Tethys Web Client Manual

Tethys, Antioch mosaic, 3rd century from Baltimore Museum of Art

Ally Rice, Scripps Institution of Oceanography

Marie A. Roch, San Diego State University

Table of Contents

1 Introduction 2
2 Simple Queries 3
2.1 Orienting yourself to the controls 3
2.2 Bounding Box Control and Time Filter 3
2.2.1 Time Controls 3
2.3 Species, Site, Project, Deployment Controls 4
2.4 Submit or Refine Queries 5
2.5 Refine queries in advanced viewer, Save Results, Loading Queries 6
2.6 Viewing Results in an R session. 7
2.7 Viewing the Results in MATLAB 7
2.8 Viewing the Results of Simple Queries. 8
3 Advanced Queries 8
3.1 Controls for the Advanced Queries View 9
3.1.1 The Build Tab 9
3.1.2 Viewing Your Results 12
4 Create a Source Map 14
4.1 Connect to a data source 15
4.2 Set the configuration 17
4.3 Navigate the schema 18
4.4 Field mapping 19
4.5 Create source map 23

1 Introduction

This document provides examples of common tasks using the web interface to Tethys. The Tethys Web Client is a web application that provides a more visual interface to a subset of the functionality of the MATLAB client. It provides maps and query tools, allows time and spatial filters to be built into the queries, and provides a set of tools for saving queries for later modification, refining simple queries, and exporting query results in MATLAB, XML, and R data formats. A tool for generating source maps is also available through the web client.

This manual will walk through how to use the web client to form simple and advanced queries, as well as how to use the source map tool. The drop-down menu located at the top-right of the web client page also provides a link to tutorials, as well as lists of the documents currently uploaded to each collection in Tethys (i.e., Calibrations, Deployments, Detections, etc.). By clicking on a document within one of the collections you can view the XML or raw data contained in the document.

2 Simple Queries

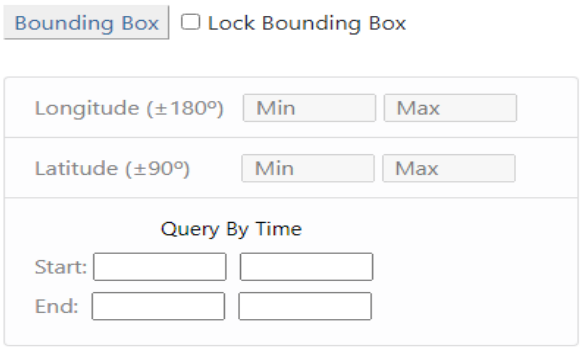
2.1 Orienting yourself to the controls

The web interface will automatically open to the 'Simple Queries' tab. These tab controls are toward the top of the web page under the Tethys banner. There is a Simple Queries and an Advanced Queries tab.

In the Simple Queries pane, all of the controls are aligned vertically alongside the left-hand side of the pane. These controls help to build and view the results of a simple query. Additionally, there are controls for refining the query, saving the query results in different formats, and loading queries that you have saved.

2.2 Bounding Box Control and Time Filter

The first section contains an optional set of bounding box controls that will allow you to refine your query spatially. All queries can be filtered on a graphic bounding box on the map that you can control.

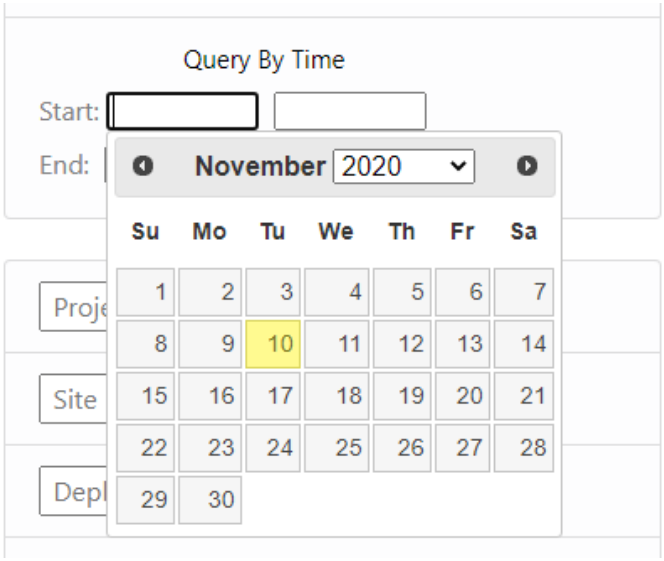


To activate the controls, click the 'Bounding Box' button at the top. You will notice a transparent red-outlined square appear on the map. You can grab the box on the map and drag the entire box. You can also grab the corners or the sides to resize and shape the bounding box area. You will notice when you drag and reshape the box that the Longitude and Latitude coordinates will automatically update. By selecting the 'Lock Bounding Box' checkbox the Bounding Box will lock in place while you scroll or zoom the map. You do not have to provide any Bounding Box information for a query, it is optional. To turn off the bounding box coordinates and remove them from your query, you can simply click the 'Bounding Box' button and the coordinates will disappear.

2.2.1 Time Controls

In addition to spatial filters, queries can be narrowed to include a time range for the data. The first row of the 'Query By Time' section is the start time. By clicking on the box, a calendar date can be chosen for the Start date in the left-hand column (shown below). A time component for

the Start date can be chosen for the hour, minute, and second in the right-hand column. The bottom row contains the same controls but for the end date/time.



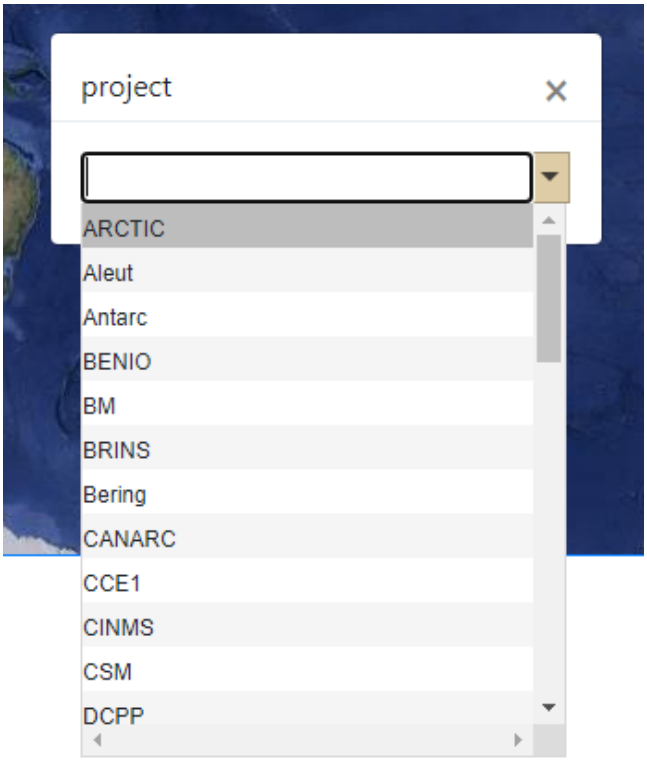
2.3 Species, Site, Project, Deployment Controls

A set of four controls appear directly below the Bounding Box and Time control.

The first control allows you to enter a Project name. The second and third controls allow you to choose a Site name and Deployment Id. The last control allows you to enter a species Id. The default species abbreviation expected is 'SIO.SWAL.v1'. The species abbreviation format can be adjusted using the Species Library drop-down just to the right of the Advanced Queries tab. For example, if you would like to just Latin names for your species Id, select 'LATIN NAMES' from the Species Library drop-down.

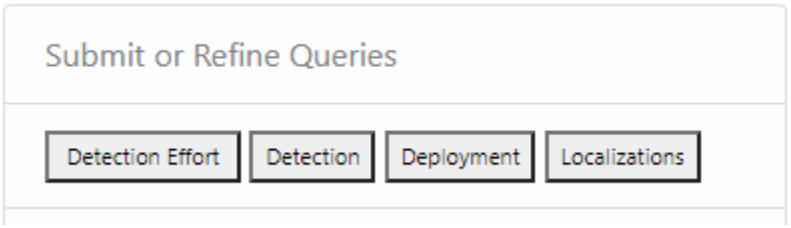
Next to each of these four controls is a small magnifying glass button. Clicking on this button will present you with a list of possible values for that control. The possible values will change depending on the values that have been filled in for the other controls. For example, when all

controls are empty, the list of sites will show all sites that exist in your database; however, if you have entered a Project, the list of sites will only include sites that are a part of the given Project. An example of possible values for Project is shown below.



2.4 Submit or Refine Queries

The next group of controls entitled 'Submit or Refine Queries' allows you to choose the type of simple query you wish to perform and optionally refine it in the advanced viewer. Other options include export/save results, save queries and load queries. To begin we will split up this into two sections easier to understand. Below is the first section of those controls:



Four buttons submit specific pre-built queries, based on the parameters entered in section 2.3 and the spatial and/or temporal filters. The queries are Detection Effort, Detection, Deployment, and Localizations.

The Detection Effort query returns information about deployments for which there was detection effort based on the specified parameters. For example, maybe you want to know at which sites within a specific project there was effort for Risso's Dolphins (Grampus griseus). Alternatively, maybe you want to know all of the species there was effort for at a specific site or across an entire project.

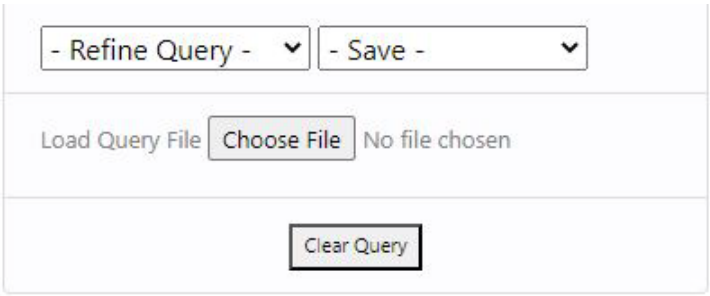
The Detection query returns detection information and deployment information for actual detections given a set of parameters. For example, you may want to know all of the blue whale detections for a project, site, and deployment combination, or you may simply want all the detections at a specific site.

The Deployment query returns deployment information for the specified parameters.

The Localization query returns information about tracks from a localization study. Information will include the timestamps and lat/longs for specific waypoints along a track as well as the bounding box for the entire track.

2.5 Refine queries in advanced viewer, Save Results, Loading Queries

The next section of controls (see below) starts with a drop-down box for refining the pre-built queries in the Advanced Queries tab. By selecting one of the query options, the view will switch to the Advanced Queries tab, already populated with your parameters. From here you can refine the query by adding or subtracting new parameters. This will be described in detail in section 3. where we discuss the Advanced Queries tab.



For now, let us focus on the Save menu. The 'Save' drop-down has four options that can be used after you have gotten the results of the query:

- 'Save Query' saves the parameters of the query you performed, so that you can re-run your query in the future. This is particularly useful with the more complex queries built in section 3.
- Save Results as JSON – Saves the query results as a JSON file using the Badgerfish convention when converting from the XML result to JSON.
- Save Results as MATLAB – Saves the query result as a MATLAB file that can be loaded with Matlab's load command.

- Save Results as R – Converts the query result to data.tree format if R has been configured on the server (see the Tethys manual for details). A more efficient method of using XML in R is to use the xml2 package. See the R client documentation for more details.
- Save Results as XML – Downloads an XML file of the results.

CAVEAT: If you are using Tethys without encryption (web server is addressed with http as opposed to https), your web browser may require you to approve the download. Some recent browsers block such downloads unless the user approves the download.

The final option in the Submit and Refine Queries box is the 'Clear Query' button. Clicking this will remove any query results and remove any values entered for any of the query parameters.

2.6 Viewing Results in an R session.

Documents can be saved to XML and read into R using packages for reading and parsing XML files (xml2, XML). If your Tethys administrator enabled R on your server, you should also be able to store results in the data.tree package format which can be loaded into R. The data.tree format is less efficient than using the xml2 library. See the R client documentation for guidelines on working with XML.

2.7 Viewing the Results in MATLAB

To view the results in Matlab, use Matlab's load function to load the .mat file. For example, if we stored the results of an effort query to effort.mat in Matlab's working folder, we could use:

```
>> eff = load('effort.mat')
```

This creates a structure array that mirrors the XML. Each deployment/detection record is an entry in a cell array (use { } for cell arrays) of eff.Result.Record:

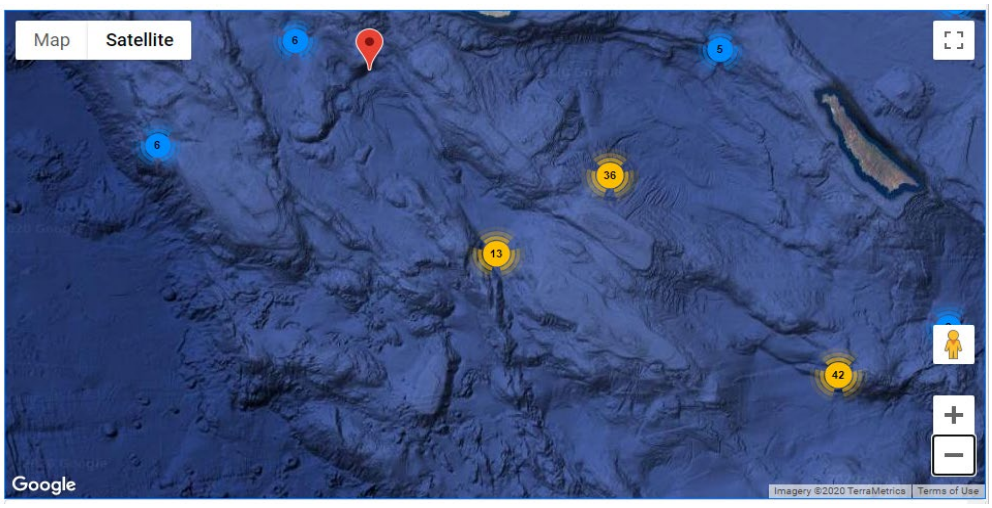
```
>> eff.Result.Record{8}  % 8th effort record

ans =

  struct with fields:

    Deployment: [1×1 struct]
    Detections: [1×1 struct]

>> eff.Result.Record{8}.Detections

ans =

  struct with fields:
```

```
        Start: 7.3694e+05   % Matlab serial date (datenum)
          End: 7.3694e+05   % use datestr to interpret
         Kind: [1×1 struct]
    Algorithm: [1×1 struct]
```

2.8 Viewing the Results of Simple Queries.

To the right of the inputs for building the query there is a map and, below it, an empty pane. These are the areas where the results of your query will appear. Once you have set your parameters and optionally set filters, clicking the query button will submit the query and a progress indicator will appear to let you know when the query is finished. When the query is finished, the results will appear on the map as colored circles and a series of results folders will appear below the map. For all queries, results over 200 MB will not be returned below the map and can instead be downloaded through the 'Save' drop-down.

In the mapped results, individual data points are represented as red balloons. Hovering over the red markers will reveal information about that data point. Yellow or blue circular markers represent clusters of data points with a number that denotes how many data points are clustered there. Clicking on the circular markers will expand to show the individual data points in the cluster.



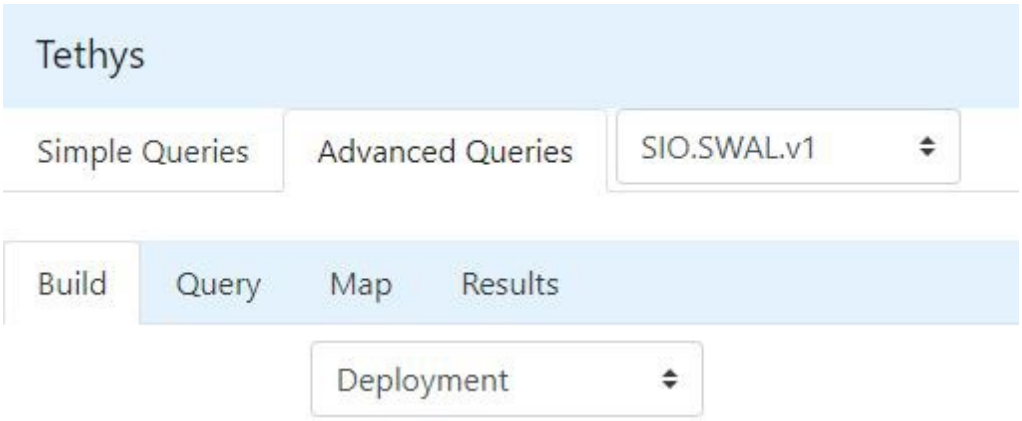
3 Advanced Queries

There are two ways to activate an advanced query. You can go there directly by using the 'Advanced Queries' tab at the top of the page or you can refine a simple query by using the 'Refine Query' drop-down. Opening an advanced query from its tab will open a new blank query while using the refine option will populate the advanced query with the parameters of a simple query.

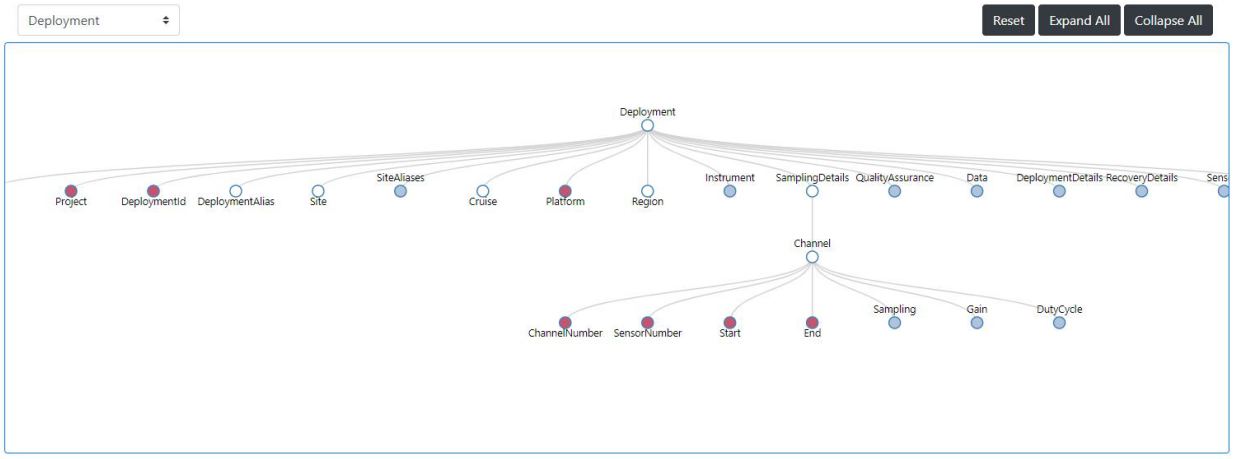
3.1 Controls for the Advanced Queries View

3.1.1 The Build Tab

Shown below is the top left of the page that opens when the Advanced Queries tab is selected either by the user directly or through the refinement tool. Below the main tabs is a second tier of tabs: Build, Query, Map, and Results.



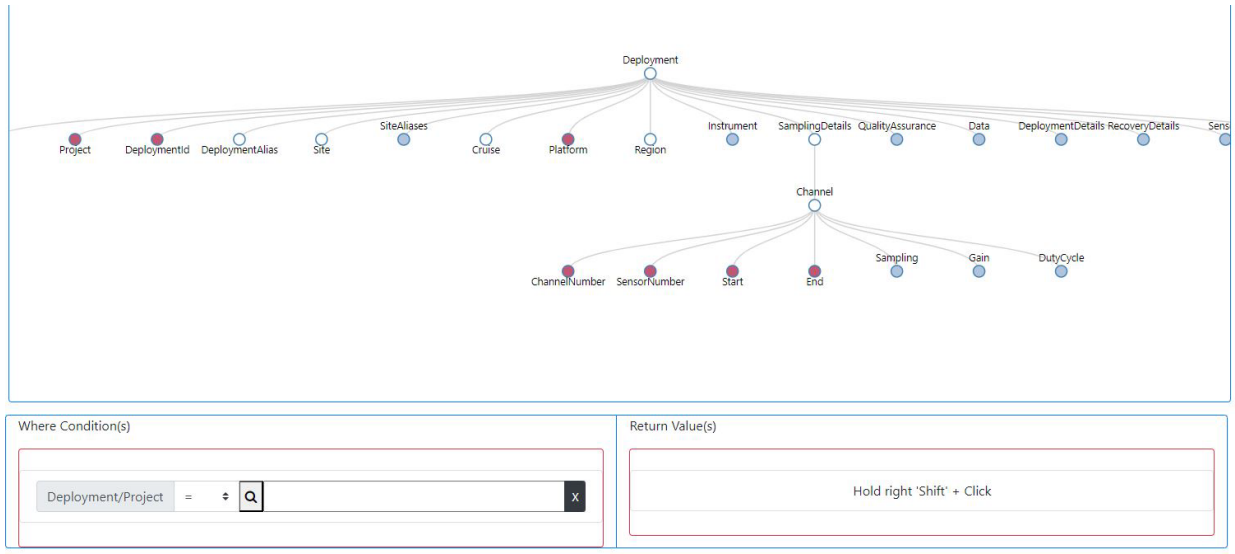
The Build tab is the default view for the Advanced Queries tab. This is where you can actually build a query using a user-friendly drag-and-drop functionality. The drop-down at the top of the Build tab allows you to choose the collection from which to choose parameters (Deployment is the default collection). For example, you may want to use details of a deployment from the Deployment collection when formulating your query, but you may also want to select a species Id from the OnEffort/Detection section of the Detections collection. Selecting a collection from the drop-down will populate a tree that depicts all the elements in the collection. Below is the tree view of the Deployment collection with the SamplingDetails node expanded.



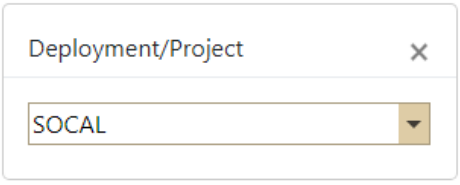
Navigating the tree is as simple as just dragging it and using your mouse wheel to zoom in or out. You'll notice that nodes on the tree have different colors. The blue nodes in the tree are expandable. Red nodes indicate elements that every record within the given collection must

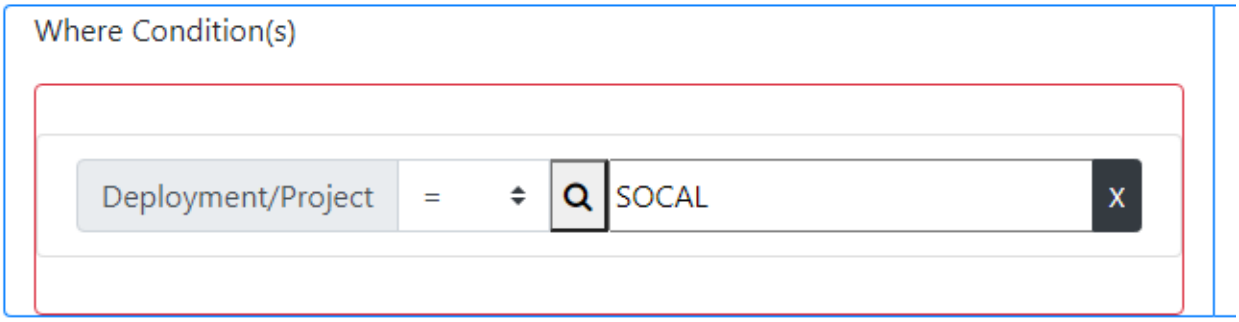
contain while white nodes are optional elements. By hovering the cursor over an element, a description of the data represented by that element will appear to the bottom left of the tree.

Below the tree view there is a single row containing two columns. The left column represents Where Conditions for your query (the parameters and filters that define what you are looking for) and the right column represents the Return Values for your query (the output you would like your query to produce). To populate these columns, hold the left Shift key down and click on a node that you would like to include in the Where clause. Let us say that you want to query on the Project name. Find the Project node (located on the far left of the tree), hold the left Shift key, and click on the Project node. A Where condition will appear in the row below the tree as shown below. Note that you can also double-click on a node and drag and drop it into the Where Conditions or Return Values columns as needed.

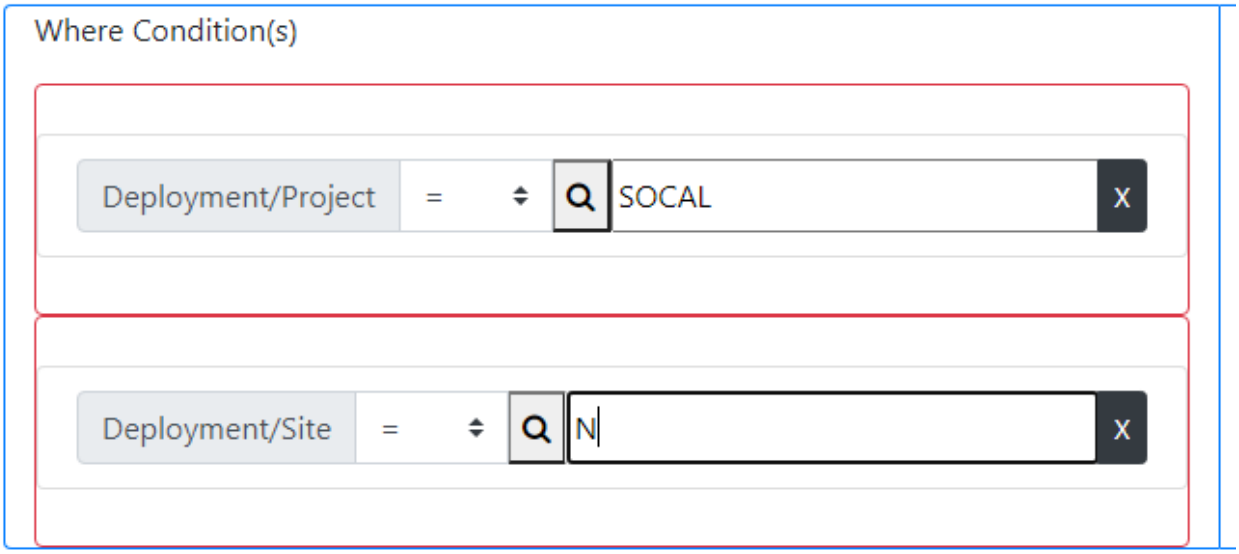


Notice that the node name will appear on the left followed by an operator drop-down. The default operator is ‘equals’ but the drop-down can be used to select a different operator. The next button is a magnifying glass; by clicking on it you will get a drop-down with all of the allowable values for this variable. Below are images of the values drop-down, and the Where Condition populated with the value chosen using the magnifying glass drop-down.

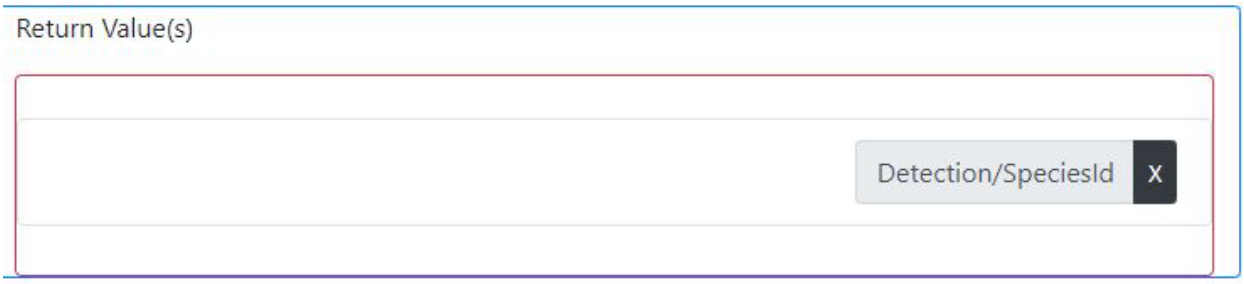




By clicking on other nodes, additional conditions can be added as new rows (shown below).



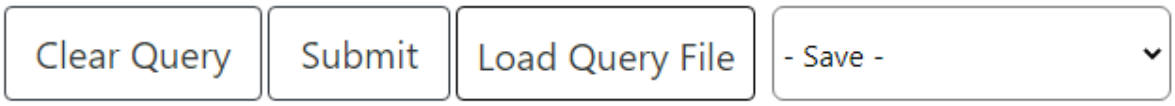
Adding Return Values follows the same procedure but when adding Return Values hold the right 'Shift' key while clicking on a node. Let us say that we would like to return some information from the Detections collection; below is what that would look like.



Notice that we have decided to return the SpeciesId from the OnEffort/Detection node for the conditions defined in the Where Conditions column. The query will return all species detected at

site N in the SOCAL project. Again, adding more return values by clicking on nodes will add new rows to the 'Return Value(s)' column.

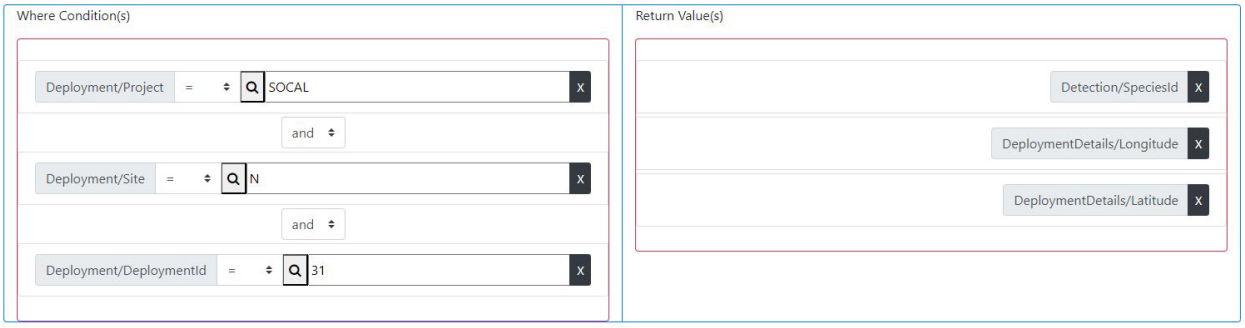
At the bottom of the page, you will see the following buttons.



Once you have finished building your query, you can click on 'Submit' to retrieve your results. Notice that you can clear a query, submit a query, or build on an older query you have saved, and load those parameters into a new query by clicking on 'Load Query File'. You can also use the 'Save' drop-down to save your results in the same formats as the 'Save' drop-down described for the Simple Queries.

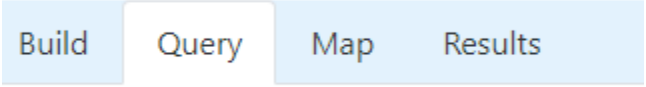
3.1.2 Viewing Your Results

Let's say that we have the following query:



It is essentially the same query as the query we were constructing previously, but with an extra Where condition for the DeploymentId and two extra Return values (the latitude and longitude for the deployments). Upon clicking the 'Submit' button, a progress indicator will appear while the query is being processed. When the query finishes, the results are sent back to the web client for viewing and downloading/exporting.

Next to the Build tab, there are three more tabs named, Query, Map, and Results.

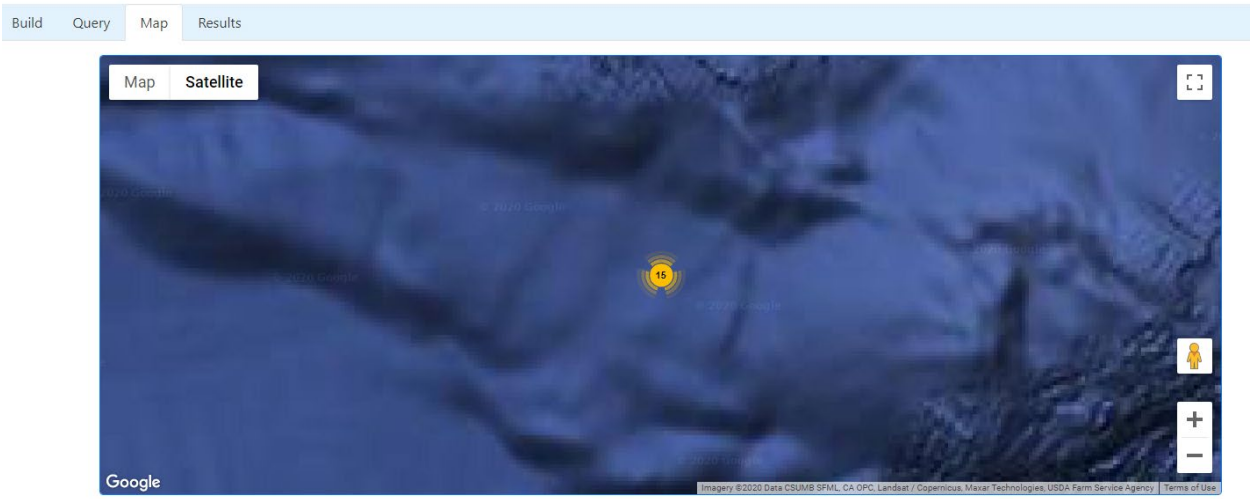


The Query tab shows the structure of the query in both JSON and XQuery format. The database uses the XQuery language and viewing the query can sometimes give you a sanity check when

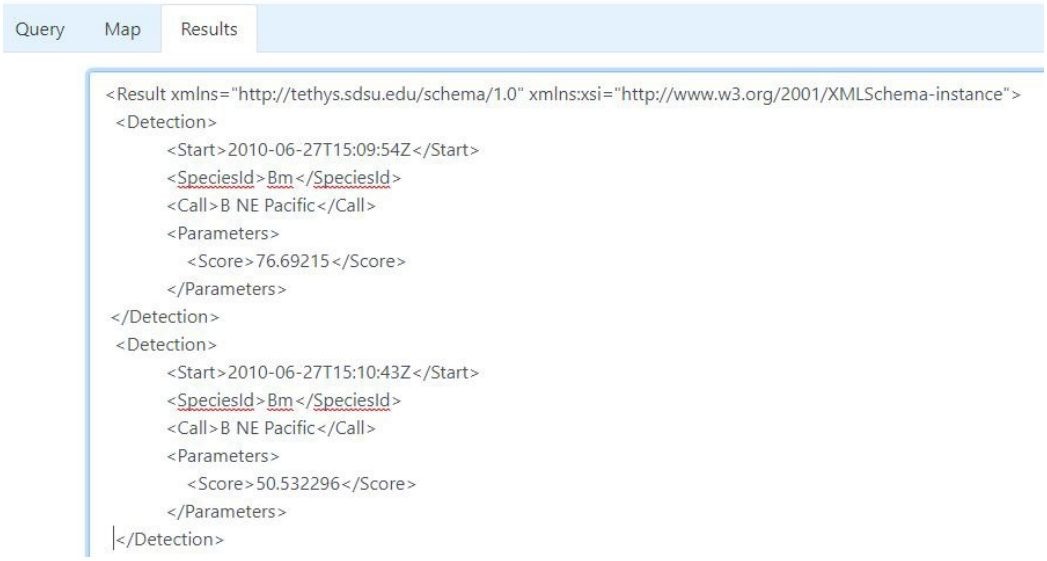
you receive results that you did not expect, or as a last-minute check before submitting your query. The following is the XQuery for the query parameters shown above:

```
import module namespace lib="http://tethys.sdsu.edu/XQueryFns" at "Tethys.xq";
declare default element namespace "http://tethys.sdsu.edu/schema/1.0";<Result
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"> {
  for $Deployment0 in collection("Deployments")/Deployment[Project = "SOCAL" and
Site = "N" and DeploymentId = 31],
  $Detections0 in collection("Detections")/Detections[DataSource/DeploymentId =
$Deployment0/Id]
  return
    <Return>{
      <Deployment>{
        $Deployment0/DeploymentDetails/Longitude,
        $Deployment0/DeploymentDetails/Latitude
      }</Deployment>,
      <Detections>{
        $Detections0/OnEffort/Detection/SpeciesId
      }</Detections>
    }</Return>
} </Result>
```

Next is the Map tab. If you include latitudes and longitudes in your Return Values, you will be able to view the coordinates of your results in the map as shown below.



The last tab is the Results tab. Any results from your query will be shown here in XML format (assuming your result set is smaller than 200 MB; if it is larger, you can easily download the results as XML from the ‘Save’ drop-down).



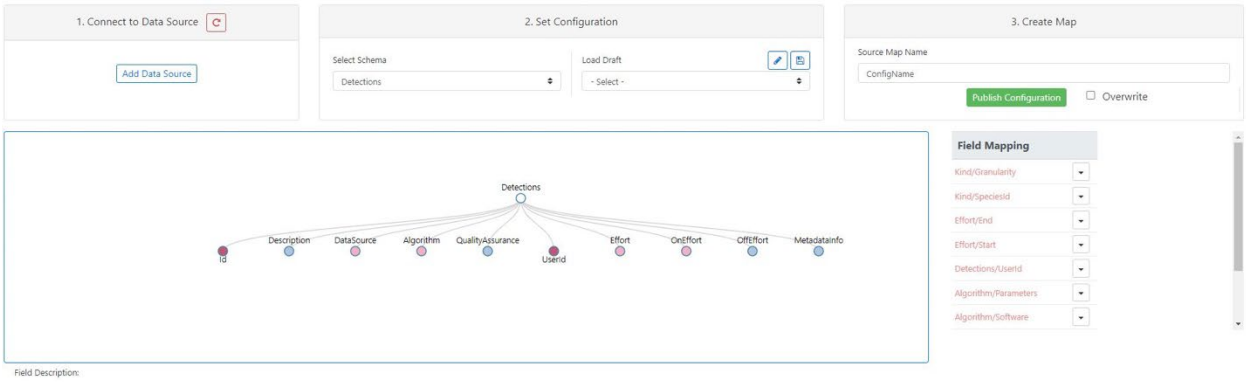
4 Create a Source Map

Source maps are XML documents that provide information on how field names from a document that you want to add to your database should be translated to the established Tethys schemata. For more detail on source maps and the schemata, see the Tethys Manual (source maps: sections 3.2.1 and 6.1; schemata: section 5).

Source maps are typically used to import data from multiple files. That is, we create a source map for a specific type of data such as detections from a spreadsheet, and then use the source map each time that we want to import data from a spreadsheet that has been populated in the same way. The source map tool does not import data, it simply builds a mapping from your data to the Tethys schemata that can then be used to import data.

The Tethys Web Client has an interface for creating source maps that can then be used when importing data into Tethys. To access this tool, click **Plan data import (sourcemap generator)** from the drop-down at the top right of the Web Client.

When the tool first opens you will see three numbered boxes across the top (Connect to Data Source, Set Configuration, and Create Map). Below these boxes will be the schema for the Detections collection and a Field Mapping box with elements from the Detections schema.



To create a source map, you will need to add the document or database you need to map to the Tethys schema, select the schema your source map will be for, and then match the fields from your source data to the fields in the Tethys schema you selected.

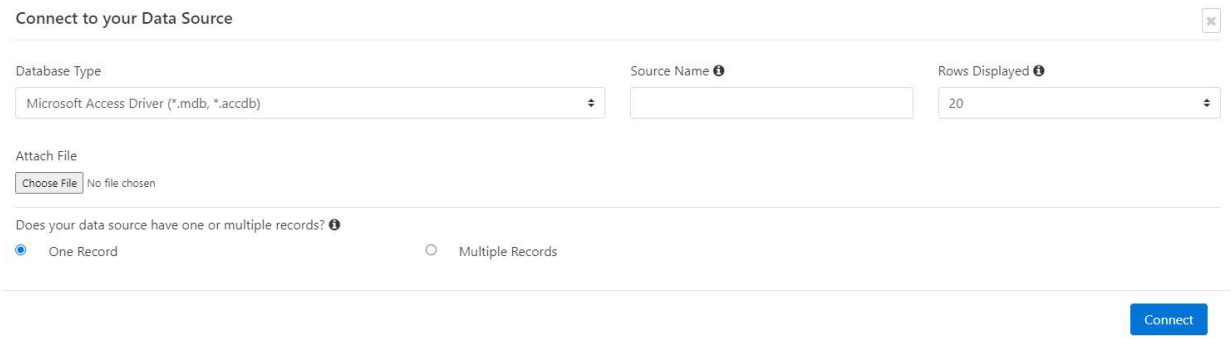
In most cases, you will be selecting Deployments, Detections, or Localizations, depending on what you are interested in mapping, but the tool can be used for any type of schema supported by Tethys.

4.1 Connect to a data source

In the Connect to Data Source box, click the **Add Data Source** button to see your data source options. You will be required to enter information about the data source you would like to use for your source map.

The **Database Type** drop-down will be automatically populated with all database types that the Tethys server currently knows about (e.g., Microsoft Access or Excel, MySQL, PostgreSQL, etc.). Select the one you would like to use. If you have a data type that the server is not aware of, you may be able to ask your Tethys administrator to install an office database connectivity (ODBC) driver that will let you use that type of resource. ODBC drivers exist for most databases and office productivity software. The remaining input options will change based on the database type that is selected. The information icons next to each field name will provide an explanation of what is needed, but we will go through two examples below.

If Microsoft Access or Excel is selected, the remaining inputs are:

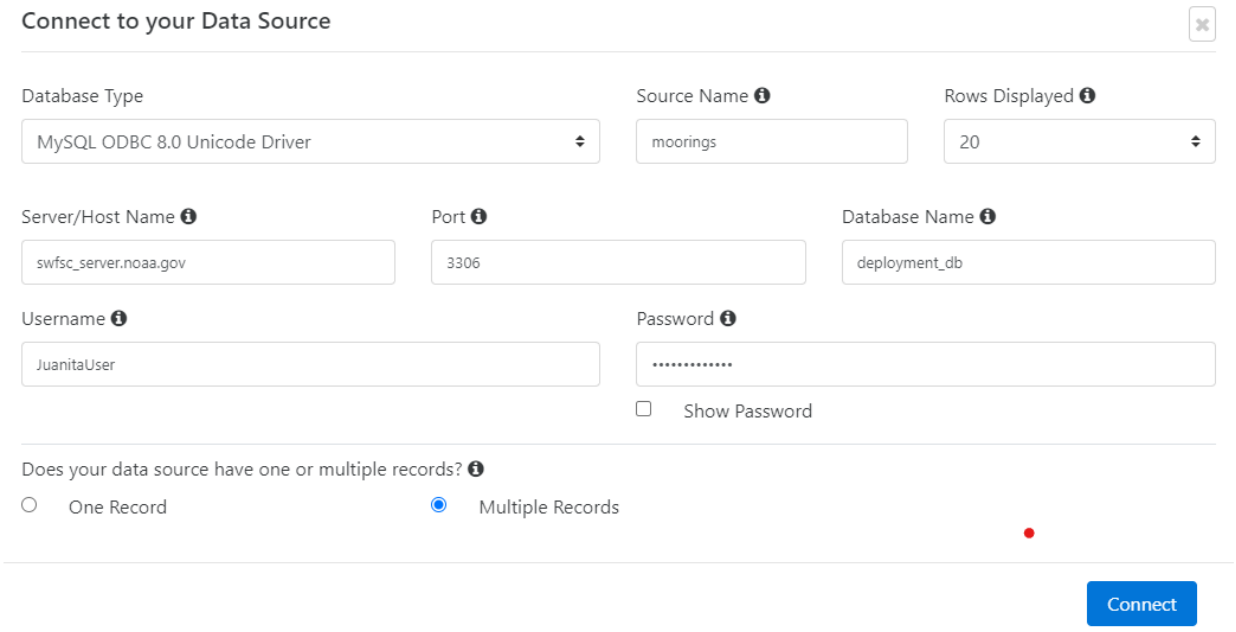


The **Source Name** is whatever you would like to call your data source. This field is important when you want to combine multiple sources, such as multiple spreadsheets or comma-separated value files. Each data source will be referred to by a name of your choice. In general, this field should say something about the type of data, not the name of a specific file.

The **Rows Displayed** drop-down allows you to select how many rows of your data source you would like to be able to view. The default is 20 as you often only need to see a few rows to understand the data contained in a data source. Additionally, it can take a long time to load all the data rows for large databases.

Finally, determine whether your data source will be creating a single record (e.g., one deployment) or multiple records. Multiple records are more common when selecting from databases. For example, if we had an instrumentation database and we wished to populate all instrument deployments, we would select multiple records.

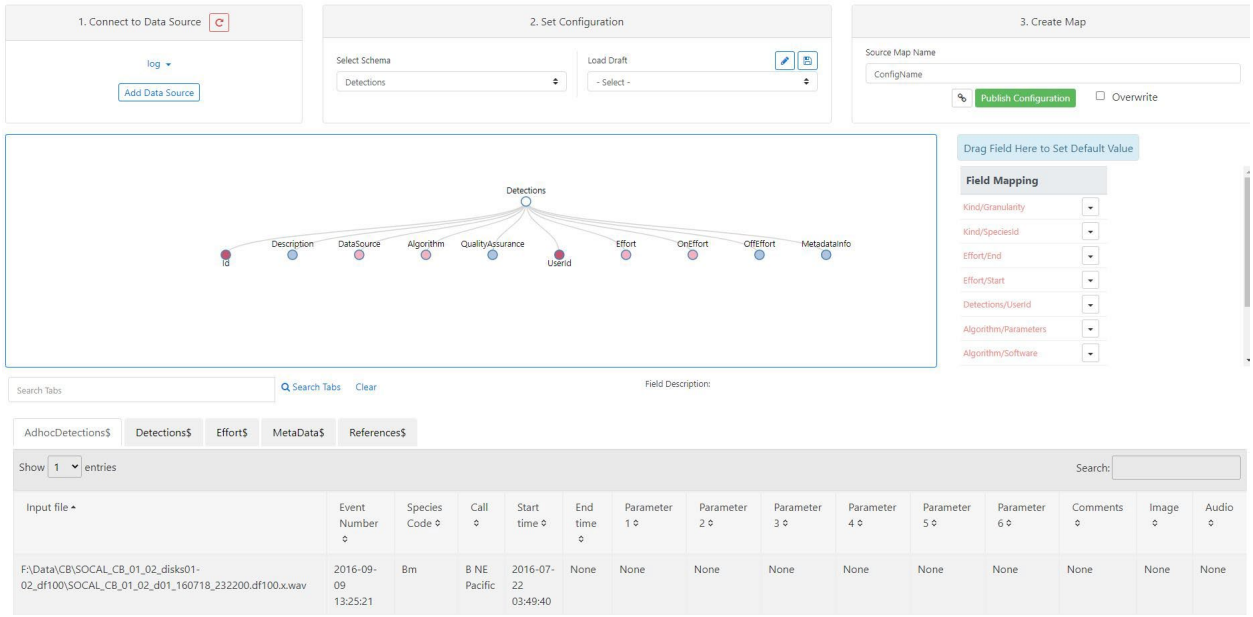
When the Data Source is a network resource such as a database server such as MySQL or a Microsoft SQL server, there are some additional inputs required (MySQL server shown):



The **Server/Host Name** is the name of the computer that hosts your data source. This is typically a computer within your domain, e.g. swfsc_server.noaa.gov. **Database Name** is the name of the database. The **Username** and **Password** are the credentials you use to access the SQL database. If you do not require a password, you can leave this field blank. If you do not know what values to place in these fields, contact the person at your organization who administers the database to which you wish to connect.

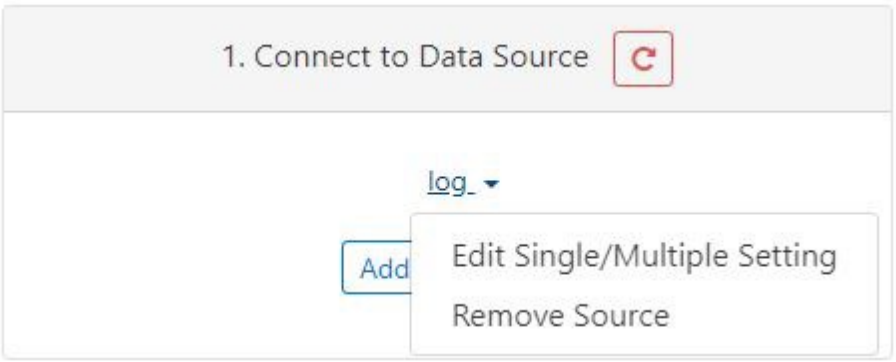
Once you have filled out all required inputs, click the blue **Connect** button. The Connect to Data Source box should now show the data source you just added, and assuming that there are no

errors, one or more tables showing representative data from the specified source should appear at the bottom of the page.



In this example, an Excel spreadsheet containing detections was added. Each tab at the top of the table corresponds to a sheet from the imported spreadsheet. The sheet names have dollar signs appended to them (this is an artifact of how Microsoft handles the spreadsheets). You can add additional data sources as needed, and information from all data sources will appear in this table. Note that if you are adding multiple data sources, the source names must be different. By hovering the cursor over a tab, a tooltip will indicate which data source a tab is from.

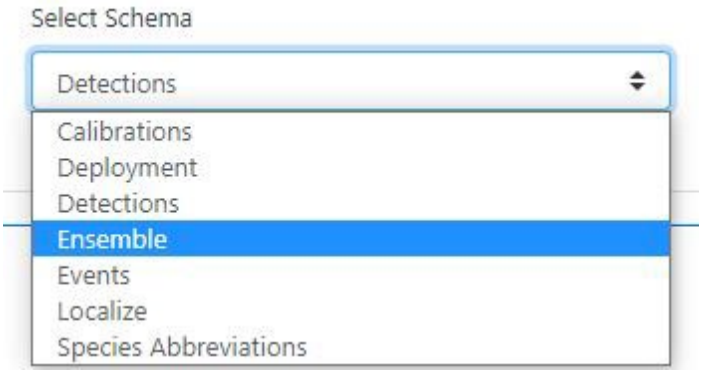
You can use the drop-down next to the data source name to modify whether it represents one or multiple records or to remove it entirely.



4.2 Set the configuration

When a document is imported into Tethys, it is added to a collection for which there is a defined schema (rules about specific fields that need to be present and how different fields are related to one another). In the Set Configuration box, you should use the **Select Schema** drop-down to

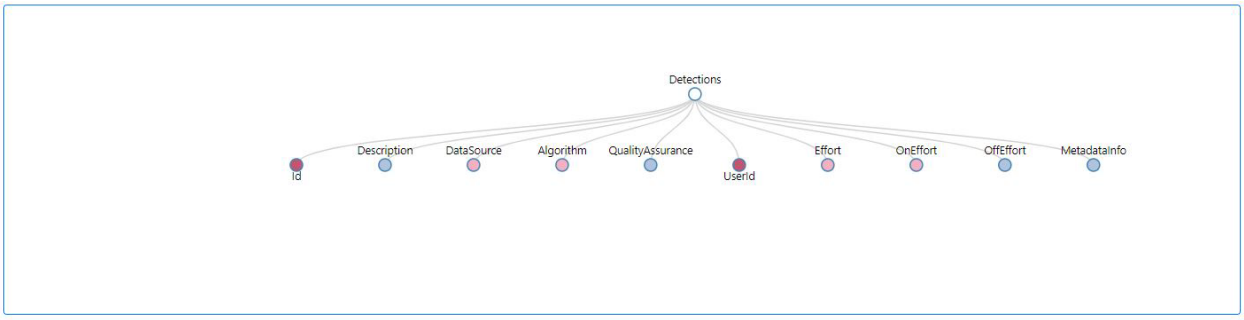
select the collection that your source map will be for. For example, if you have a database of deployment information that you want to be able to translate into the Tethys schema, you would choose the Deployment schema. The schema map and Field Mapping box will change based on what is selected here.



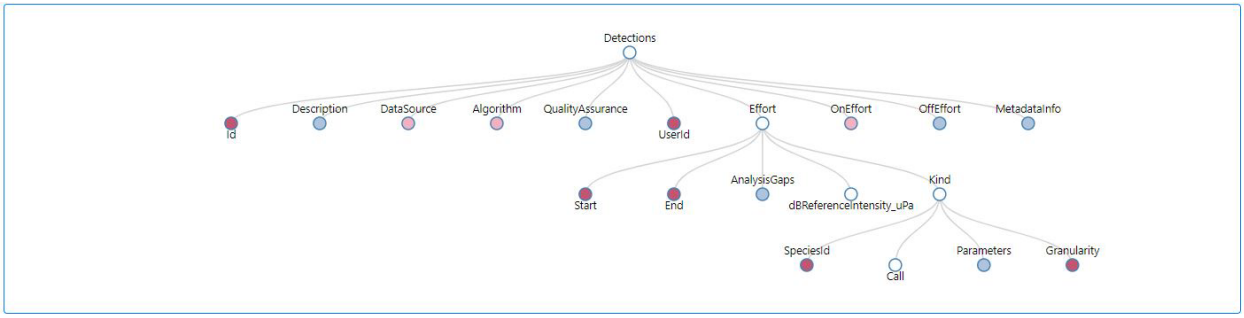
If you want to continue working on a source map that you saved as a draft previously, you can use the **Load Draft** drop-down to select the file. This will automatically update the schema selection to whichever was used previously.

4.3 Navigate the schema

You can navigate the schema visualization by clicking and dragging to move around and by scrolling on your mouse to zoom in and out. Below is an example of the Detections schema.



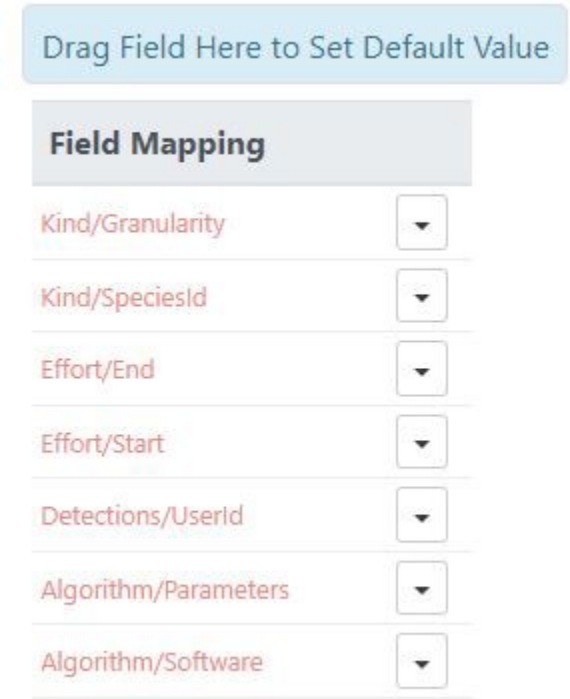
Required fields are indicated by red circles, while optional fields are indicated by white circles. Fields that contain subfields are indicated by pink and blue circles. A pink circle indicates that a subfield is required, while white circles have no required subfields. Click once on a circle to expand that field and see any subfields. The example below is of the same Detections schema, but the Effort field has been expanded, as well as the Kind field within the Effort field. To collapse an expanded field, simply click the circle again.



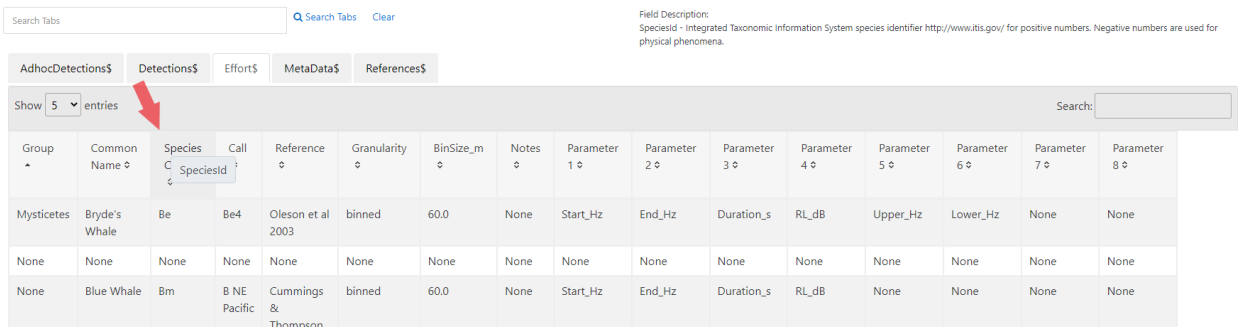
At the bottom-right of the schema visualization is a Field Description. As you hover over different schema fields, the field description will provide a brief description of what the field is meant to contain in the Tethys schema so that you know what you should map it to in your data source.

4.4 Field mapping

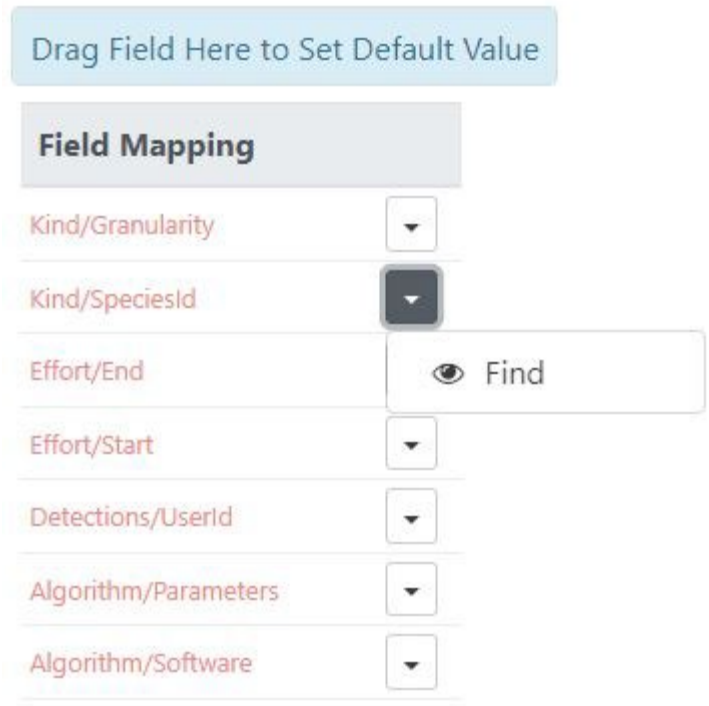
Depending on the selected schema, the Field Mapping box will be populated with the schema fields that are required for your source map. Below are some of the required fields for the Detections schema.



To make your source map, you must match the fields from your data source table to the required fields in the Field Mapping box. To do this, simply double-click on a field in the schema visualization and drag and drop it on the field you want to match it with in your data source table. In the example below the SpeciesId field from the Detections schema is being mapped to a Species Code field in the data source table.

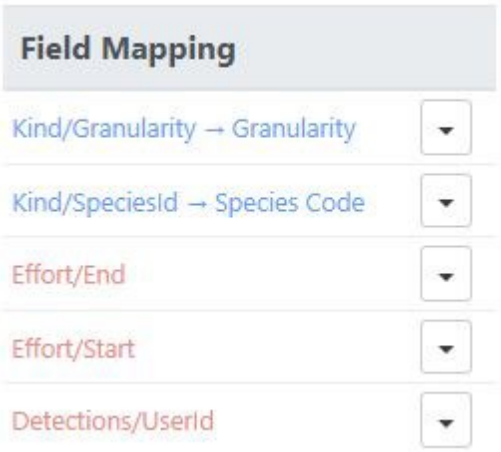


If you are having trouble locating a required field in the schema visualization, you can select the drop-down next to the field in the Field Mapping box and click **Find**. This will highlight the field in yellow and the tree will move to center on the field. If hidden, the tree will expand to reveal the field.

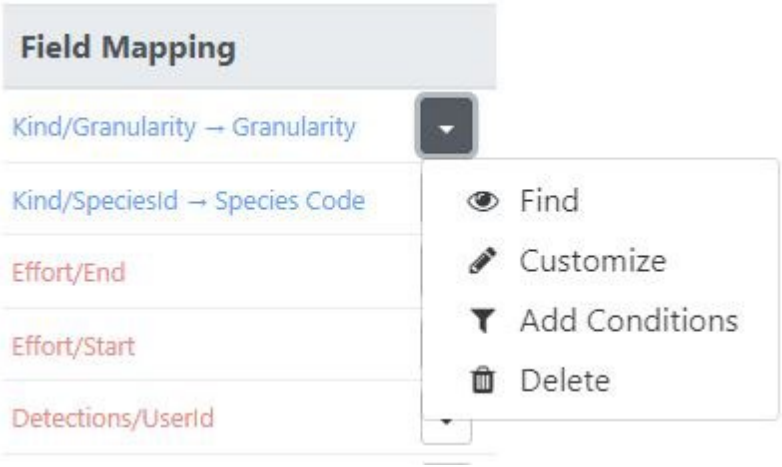


Likewise, if your data source table has many tabs, you can use the search bar located above the table to search for specific tabs in the table.

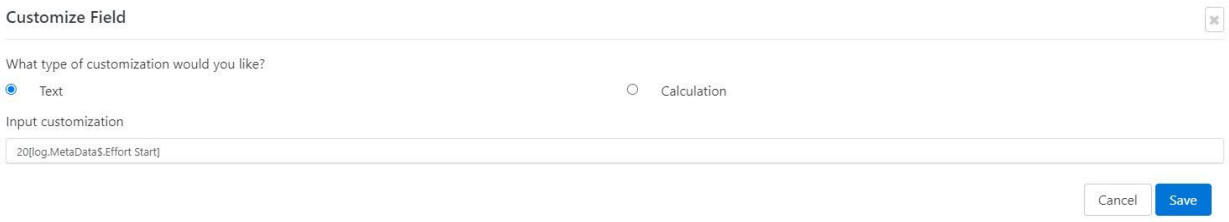
Once you drop a field from the schema visualization onto your data source, the Field Mapping box will indicate that the field has been mapped. The field will turn blue and will have an arrow pointing to the data source field it was mapped to. The field in the schema visualization will also turn green to indicate that it has been mapped.



Once you have mapped a field, there are various way you can modify it. You can see these options by selecting the drop-down next to the field name. When unmapped, this drop-down has one option: Find. Once the field has been mapped there will be additional options to Customize, Add Conditions, and Delete.



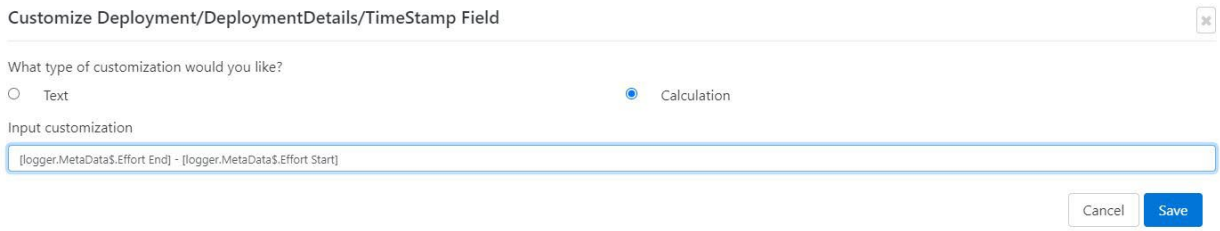
If you select **Customize** you will be able to modify the field using text or a calculation and the input box provided. A text customization can be used when data need customization. For example, if dates were stored with two-digit years, you could create a customization to add the first two digits of the year:



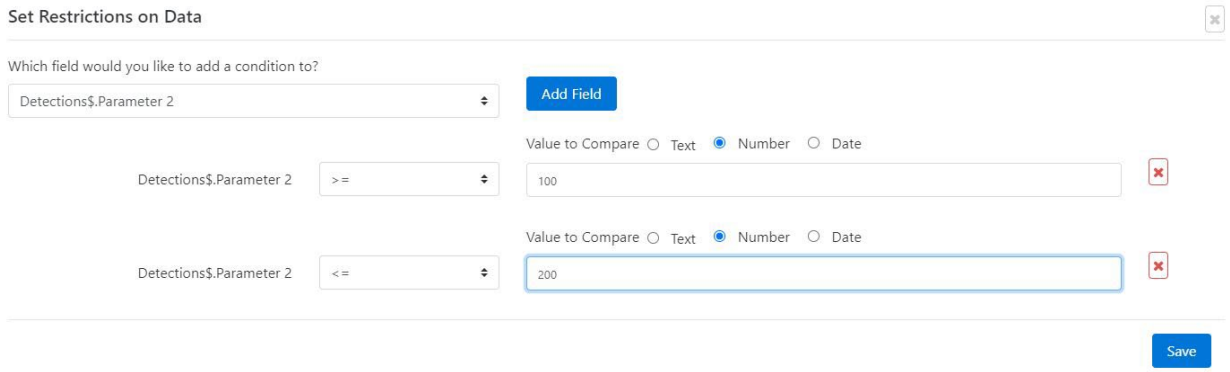
Customization can also be used to combine fields. If the MetaData tab contained separate date and time fields (StartDate and StartTime), you could drag a Tethys timestamp onto both fields and customize the content as follows:

[logger.MetaData$.StartDate][logger.MetaData$.StartTime]

A calculation customization can be used if you need to perform a mathematical transformation of the data. Any valid SQL expression can be used. This includes simple arithmetic, such as converting a field from minutes to seconds or computing the number of days between a start and end date using SQL's datediff function. Perhaps you want to convert hours to minutes or, as in the example below, you want to determine a duration.



If you select **Add Conditions** you will be prompted to create a condition about which data will be imported into Tethys. For example, let's say you only want to pull deployments from a specific range of latitudes. You would add two conditions to define this range. When you add your condition, you also need to select whether it is text, a number, or a date. In the example below a condition has been set for a specific detection parameter so that it must be between 100 and 200 to be included.



You can also choose **Delete** to remove the match you made.

If you do not have a field in your data source that you want to match to the schema, you have the option of setting a default value. You can drag and drop the field onto the **Drag Field Here to Set Default Value** label above the Field Mapping box. You will be prompted to enter your default value and click **Save**.

This is also useful if there will be null values in your data. By setting a default you ensure that any rows without data for the field will be filled in with the default value.

Although only required fields are listed in the Field Mapping box, you can include any fields from the schema in your source map. If you add any optional fields, the Field Mapping list will update to include the field and any fields that are required as a result of selecting that field. For example, when creating a source map for the Detection schema, the OffEffort field is not required, but if you want to include a field from OffEfffort, you are required to include Start and SpeciesId.

If you have already mapped a field and you map it to another field in your data source, you will be prompted to override the previous mapping or add the additional field. In some cases, you may want a schema field mapped to multiple data source fields. For example, perhaps you want to combine a date field and a time field from your source map into one column. To do this you would need to drag the schema field that will represent the combined date and time to both your data source's date field and its time field. When you add the time field and are prompted with the override or add field option you would choose **Add Field**. You will then be prompted to customize the field, as described previously.

4.5 Create source map

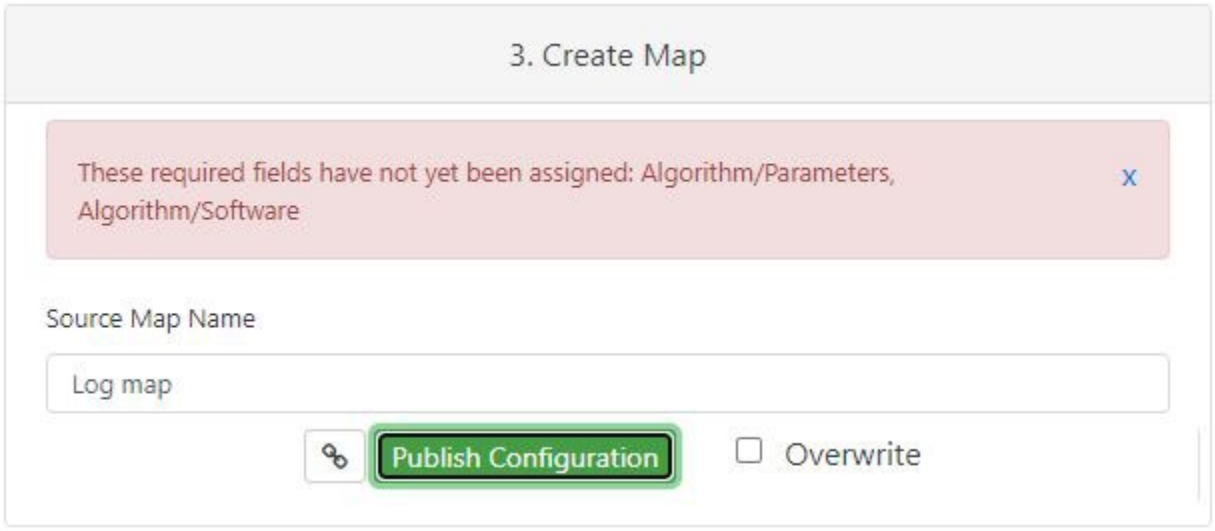
Once all the required fields in the Tethys schema have been mapped to fields from your data source (you can verify this by clicking the chain icon next to the green **Publish Configuration** button), you can create your source map by selecting the green **Publish Configuration** button in the Create Map box.

This will publish your source map to the Tethys server. Like any document in Tethys, if you wish to see the XML that was produced, type your server and port name into a web browser followed by a slash and the collection name, which in this case is SourceMaps. A list of source maps by name will be shown and clicking on the one you published will show the XML-based source map that was created for you, which will have the following structure:

```
<Mapping>
	<Name>MyMap</Name>
	<DocumentAttributes>
		List of <Attribute> elements that will be applied to the document
	</DocumentAttributes>
	<Directives>
		Elements that provide document structure and specify how
		fields in the source spreadsheet, database, etc. correspond to elements in the
		resulting XML document.
	</Directives>
</Mapping>
```

You can upload this source map to Tethys and then when you import documents into Tethys in the future, you will be able to use this source map.

If you did not assign all of the fields listed in the Field Mapping box to fields in your data source, you will receive an error message like the one below that will list which fields need to be assigned before your source map can be created.



If you have not finished working on your source map but would like to save your progress, you can save a draft by selecting the disk icon within the Set Configuration box. This will save your in-progress configuration to the *ImportConfig* folder within the folder where your database is stored. **Note that this means that anyone who has access to the database server would have access to this file.** When you want to continue working on your source map, you can return to the source map tool and click the **Load Configuration** drop-down in the Set Configuration box. This will generate a list of all configuration files for your database, and you can select the one you want to modify. Once selected, the tool will populate the correct schema and field mappings. To remove a draft, select the pencil icon within the Set Configuration box. This will populate a list of all saved drafts that you can then remove.