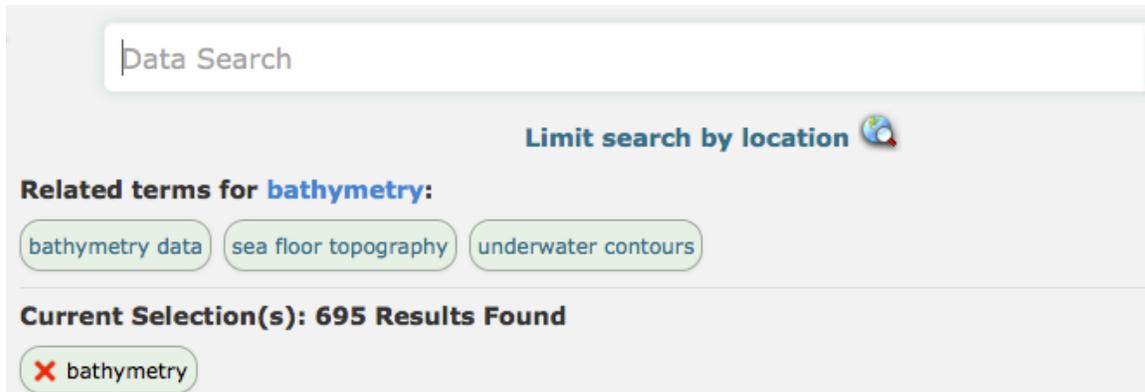
SEARCHING THE CATALOG USING THE TEXT BOX SEARCH

Let's say you're interested in USGS bathymetry datasets, so you type the term into the text search box:

The screenshot shows the search interface with the term "bathymetry" entered in the search box. Below the search box is a button labeled "Limit search by location" with a location icon. The results section shows "Current Selection(s): 4451 Results Found" and "All Catalog Holdings".

Remember, from the Catalog home page, you are seeing "All Catalog Holdings" in the results. Your query is going to reduce the number of results to those that contain your query.

The Catalog returns the following response:



The screenshot shows a search interface with a search box containing the text "Data Search". Below the search box is a button labeled "Limit search by location" with a globe icon. Underneath, there is a section titled "Related terms for bathymetry:" with three buttons: "bathymetry data", "sea floor topography", and "underwater contours". Below this is a section titled "Current Selection(s): 695 Results Found" with a button labeled "X bathymetry".

The total number of datasets in the Catalog has been reduced to those containing the term 'bathymetry' in one of the key metadata fields given priority in the indexing process.

You have a few choices for proceeding from this set of results. You can

- **Browse** through the results presented in the list of relevant datasets.
- Use a **Filter** from the list in the left column to add another parameter that will further narrow your search (this will AND your term 'bathymetry' to the filter selected).
- Add an **additional text query** term to 'bathymetry' in the text search box (this will AND the added term to your existing term, 'bathymetry').
- Create a **geographic polygon** that will restrict your search for 'bathymetry' to datasets for a certain location.
- Use the "**Related Terms**" to explore related concepts to your term 'bathymetry.'
- Delete your 'bathymetry' query and start over with a brand **new search**.

ADDING AN ADDITIONAL QUERY TO YOUR TEXT SEARCH

In the previous example, a text search on bathymetry produced 695 results, too many items to easily scan. Perhaps you are interested specifically in bathymetric data for Monterey Bay. You can search within the original 'bathymetry' results set by typing an additional query into the text search box:

Monterey Bay

Adding an additional query to search within the original subset of 695 records found for 'bathymetry';

Related terms for bay:

Current Selection(s): 695 Results Found

✕ bathymetry

Initial query and results set

Related terms for Bay:

Current Selection(s): 7 Results Found

✕ bathymetry AND Monterey Bay

The addition of 'Monterey Bay' was ANDed to the original query statement to greatly reduce the returned set to only 7 specific records.

These 7 final results include records with both terms appearing in one or more key metadata fields.

STARTING A BRAND NEW SEARCH

To start a new search, you will first need to clear out any previous search parameter displayed. The easiest way to confirm that you are executing a new search from the default state is to make sure that the words “All Catalog Holdings” displays beneath “Current Selection(s)” and that there are no query statements visible:

Data Search

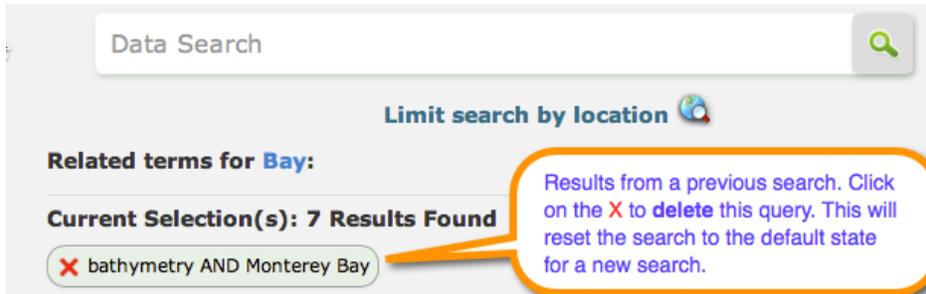
Limit search by location

Current Selection(s): 4451 Results Found

All Catalog Holdings

All holdings in the Catalog are displayed, and no query has yet been run to restrict the number of records displayed.

If a previous search has been executed but not yet cleared, you will see this:

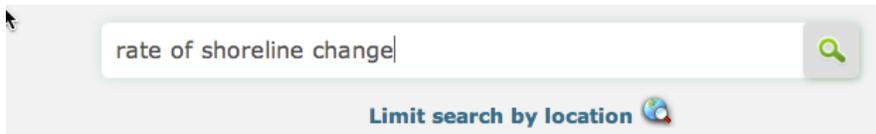


Delete all query strings by clicking on the **X** to remove them. This will reset the search to the default state for a new query.

STRUCTURING A COMPOUND QUERY STATEMENT

By default, compound queries are ANDed together to narrow the results set. In other words, the returned results must contain all of your query terms in the key metadata fields.

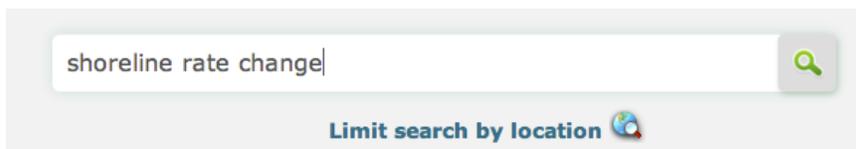
For example, if you are looking for datasets that address the degree of the nation's shoreline change, you might try the following text query:



Which might produce the following results set:



Or you could have worded it slightly differently to eliminate the 'of':



Notice that the results set is the same.

Related terms for rate:

Current Selection(s): 261 Results Found

✗ shoreline change rate

The search simply ANDs all of the concepts together, ignoring the 'of' in the first example. All records that contained all three terms (shoreline AND rate AND change) in any of the key metadata fields were matched and returned. The matching records may or may not contain the exact phrases "rate [of] shoreline change" or "shoreline change rate"; they may instead contain the words 'rate' and 'change' in the title, and contain 'shoreline' somewhere in the Description field, for example. This means that some of the results in our example above may not deal specifically with rates of shoreline change, but could instead include a record on rates of change in seagull populations in shoreline versus urban areas.

If you are looking for a very specific concept, it often helps to use quotation marks around the phrase – a bound query – to search for the exact wording occurring in one or more key metadata fields.

From our previous example:

The exact query "rate of shoreline change"



"rate of shoreline change" 

[Limit search by location](#) 

Produces very different results than it did without the quotation marks around the phrase.

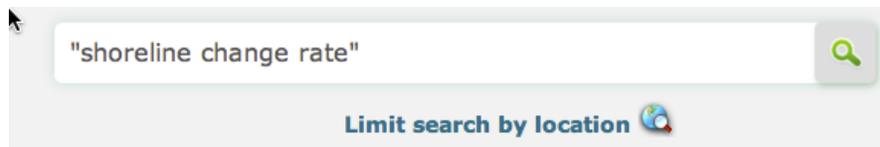
Related terms for change":

Current Selection(s): 3 Results Found

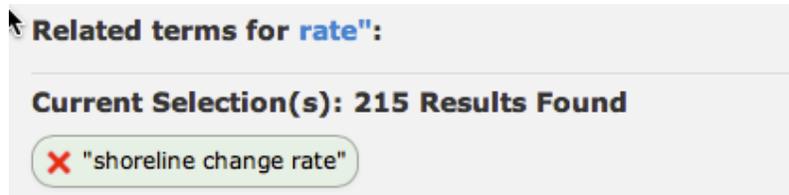
✗ "rate of shoreline change"

In this case, all 3 items found contain the phrase "rate of shoreline change" within the Description field for the respective datasets.

It's a good idea to try potential variations on a bound phrase search to find all possible conceptual matches.



In this second bound phrase example, the number of returned results containing this exact phrasing in one or more key metadata fields is much larger: 215!



The exact phrase appears in titles and/or Description fields for these records.

GEOSPATIAL SEARCHING TECHNIQUES

There are two ways that the geospatial aspects of the data are described in the metadata:

- **Geospatial keywords** that describe the area of study (e.g. Alaska, Green River, Bakken Formation, Acadia National Park, Gulf of Mexico). You can search for geospatial keywords in key metadata fields by entering them into your text query.
- **Bounding coordinates** that give the limits of coverage of a dataset: western-most, eastern-most, northern-most, and southern-most. You can restrict your search results to a certain general geographic area by drawing a bounding box on a map, alone or in combination with other search parameters.

When should I use the geospatial keyword search technique?

Geospatial keyword searches using the text box search are recommended when you are looking for data from a very specific location. Research studies that are very localized are usually described by metadata that contain those location terms in the Title, Description, and Place Keyword fields. Because these fields are part of the search index, a text search will return them, and they should appear at or near the top of relevance-ranked results.

The key thing to remember is that you are searching on words in the metadata. The Catalog currently does not translate your text place keywords into coordinates, so your keywords must match the textual metadata in order to return search results. Bounding coordinates in the dataset metadata must be searched using define-a-polygon search that is executed by clicking on “limit search by location” beneath the text search box.

When should I use the bounding coordinates search technique (“limit search by location”)?

Searching against the bounding coordinates in the metadata is recommended when you are looking for datasets for a more general geographic area, and not a very specific *named place*. You can use a ‘search by location’ alone, or in conjunction with a text search or a facet search.

The ‘search by location’ option, when selected, allows you to create a polygon or bounding box to define a specific geographic area of interest. When this feature is activated, you have the choice to pan, zoom, and draw your own bounding box on the map to specify your area of interest, or to use dropdown menus to select pre-defined polygons for U.S. states, for countries, and for oceans and major water bodies of the world.

A brief tour of the ‘limit search by location’ map feature:

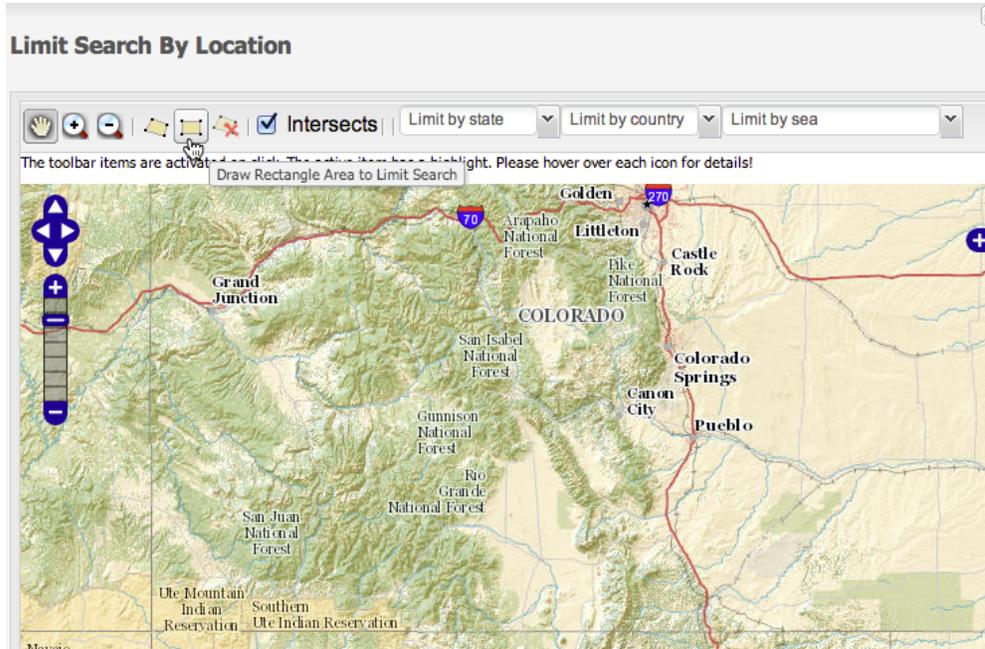
The screenshot shows the 'Limit Search By Location' interface. The title bar reads 'Limit Search By Location'. Below the title bar is a toolbar with icons for pan, zoom in, zoom out, a rectangle tool, a polygon tool, and a delete tool. A checkbox labeled 'Intersects' is checked. To the right of the toolbar are three dropdown menus: 'Limit by state', 'Limit by country', and 'Limit by sea'. The main area is a map of North America. Several callouts provide instructions:

- Pre-defined polygons (easier than drawing your own)**: Points to the dropdown menus.
- Zoom the map in and out**: Points to the zoom in and zoom out icons.
- Re-center the map in the window**: Points to the pan icon.
- Draw your own polygon**
- Draw a bounding BOX
- Delete your polygon/box: Points to the rectangle, polygon, and delete icons.
- If 'Intersects' is unchecked: only datasets with bounding coordinates completely inside your polygon will be returned. (narrower search)**
- If 'Intersects' is checked: datasets with bounding coordinates either completely inside your polygon OR touching a section of your polygon will be returned. (broader search)**

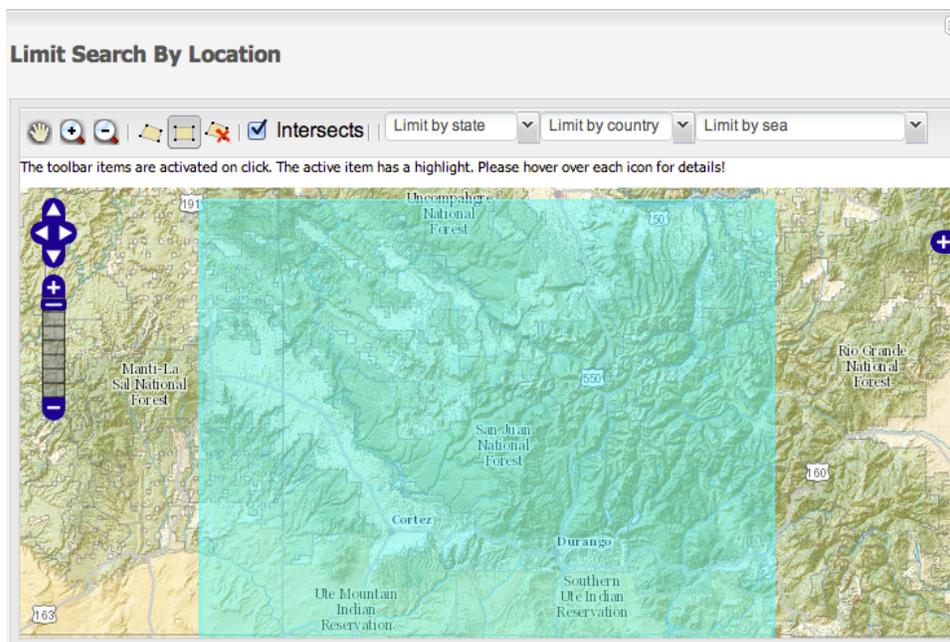
Example uses of the 'limit search by location' feature:

Example #1:

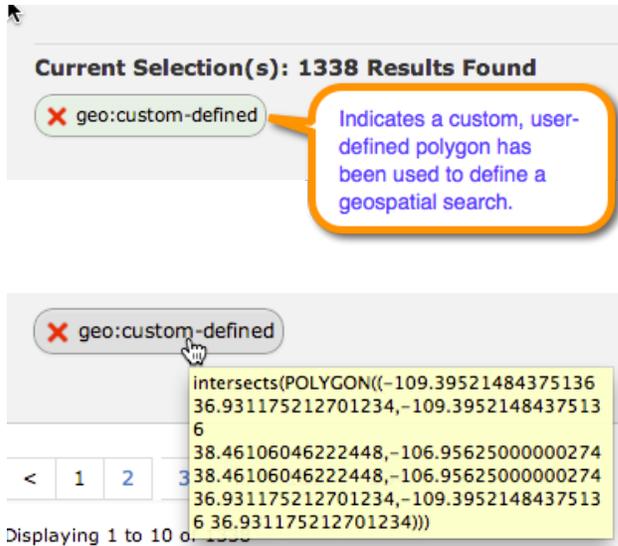
You're interested in USGS datasets related to southwestern Colorado.



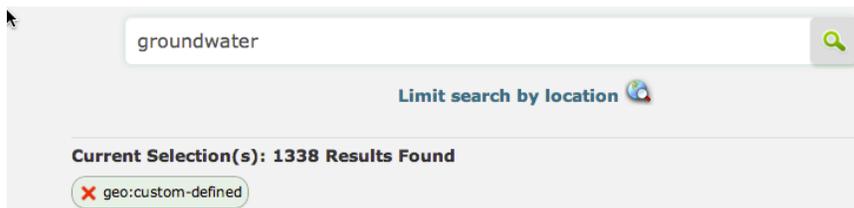
You draw a bounding box to specifically define this area, then click on the X in the upper right corner of the window to close the map box and return to the main screen:



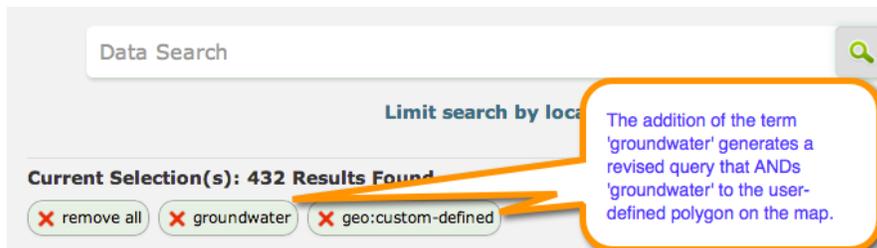
From this specific area of interest in southwest Colorado, the Catalog has located 1338 datasets whose bounding coordinates are either contained within, or overlap the defined area:



Perhaps you're specifically interested in datasets for this region that pertain to groundwater. You can further refine the initial 1338 results returned for this region by submitting a text search for a keyword or phrase:



The addition of the term to the query narrows the results set from 1338 results to 432:



You can add additional terms to further narrow the results.

To **revise** your selected or custom-defined polygon, you must first **remove** your existing geo: selection. In the above example, you would need to click on **X geo:custom-defined**

to delete it; you would then click on 'limit search by location' and select or define a new polygon of interest.

To initiate a completely new search, click on **X remove all**.

Example #2:

The option to 'limit search by location' can also be used in conjunction with the filters (or 'facets') that appear in the left column of the Catalog to focus in on records which contain a specific Theme Keyword or Scientist name in the metadata, or which are produced by a particular USGS Mission Area or program ('Data Source').

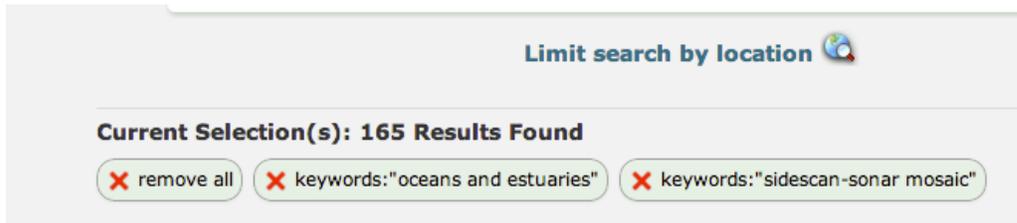
For example, you might be interested in the datasets described by the Keyword phrase 'oceans and estuaries,' so you select it:

The screenshot shows a search results page. On the left, under 'Filter By: Keywords', the 'oceans and estuaries (942)' filter is selected. The main content area displays a 'Featured Dataset' titled 'USGS National Geologic Map Database Collection'. The dataset details include: Data Source: USGS National Geologic Map Database; Mission Area: Core Science Systems; Begin Date: Unknown; End Date: Present. A description states: 'The National Geologic Map Database (NGMDB) is a Congressionally mandated national archive of geoscience maps, reports, and stratigraphic information. According to the Geologic Mapping Act, the U.S. G more'. There are buttons for 'html' and 'View Metadata'.

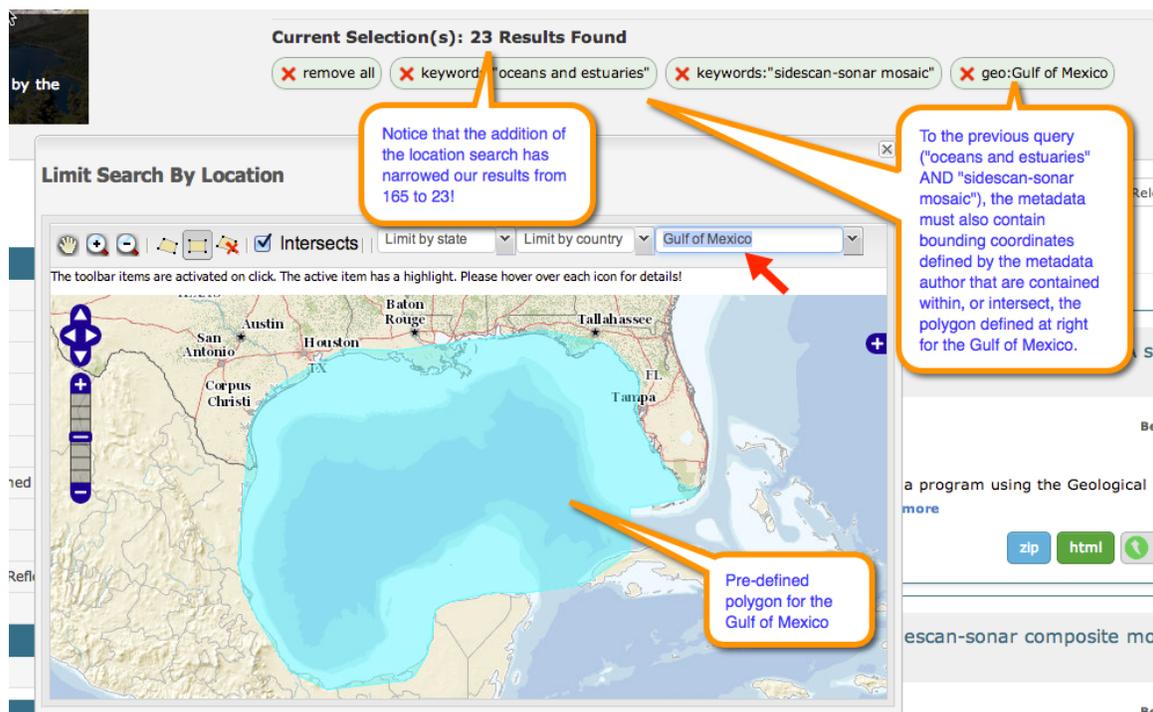
The filters in the left column 'adjust' to show a new selection of filters relevant to all of the records in the Catalog that contain 'oceans and estuaries' as a metadata Theme Keyword. Your attention is directed to the subset of records that also contain the Theme Keyword phrase 'side-scan sonar,' and you click on it to further narrow your results set.

The screenshot shows a search results page with a 'Mission Area: Water Resources' banner. It indicates 'Current Selection(s): 942 Results Found' with a filter for 'keywords:"oceans and estuaries"'. The left 'Filter By: Keywords' list shows 'side-scan-sonar mosaic (165)' selected, indicated by a red arrow. The main content area displays a dataset titled 'JPEG images of seismic data collected offshore of the Chandeleur Islands, LA, 2007'. The dataset details include: Data Source: Coastal and Marine Geoscience Data System; Mission Area: Natural Hazards; Begin Date: 06/09/2007; End Date: 06/21/2007. A description states: 'In 2006 and 2007, the U.S. Geological Survey, in partnership with Louisiana Department of Natural Resources and the University of New Orleans, conducted geologic mapping to characterize the sea floor more'. There are buttons for 'zip', 'html', and 'View Metadata'.

You've now narrowed your results set to 165 records:



Perhaps your particular area of study interest in the waters of the Gulf of Mexico; you want to know what USGS datasets exist for various possible transects in the Gulf. You can further refine the above 'filter' query by adding a geographic search limitation using the 'limit search by location' option. Because the Gulf of Mexico is among the pre-defined polygons in this tool, you can simply select it from the 'Limit by sea' dropdown menu.



This results in a much more manageable subset of records to browse: 23.

HELP WITH TEXT SEARCH KEYWORDS: RELATED TERMS

The challenge with a text search against metadata records created by hundreds of different data producers is anticipating what words they have used to describe their data. Do you use very specific terms, or broader or more generic terms? If more than one word could be used to mean the same thing (i.e. words that are synonymous), do you

have to try all of those synonyms to make sure you haven't missed a potentially useful dataset?

For example, one ecologist may use the term "non-native plants" to describe the type of plant he is studying, while another might call it a "non-indigenous plant" or an "exotic plant."

In the near future, the USGS Science Data Catalog will address this problem by embedding thesauri behind the search interface to automatically search against known synonyms for common scientific concepts; for instance, if you enter "non-native plants," the Catalog will return records with that phrase in the key metadata fields, but it will also return records using "non-indigenous plants" and "exotic plants" as well. *At this time*, however, we suggest that you try running your searches with synonyms known to you entered into the text box search, to ensure that you aren't missing any pertinent datasets.