


Mapping your dataset to WQX Web

Water quality data collecting organizations store their collected water quality data in a format that is optimal for their data entry, QA, and analysis purposes. More often than not, that data format does not look like the WQX web template format. Frequently, organizations which receive 106 grant funding believe they must change their dataset to fit into the WQX web template to be entered into WQX. In these cases, organizations are copying and pasting information into the WQX web template and filling in columns they do not use just to be in compliance with their 106 grant conditions. Within WQX Web, there is an option to create import configurations that match your dataset, rather than making your dataset match WQX formatting. This document was developed to assist with the mapping of your desired dataset format to an import configuration and submitting data to STORET.





















Create Generated Values: Generated values are metadata values needed for successful WQX submissions that are not in your original dataset. Such values may include: Organization ID, Project ID, Activity Type, Activity Media Name, Result Status ID, Result Value Type, Sample Collection Equipment Name, Sample Collection Method ID, & Activity Start Time Zone. You will use generated values if you want every record (i.e. row) in your dataset to contain the information. Ex: in the below screenshot, every result in your dataset will have been collected using a probe/sensor. If that is not the case for your dataset, you should insert the equipment type into your raw dataset and it should NOT be a generated value. Another example is Activity Media Name. The collecting organization knows the data is water data and do not indicate it in their raw dataset. But because WQX allows more than just water data, it needs to be specified before the data can go into the system. A simple way to do that is using generated values.

Generated Values (not in your import file):

| Pick | Entity | Entity # | Element | Default Value |
|---|------------------|----------|----------------------------------|-------------------------------|
|  | Organization | All | Organization ID | WQXTEST |
| | Activity | All | Activity Media Name | Water |
| | Activity | All | Activity Type | Field Msr/Obs |
| | Activity Project | All | Project ID | (none) |
| | Result | All | Result Status ID | Final |
| | Result | All | Result Value Type | Actual |
| | Activity | All | Sample Collection Equipment Name | Probe/Sensor |
| | Activity | All | Sample Collection Method ID | (none) |
| | Activity | All | Activity Start Time Zone | (none) |

Suggested minimum data elements in your original file: All raw datasets, no matter the format, have to contain minimum elements to be inserted into WQX. Those are: Monitoring Location ID, Activity ID, the Characteristic you are measuring, the result value, the Start Date, and Start Time. Information such as units for your result are required but can be defaulted in the import configuration as shown with the temperature result below. NOTE: The temperature example is shown using “Expert Mode”. Do not be intimidated by this language. It is simply a way of saying “My data is not in the WQX template”.

Regular Columns (starting at row 2 of your import file)

| | Column | Pick | Entity | Entity # | Element | Default Value | Format | Translations |
|---|--------|---|----------|----------|------------------------|-----------------|------------|-------------------------|
|    | A |  | Activity | 1 | Activity Start Date | | YYYY/MM/DD | 0 Edit |
|    | B |  | Result | 1.1 | Characteristic Name | Temperature_air | | 12 Edit |
| | | | Result | 1.1 | Result Unit | deg_F | | |
| | | | Result | 1.1 | Result Value | | | |
|    | C |  | Activity | 1 | Activity ID | | | 0 Edit |
|    | D |  | Activity | 1 | Monitoring Location ID | (none) | | 0 Edit |
|    | E |  | Activity | 1 | Activity Start Time | | | 0 Edit |

In the above screenshot, you can see in Column B, there is Temperature, the result unit, and the result value. This is a beneficial way to capture your data when it is formatted like the example on the right. The example shows a vertical data structure.

| C |
|----------------|
| Air Temp Deg C |
| 22.86 |
| 22.86 |
| 22.86 |
| 23.24 |
| 23.24 |
| 23.24 |

For most data collecting organizations, the only element in the above list which is not normally in the original data file is the Activity ID. This column commonly needs to be added to an organization's dataset to be WQX compliant. This field needs to be unique across time, and sample type. Ex: If you are using both a lab and a probe/sensor for data collection, you will need to use a different activity ID for samples collected by a probe than those analyzed by a lab, even though they were collected at the same date, time, and location. It is suggested to create a formula to generate unique Activity IDs. Those formulas usually include: Monitoring Location ID, Date, Time, and any other activity unique elements. If you look at the "Entity" columns in both figures above, you can see that elements are labeled "Activity", "Result", "Activity Project", etc. If an element is labeled "activity" and exists in your "Regular Columns" section of your import configuration, your Activity ID needs to be unique to all the different values under each of these activity information elements. NOTE: Unique Activity IDs are very important because it is the most atomic level of the dataset. This means, if you reuse an activity ID when submitting data later, it will overwrite the data previously submitted. The system thinks you are updating that old information.