wpdx
water point
data exchange

Fall 2020

WPDx Ingestion Engine: Step-by-step guide for uploading data

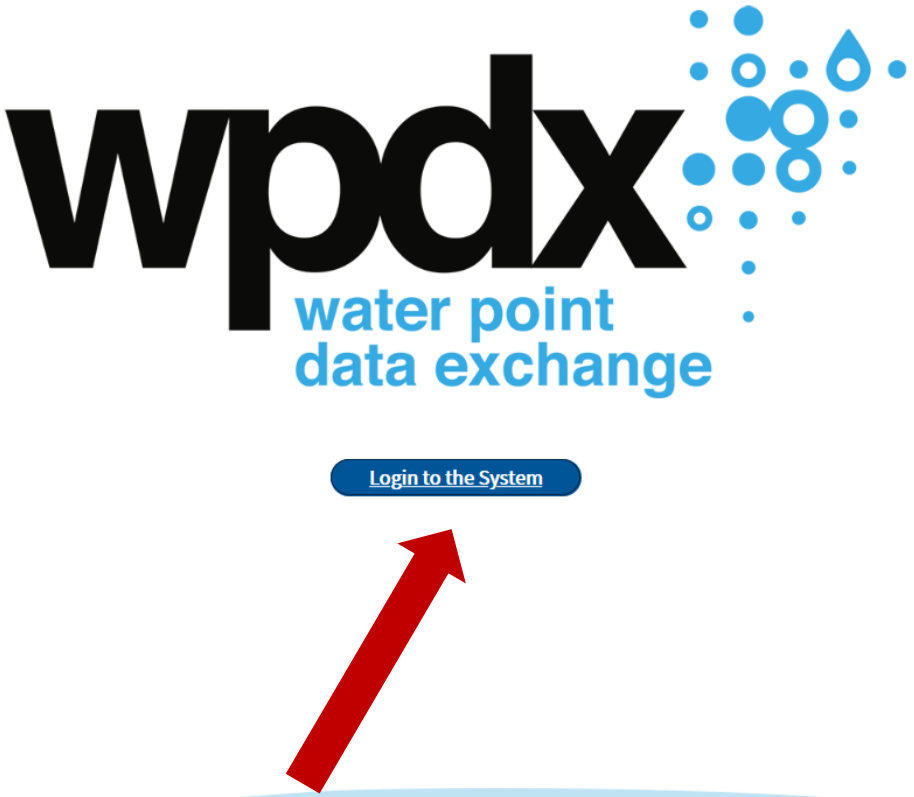
Before you start…

- Review your data to determine which columns of your dataset match with the WPDx Data Standard. You may need to make some minor edits or combinations of columns within your dataset, depending on how your data is organized.
- The "Ingestion Notes" file may be useful to use to prepare for mapping your organization's data to the WPDx Data Standard.
- Have your data file or weblink to the dataset available.

wpdx water point data exchange

Request Access:

- upload.waterpointdata.org
- Select “Login to the System”
- Requires a Google account



Uploading data: Options

There are two options for uploading data:

- Upload a physical file (.xlsx, .xls, .csv) from your computer
- Provide a web link to an API endpoint, Google Sheet, Dropbox or other online system

wpdx water point data exchange

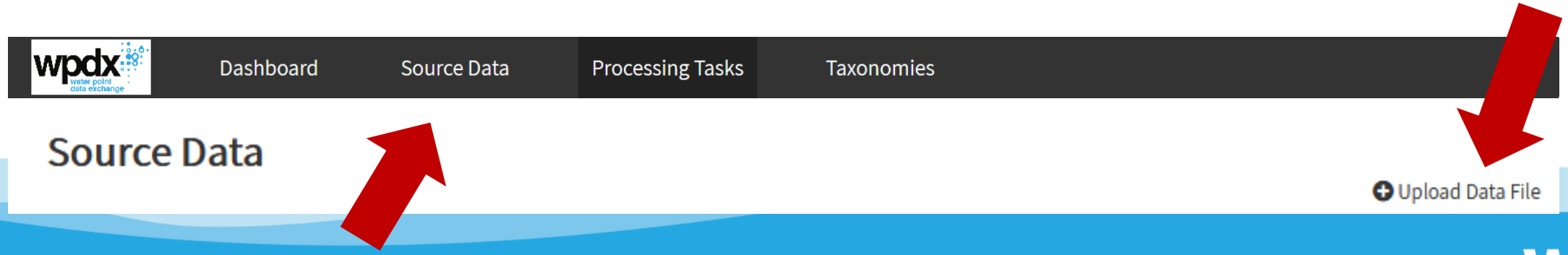
Uploading data: Web link

- Must provide web link with permissions
  - For **Akvo**, request an API endpoint from your program manager. For more details, please see here.
  - For **mWater**, create a datagrid formatted per the WPDx standard. This creates a permanent URL. Click on "Download as XLSX" and copy the download link. Use this in the direct URL box at the beginning of a new processing task. For more details, please see here.
  - For **Dropbox**, copy the download link and not the sharing link to use in the WPDx ingestion engine.
  - For **Google Sheets**, ensure that the document is shared publicly.
  - For custom data platforms, please contact us to determine how we can best connect.
- If the web link does not work, an alternative method is to simply download the source file and upload via the ingestion engine.
- Skip to slide 7

wpdx water point data exchange

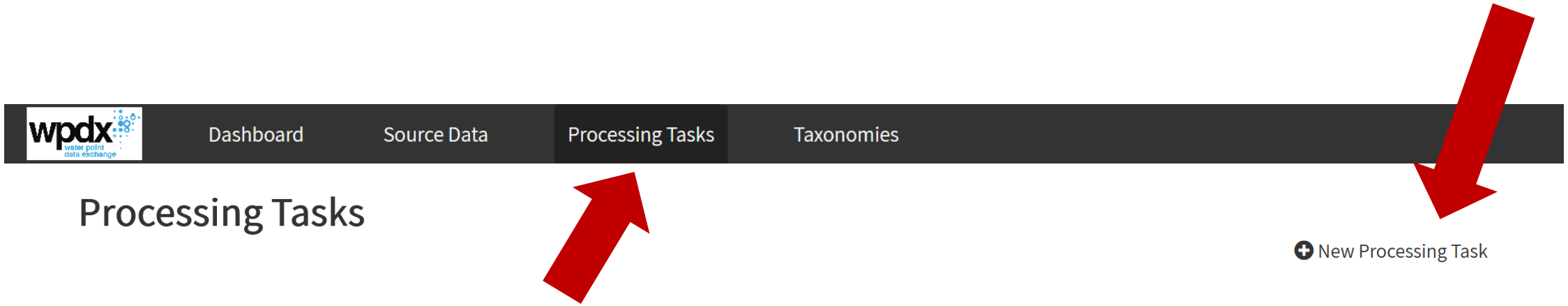
Uploading data: Source Data tab

- Select “Source Data” tab
- Before you upload the file, to help keep the repository organized, please rename the file using the following format:
  - OrgName_Country_MonthYear
  - i.e., GWC_Uganda_Jan2020
- Select “+Upload Data File” and browse to datafile and click “Open”
- “File Upload Successful” message will appear at top of screen



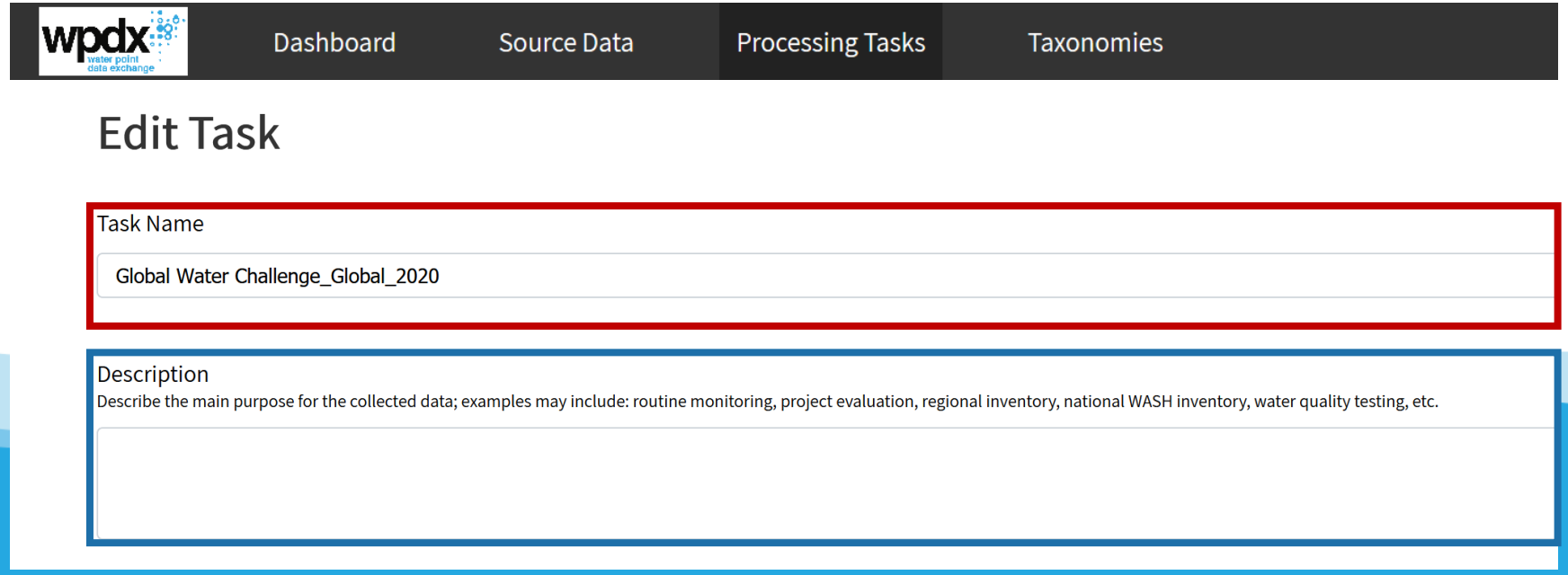
Uploading data: Processing Tasks tab

- Select “Processing Tasks” tab
- Select “+ New Processing Task”



Processing Task: Task Name and Description

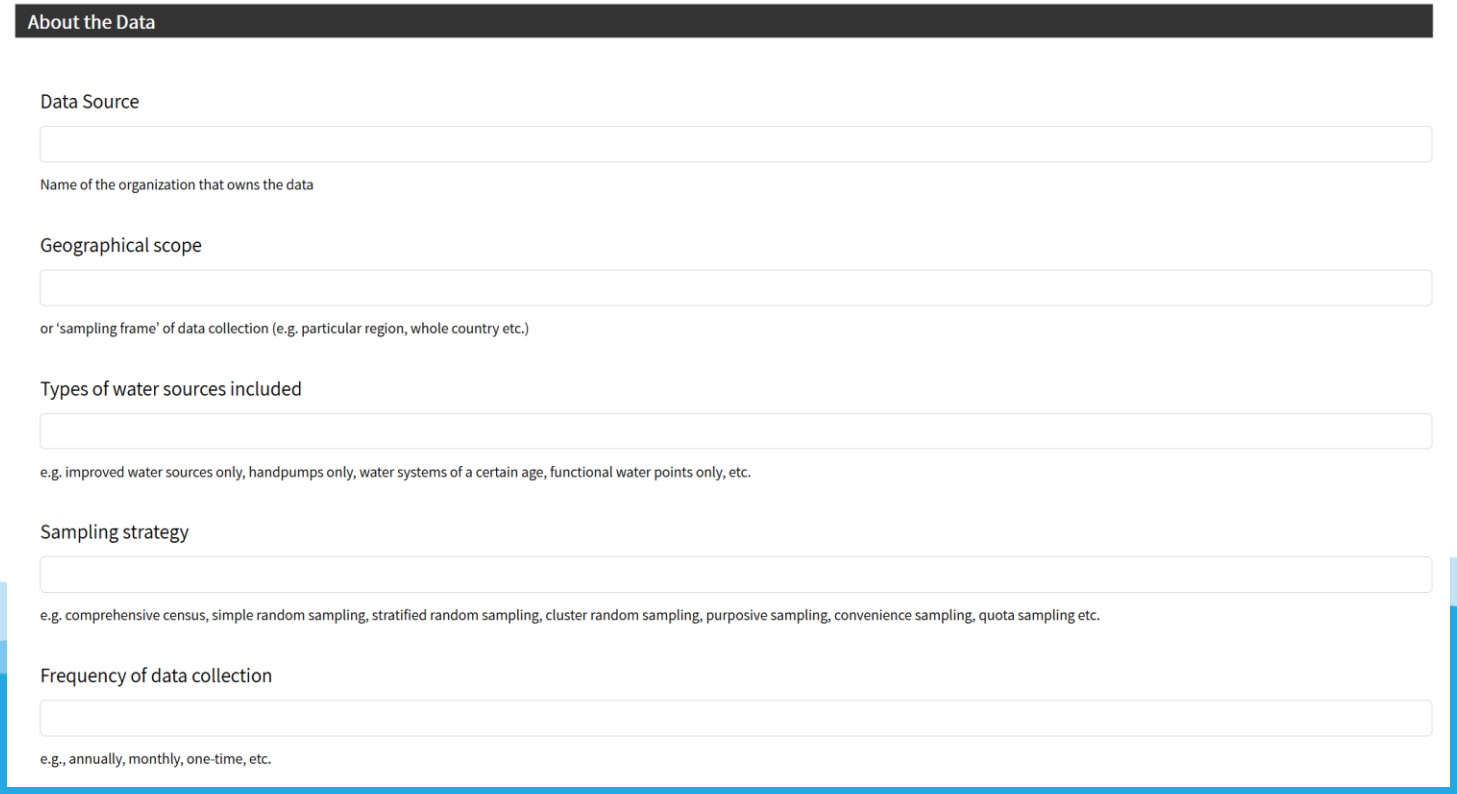
- Enter Task Name in the following format:
  - OrgName_Country/Region_Month/Year of data
  - For example, Global Water Challenge_Global_2019
- Processing tasks can be edited if additional information becomes available
- Provide the main purpose for the collected data under Description



wpdx water point data exchange

Processing Task: Metadata

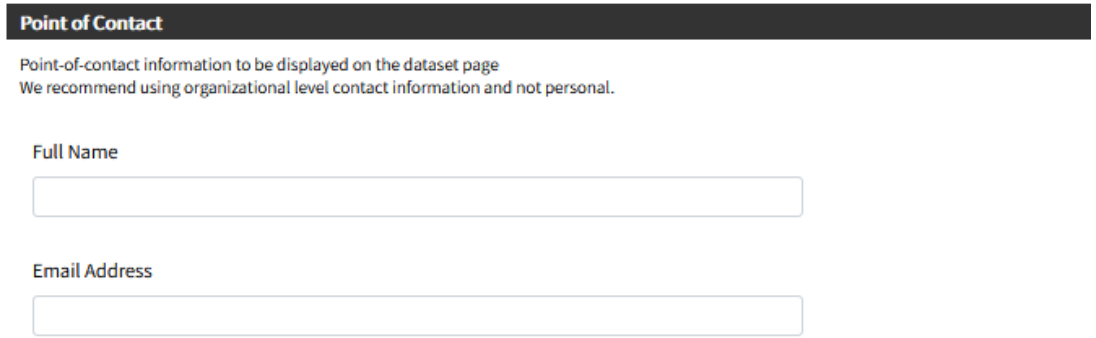
- Complete the metadata prompts to provide a detailed overview of the data within your dataset.
- The metadata will be visible on the data page for your dataset within the WPDx data catalog.



wpdx water point data exchange

Processing Task: Point of Contact

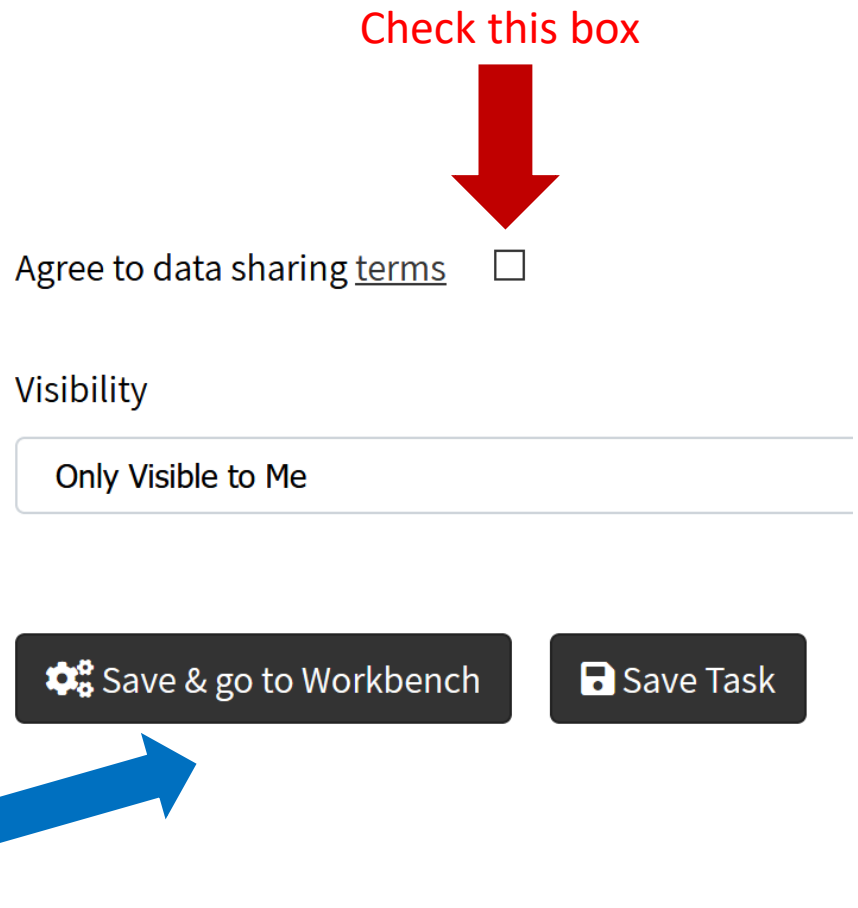
- Complete Point of Contact details for dataset.
- To protect privacy, one option is to use an organizational level email (i.e., data@name.org) which can be forwarded by your organization to relevant contacts.



wpdx water point data exchange

Processing Task: Agree to Data Sharing Terms

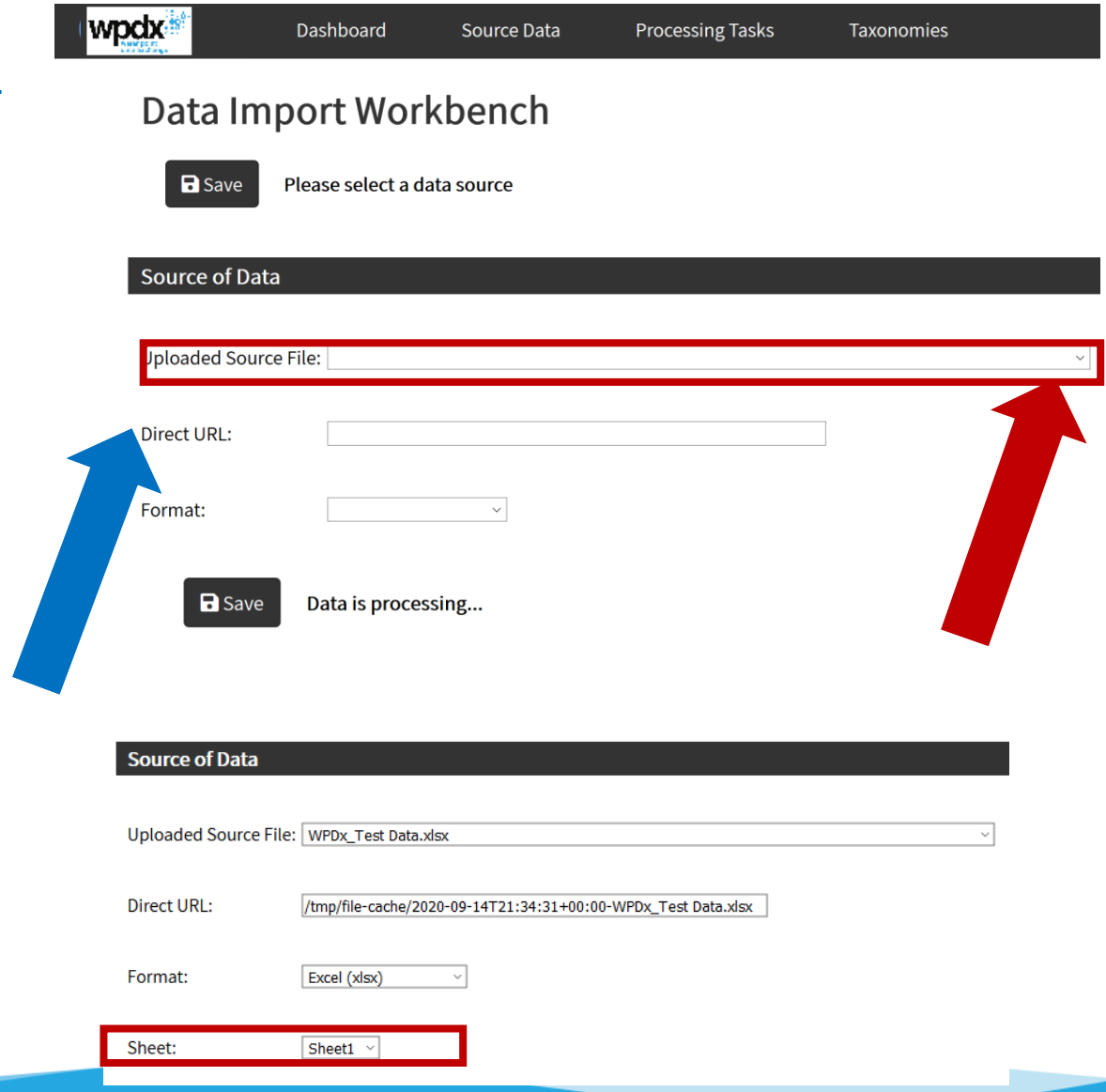
- Check box to agree to Data Sharing Terms
- Leave visibility as “Only Visible to Me”
- Select: Save & go to Workbench



wpdx water point data exchange

Data Import Workbench

- Select source file from dropdown
- Allow data to process (this may take a few minutes). The Direct URL and format boxes will auto-populate.
- If there are multiple sheets in your file, make sure the correct one is selected.
- Scroll down to continue (the “Data is Processing” message may still appear)
- If using a web address, enter directly in Direct URL text box and select the appropriate format option.
  - For JSON formats, be sure to leave the JSON Path field blank.



wpdx water point data exchange

Data Structure

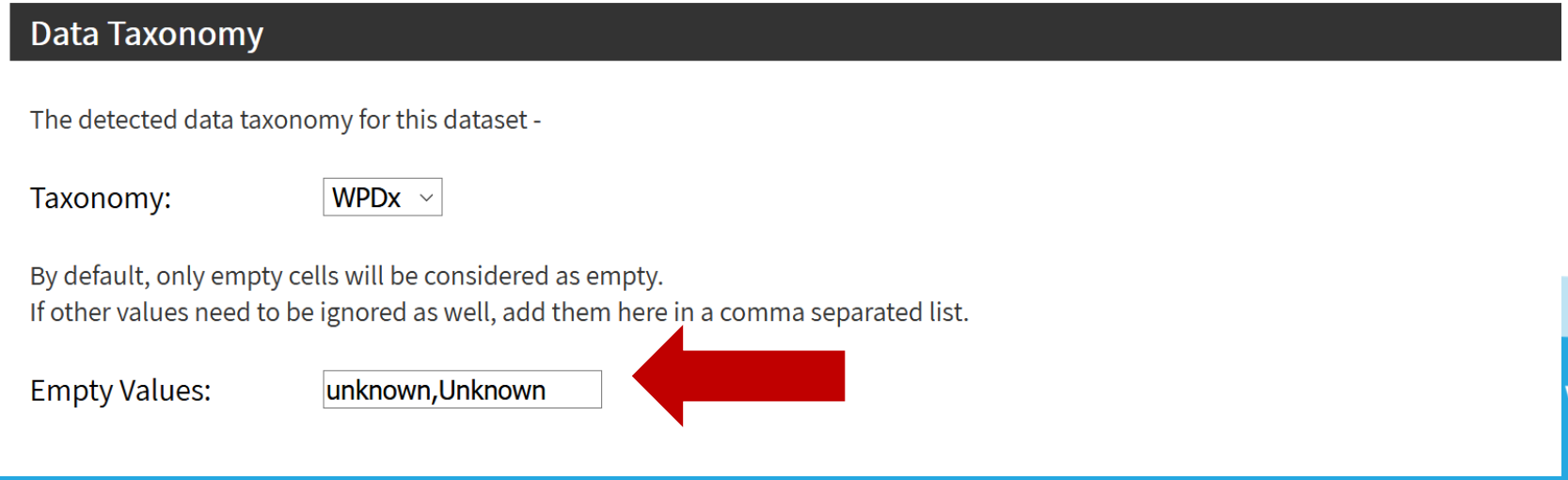
- If your dataset is formatted to include only the column headers and the data, leave Skip Rows/Columns as “0”
- If there are additional rows or columns which should be skipped (i.e., additional headers or title cells) enter the number of rows/columns to skip.

	A	B	C	D	E	F	G	H	I
1	Data for WPDx								
2	Collected in 2020								
3	Latitude	Longitude	Is Water A	Date of Vis	Source Typ	Technolog	District	Subdistrict	Unique ID
4	12.34567	98.6543	Yes	1/1/2020	Borehole	Afridev ha	North	Northwest	B123
5	12.45678	98.5432	No	1/1/2020	Protected	Gravity sch	South	Southwest	PS321
6	12.45678	98.5432	No	1/1/2020	Protected	Tap stand	South	Southwest	T456
7	12.56789	98.4321	Yes	1/2/2020	Borehole	Handpump	West	Southwest	B234
8	12.6789	98.321	No	1/2/2020	Shallow w	Handpump	East	Northeast	SW123
9	12.7891	98.2109	Yes	1/2/2020	Borehole	Handpump	North	Northeast	B345
10	12.7892	98.2108	Yes	1/2/2020	Borehole	Tap Stand	North	Northeast	T123
11	12.7893	98.2107	Yes	1/2/2020	Borehole	Tap Stand	North	Northeast	T124
12	12.34598	98.98765	No	1/3/2020	Well	Handpump	South	Southwest	W123

wpdx water point data exchange

Data Taxonomy

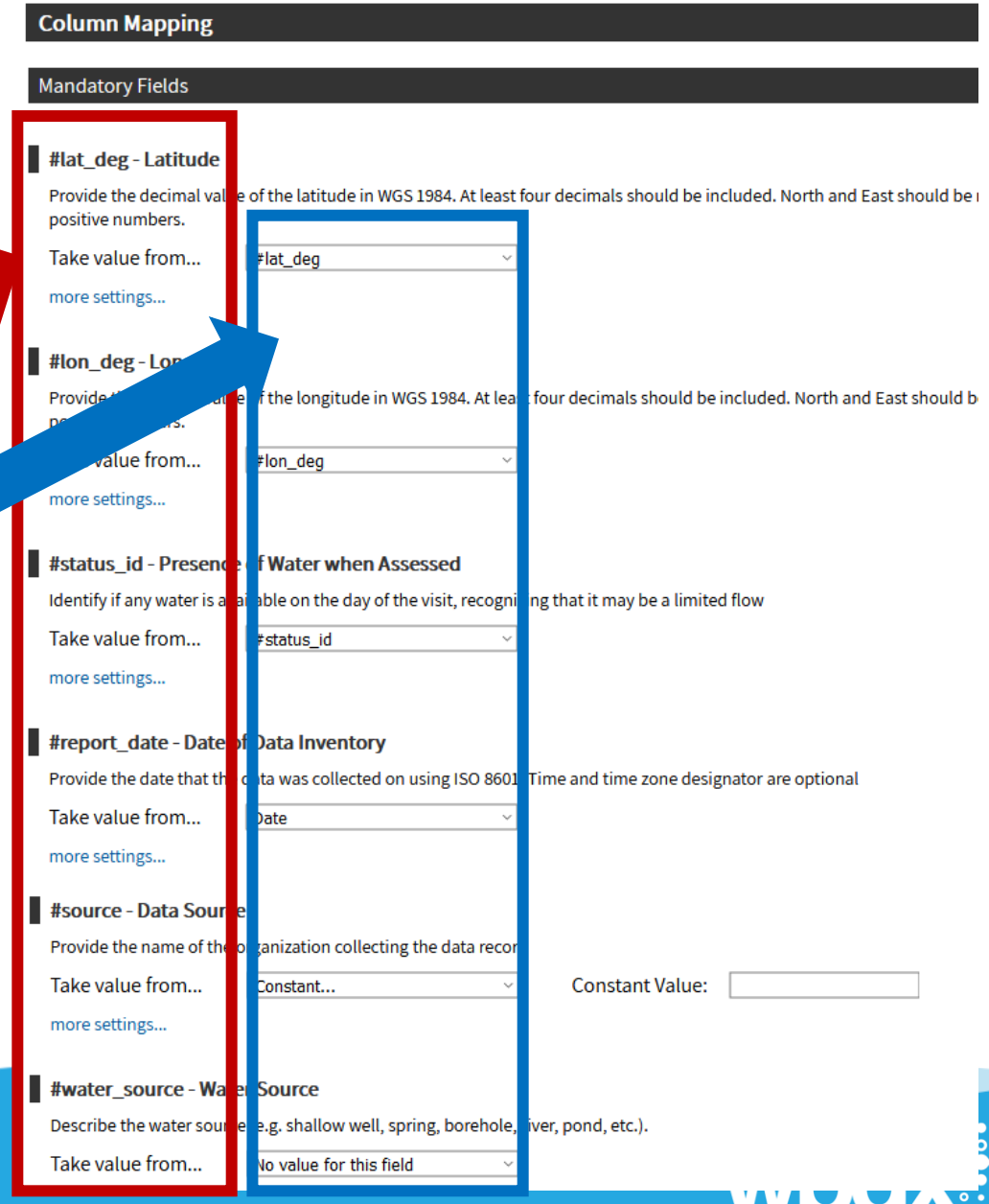
- Leave Taxonomy as WPDx
- For “Empty Values”
  - If your dataset includes terms for blank/unknown values which should be ignored (i.e., Unknown, N/A, etc.), please enter those terms in the text box.
  - Use a comma as a separator between terms. Do **NOT** include any blank spaces between commas and terms.
  - For example: “unknown,Unknown,N/A,0,null,blank”



wpdx water point data exchange

Data Mapping

- This section allows you to map your organization's data to the WPDx Data Standard.
- The WPDx Data Standard parameters are in bold on the left
- The dropdown menus contains your organization's data column headers from the source file.



Data Mapping: Getting Started

There are two methods to complete the data mapping process:

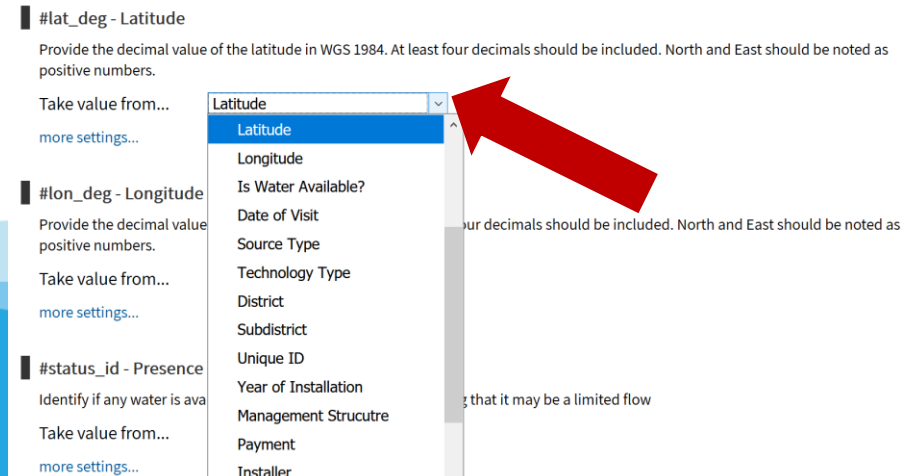
Primary method..

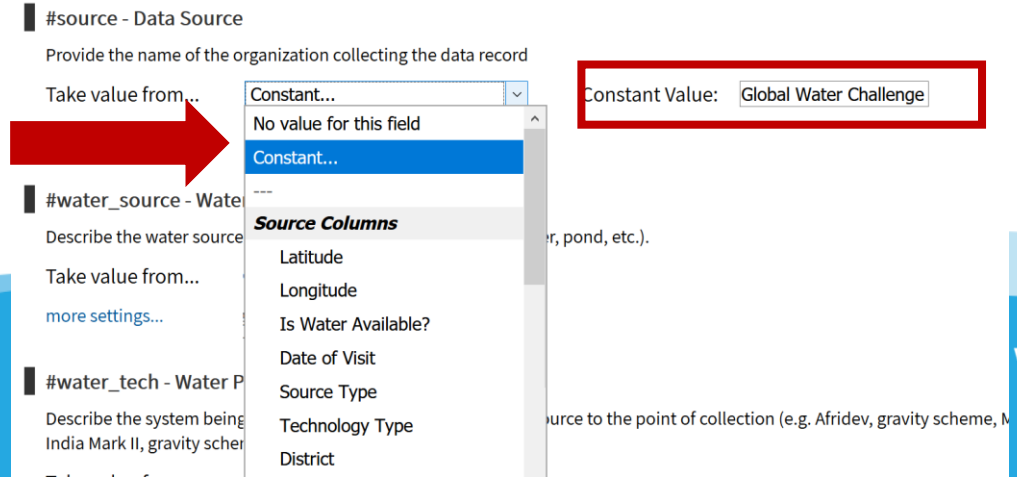
- Using the dropdown menu, scroll to select the column header from your dataset which matches the WPDx standard.
- Some parameters may pre-populate, especially if your dataset is labeled with the WPDx #titles. Verify these selections.
- Note: you cannot map the same column to two different standard parameters.

Optional method..

If there is a parameter which is not in your dataset, but for which a common value can be applied to all datapoints, Select "Constant…" from the dropdown.

- Examples
  - #source – Data Source → Constant: Name of Org
  - #country_id - Country → Constant: "UG" or "GH"
  - #orig_lnk - Public Data Source URL → Constant: URL

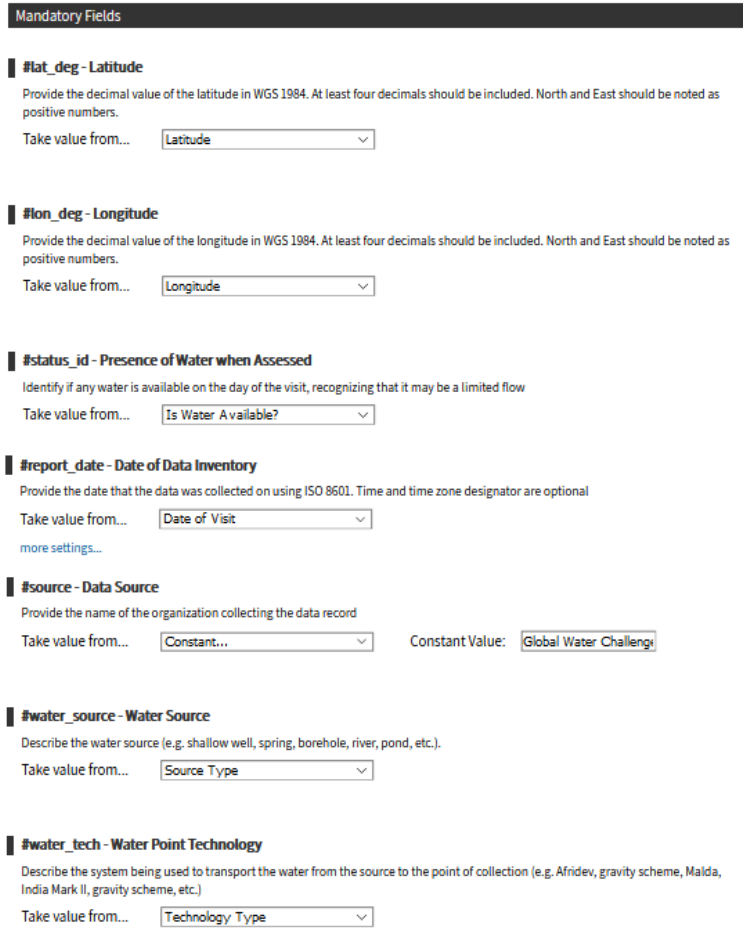




Data Mapping: Required Fields

There are 6 mandatory parameters:

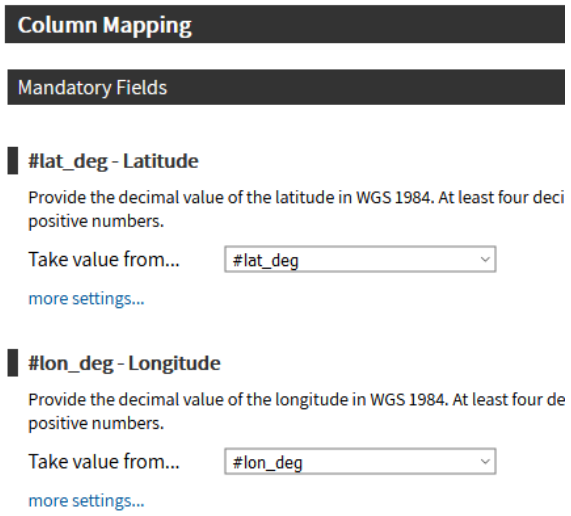
- #lat_deg - Latitude
- #lon_deg - Longitude
- #status_id – Presence of Water when Assessed
- #report_date – Date of Data Inventory
- #source – Organization providing data
- #water_source – Water Source ***AND/OR***
- #water_tech – Water Point Technology

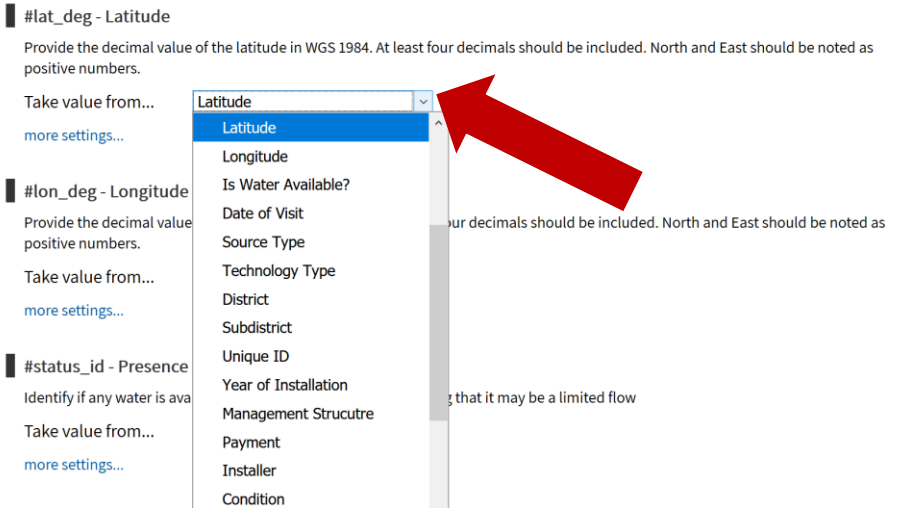


wpdx water point data exchange

Data Mapping: #lat_deg and #lon_deg

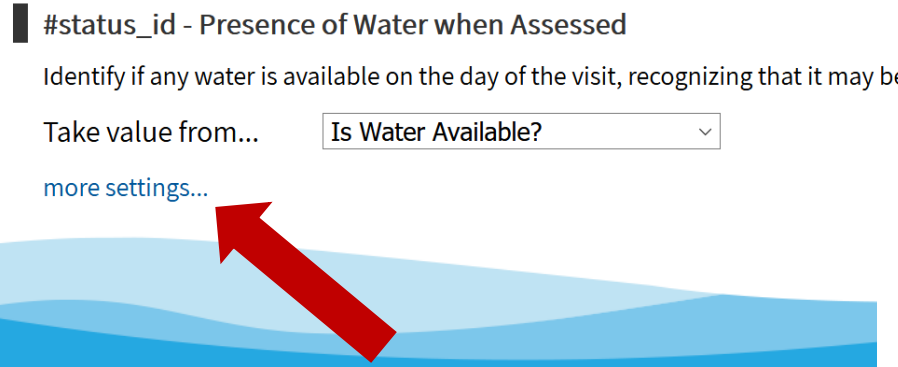
- Latitude and longitude must be in decimal degrees in WGS84.
- Select the appropriate column header which matches with #lat_deg.
- Go the next dropdown and make the selection to match #lon_deg

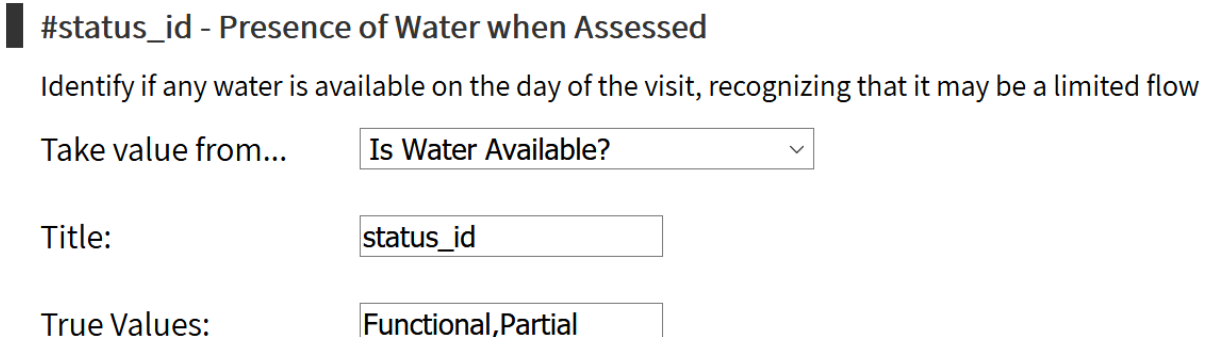




Data Mapping: #status_id

- Select the appropriate column header from the dropdown
- Default values include Yes/No. “Unknown” values (see slide 14) will be converted to a blank cell in the WPDx Global Data Repository
- If your dataset does not include Yes/No, but instead terms such as “Functional/Partial/Non-functional” select “more settings..” and enter those terms.
  - True Values = terms which indicate the water point IS functional
  - False Values = terms which indicate the water point is NOT functional
  - Do not leave any spaces between terms, just a comma (i.e., Yes,functional)

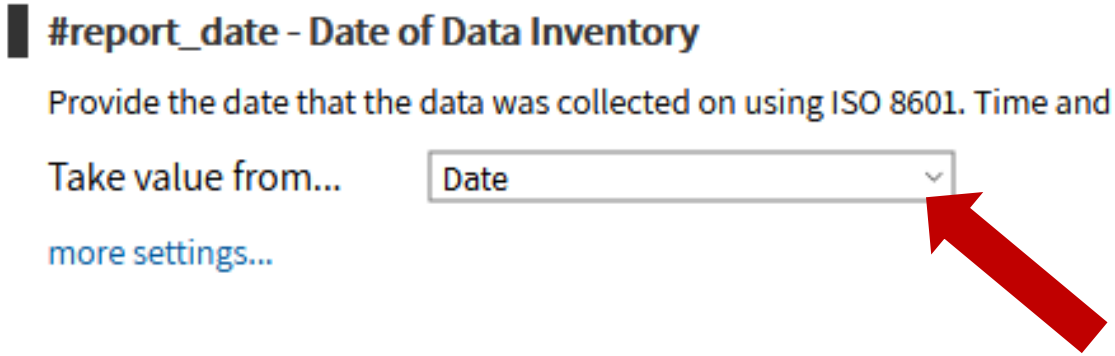


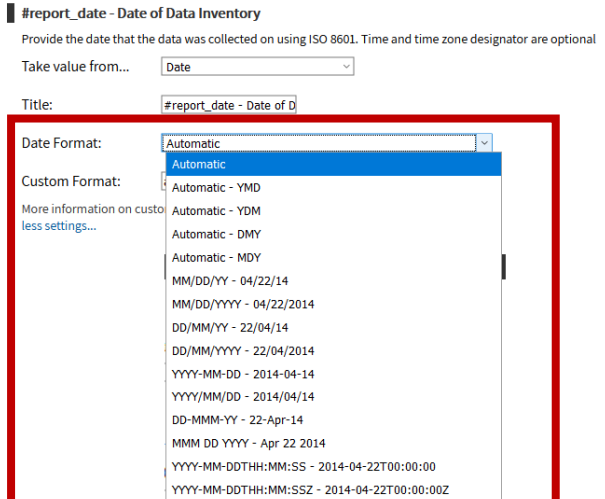


wpdx water point data exchange

Data Mapping: #report_date

- Select the appropriate column header from the dropdown
- The system will automatically detect the format of the dates in your dataset
- If there are errors indicated, select “more settings…” and choose a specific format. (This should only be an issue in rare circumstance)

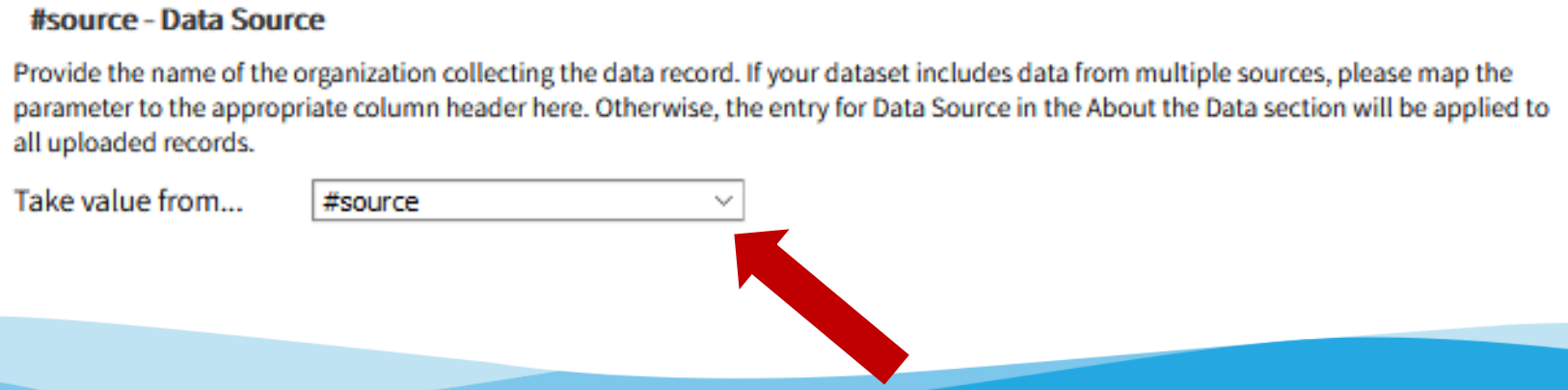




wpdx water point data exchange

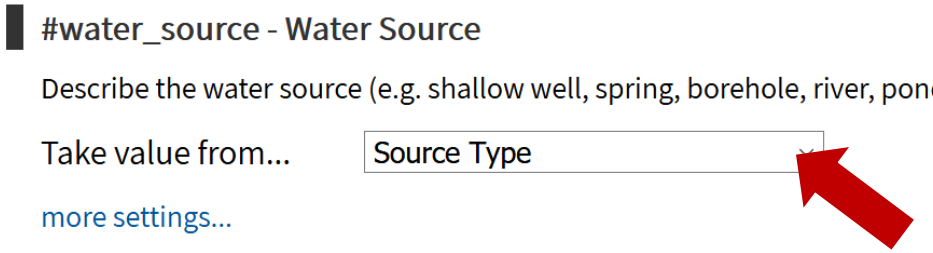
Data Mapping: #source

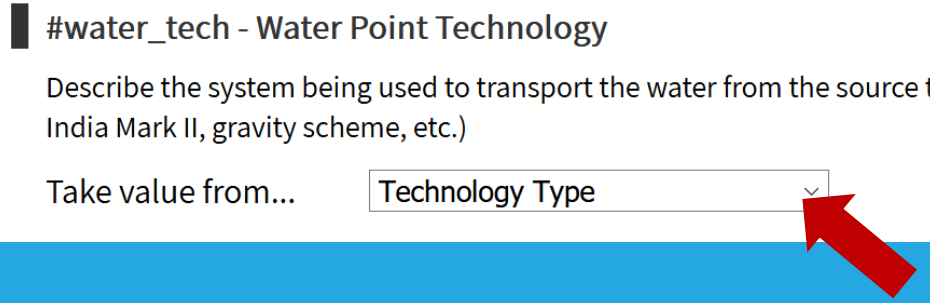
- Provide the name of the organization providing the data.
- If your dataset includes data from multiple sources, please map the parameter to the appropriate column header.
- Otherwise, the entry for Data Source in the About the Data section will be applied to all uploaded records.



Data Mapping: #water_source & #water_tech

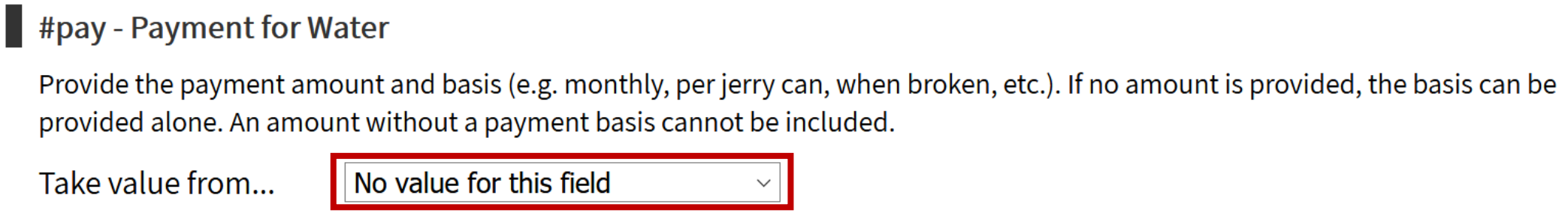
- At least one of #water_source or #water_tech must be mapped for the upload to proceed.
- Select the appropriate column header/s from the dropdown
- If the information is constant for all values, you can instead select “Constant.. “ and enter in the appropriate value in the text box.





Data Mapping: Optional Fields

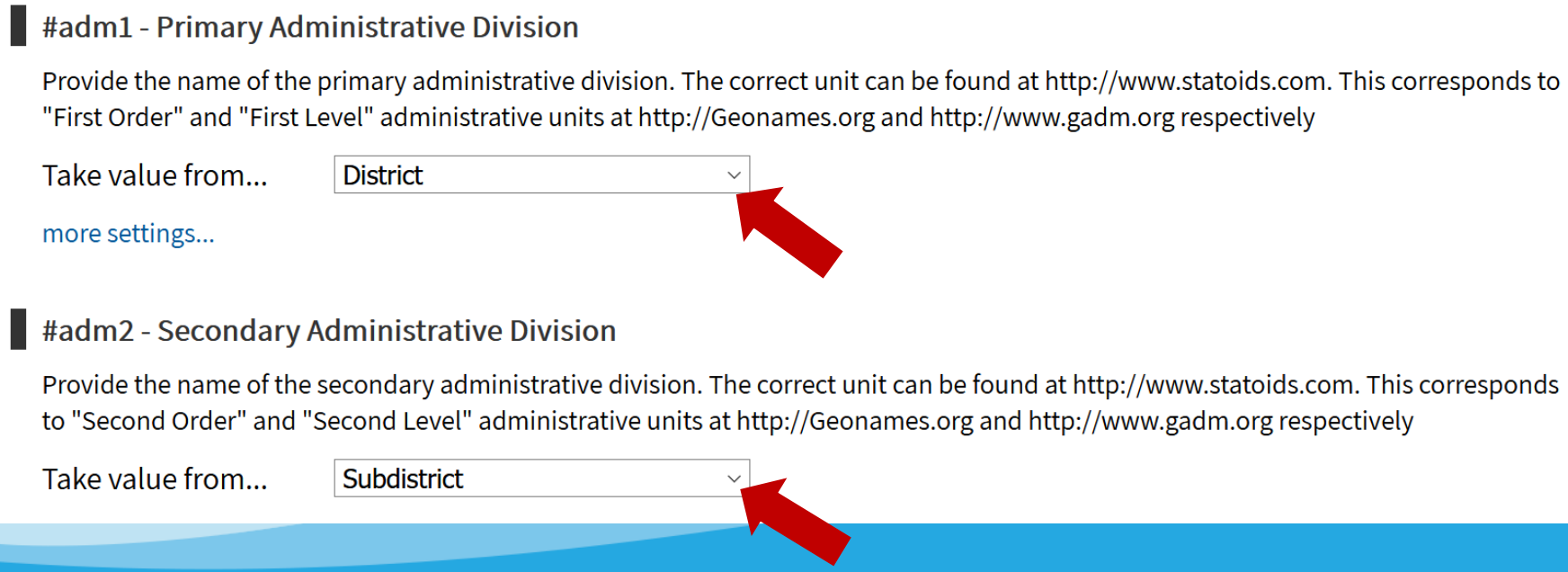
- The “Optional Fields” are not required, but they do help to provide a more robust dataset for understanding the status of the local water sector. **Please map as many of the WPDx parameters as possible.**
- For any parameters which do not align with your dataset, you can select “No value for this field” (this is the default selection) and go on to the next parameter.
- For example, if your dataset does not include any information on payment:



wpdx water point data exchange

Data Mapping: #adm1 and #adm2

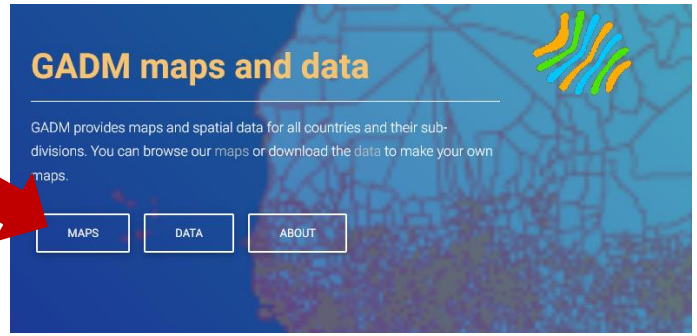
- #adm1 and #adm2 are official administrative division designations
  - If you have questions, look at GADM.org (see tutorial on next slides) or statoids.com to determine the appropriate designations.



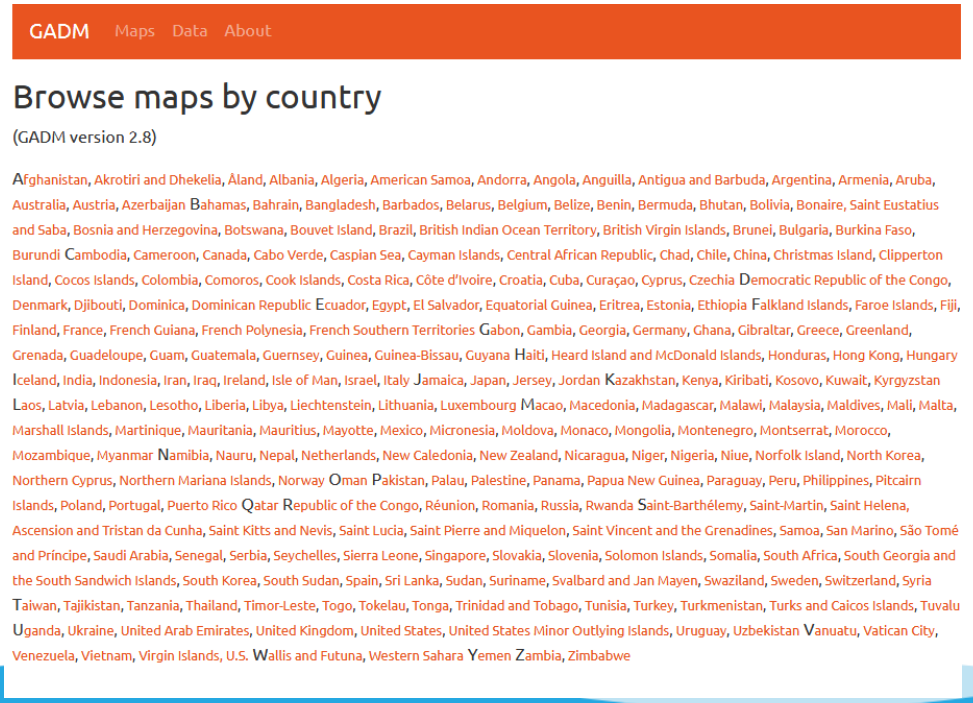
wpdx water point data exchange

GADM.org: Check administrative divisions

1. Go to GADM.org and Select “Maps”



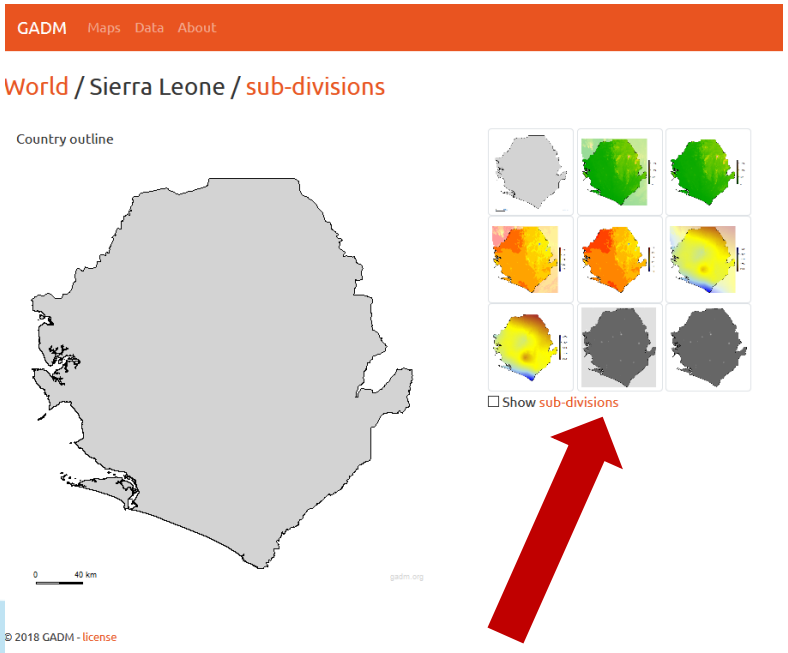
2. Click on country of interest



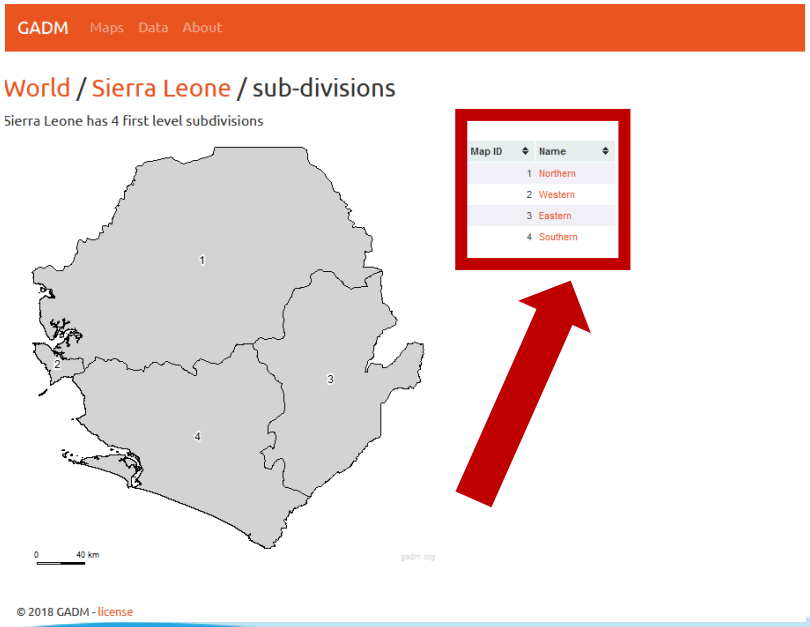
wpdx water point data exchange

GADM.org: Check administrative divisions

3. Select “Show sub-divisions”



4. This creates a map and a list of first-level subdivisions

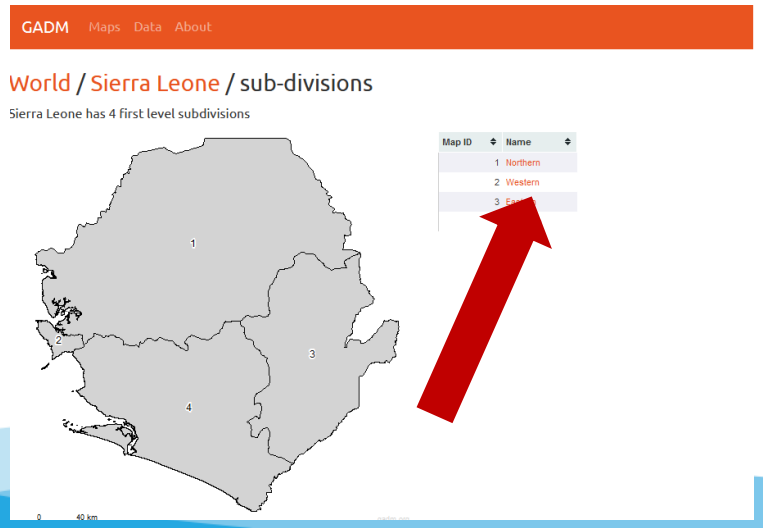


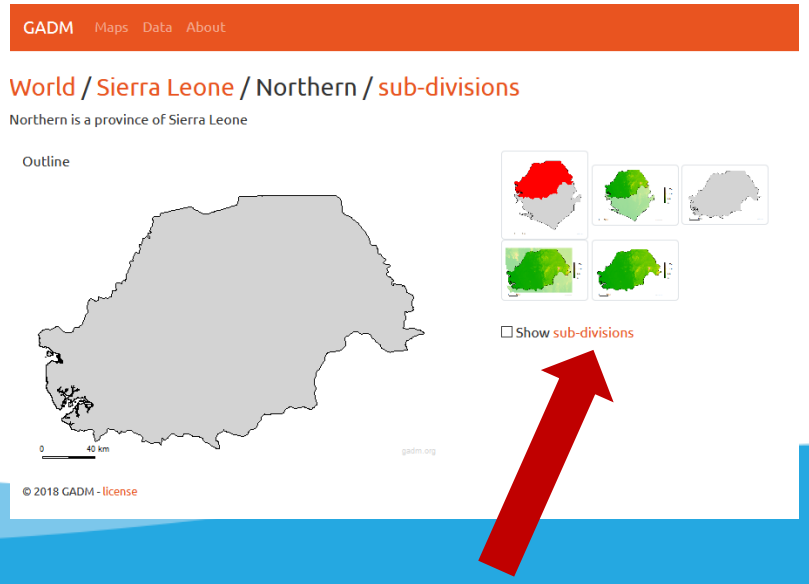
GADM.org: Check administrative divisions

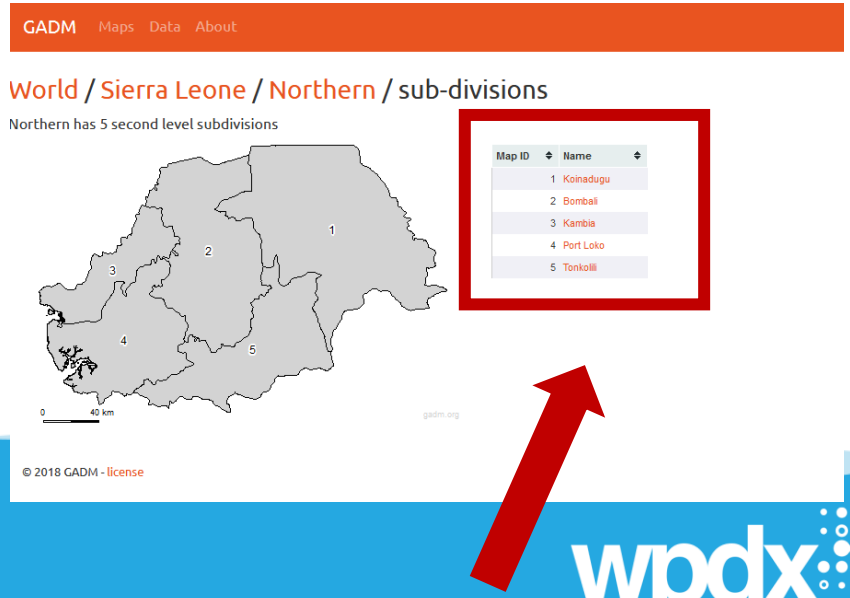
5. Click on one of the first level sub-divisions

6. Click on “Show sub-divisions

7. This creates a map and list of second level subdivisions



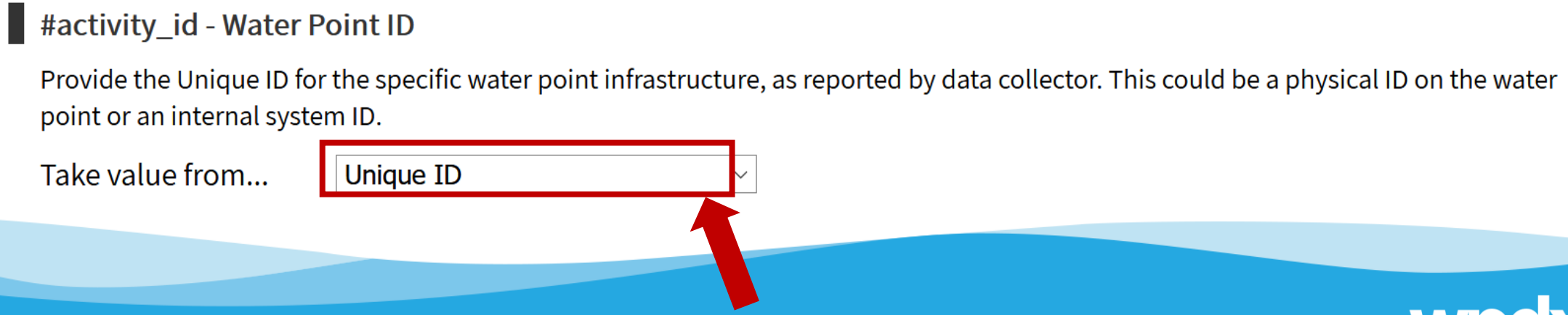




wpdx water point data exchange

Data Mapping: #activity_id

- Select the appropriate column header from the dropdown
  - If a locally or globally recognized standardized identification number exists (i.e., a physical well ID number of barcode) within your dataset, please use that column, **OR**
  - If your organization has a unique id system which would allow water points to be matched within your organization over time, please use that column

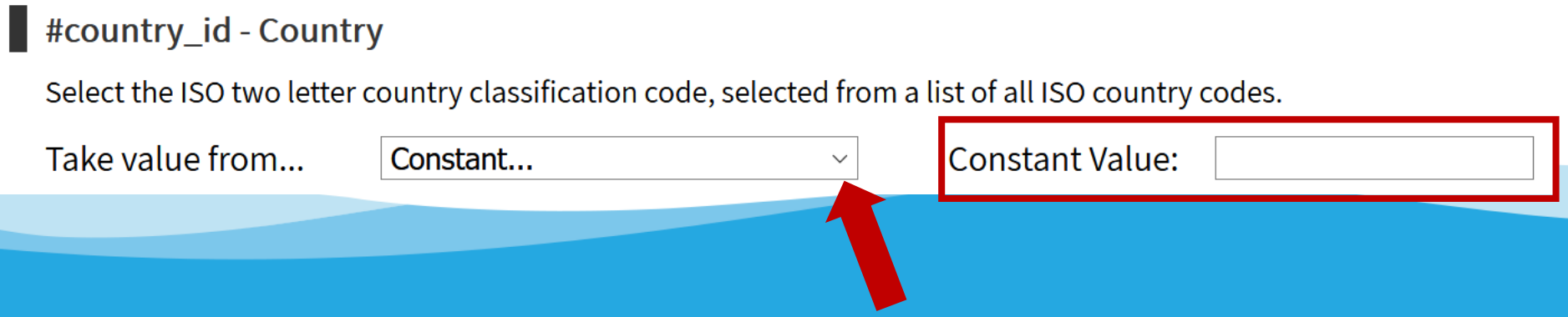


wpdx water point data exchange

Data Mapping: #country_id

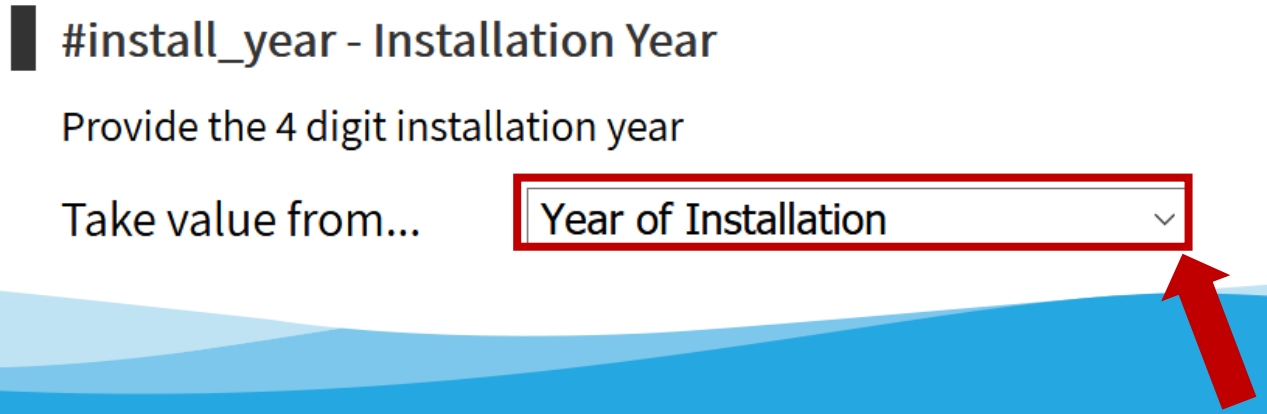
Select the ISO two letter country classification code, selected from a list of all ISO country codes.

- If your dataset includes entries from different countries, this information should be included in your data file. Select the appropriate column header from the dropdown menu.
- If your dataset only includes entries from a single file, you can select “Constant..” and enter a value to be applied to all rows.



Data Mapping: #install_year

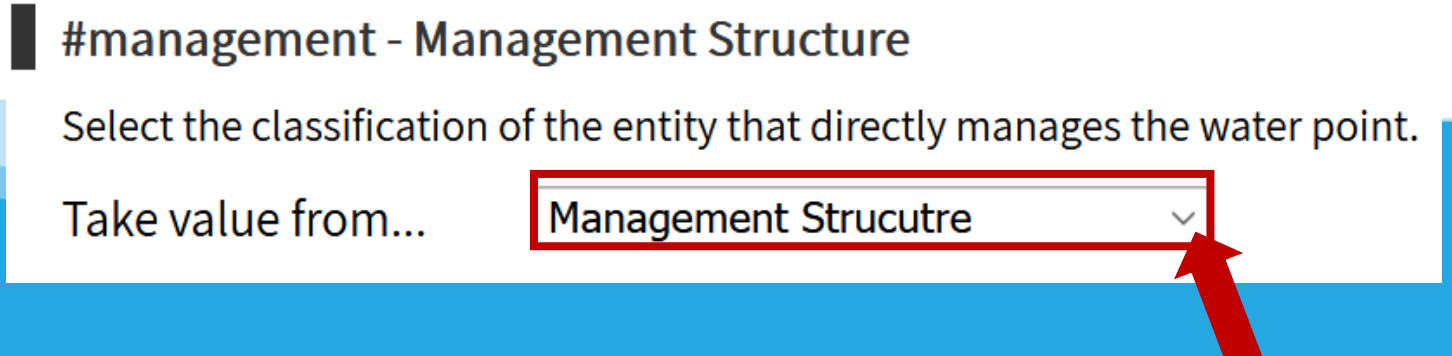
- Select the appropriate column header from the dropdown
- Note that this field accepts only a four-digit year, and not a full installation date.



wpdx water point data exchange

Data Mapping: #management

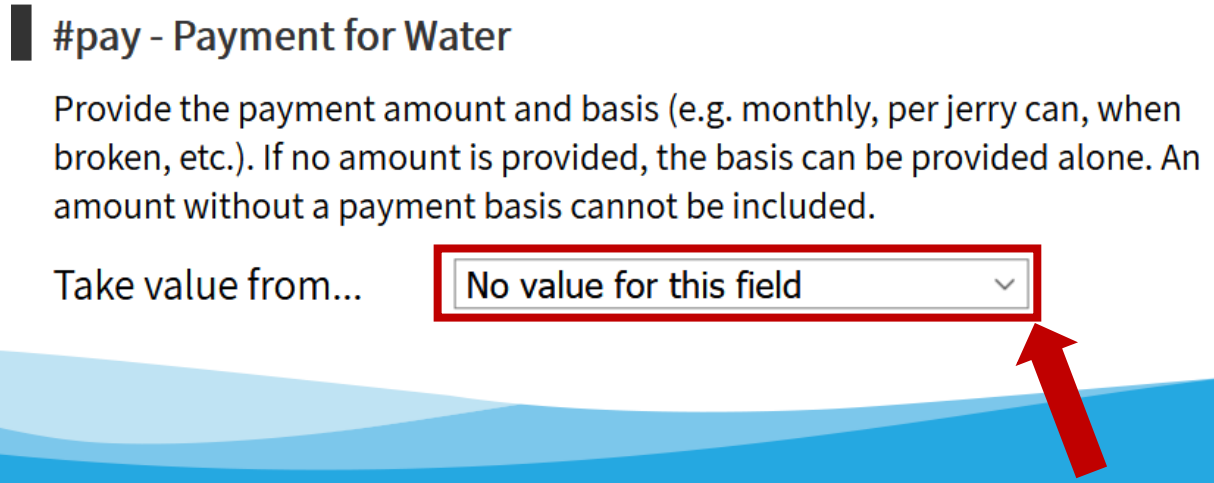
- Select appropriate column header from dropdown.
- Select the management classification of the entity that directly manages the water point. Example management types include:
  - Direct Government Operation
  - Private Operator/Delegated Management
  - Community Management
  - School
  - Healthcare Facility
  - Other Institutional Management
  - Other



wpdx water point data exchange

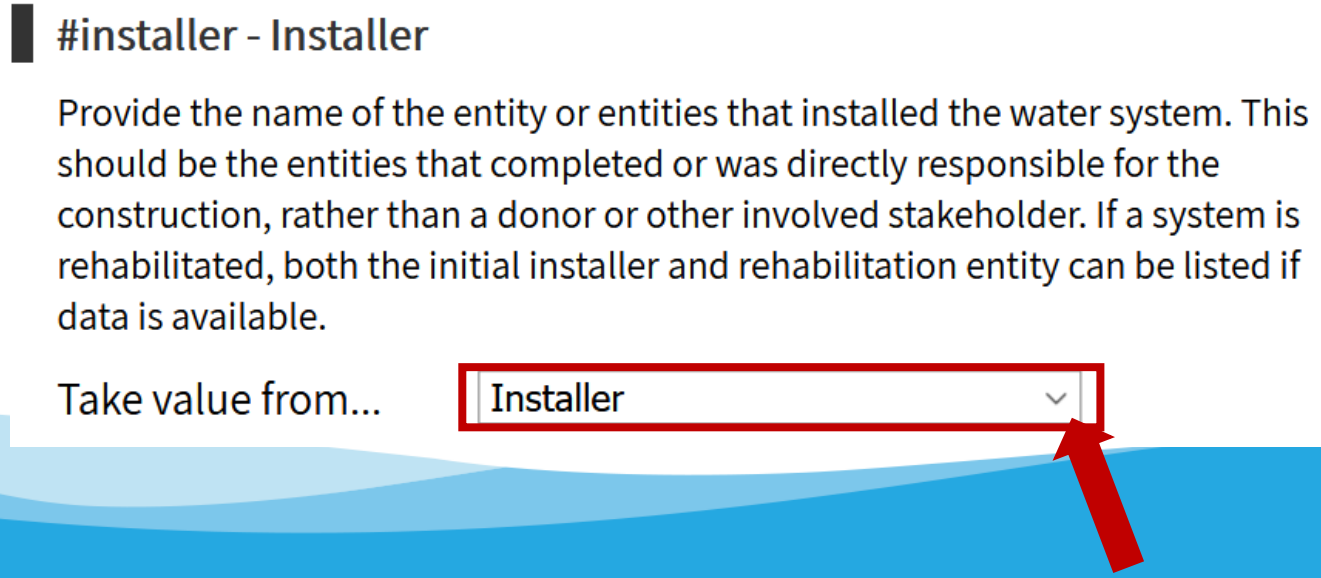
Data Mapping: #pay

- Select appropriate column header from dropdown.



Data Mapping: #installer

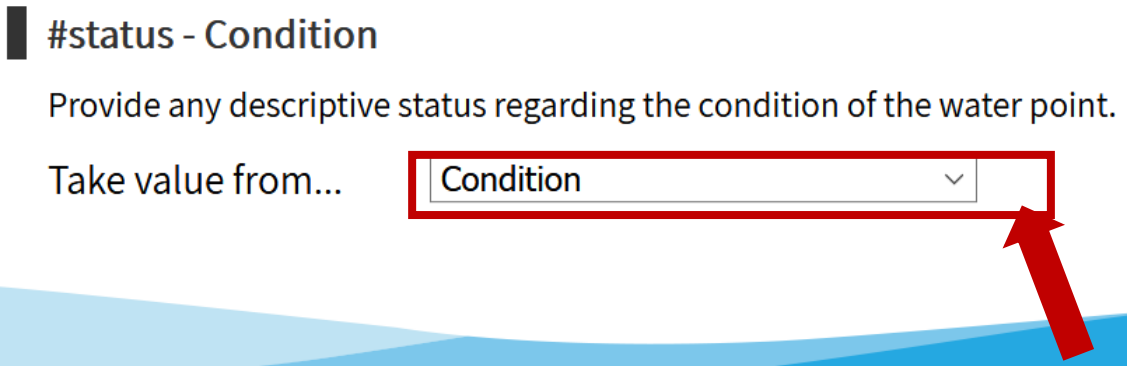
- Select appropriate header from dropdown



wpdx water point data exchange

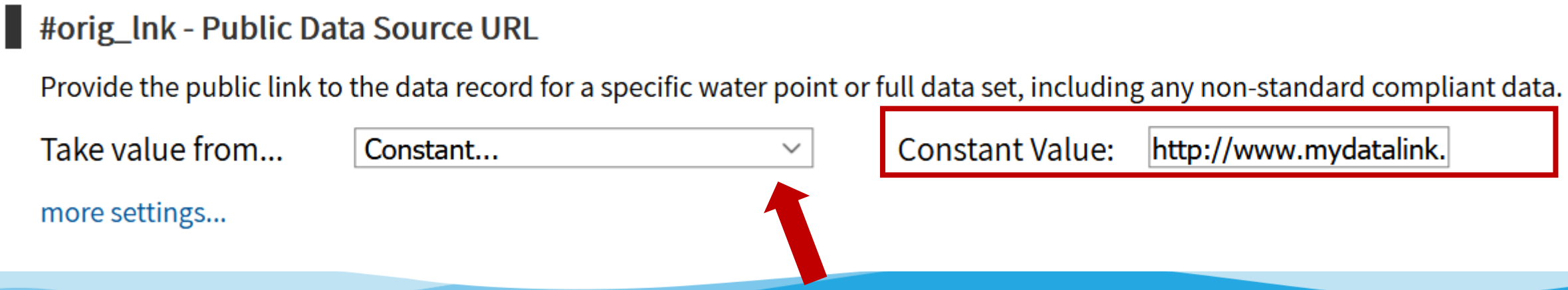
Data Mapping: #status

- Select appropriate header from dropdown



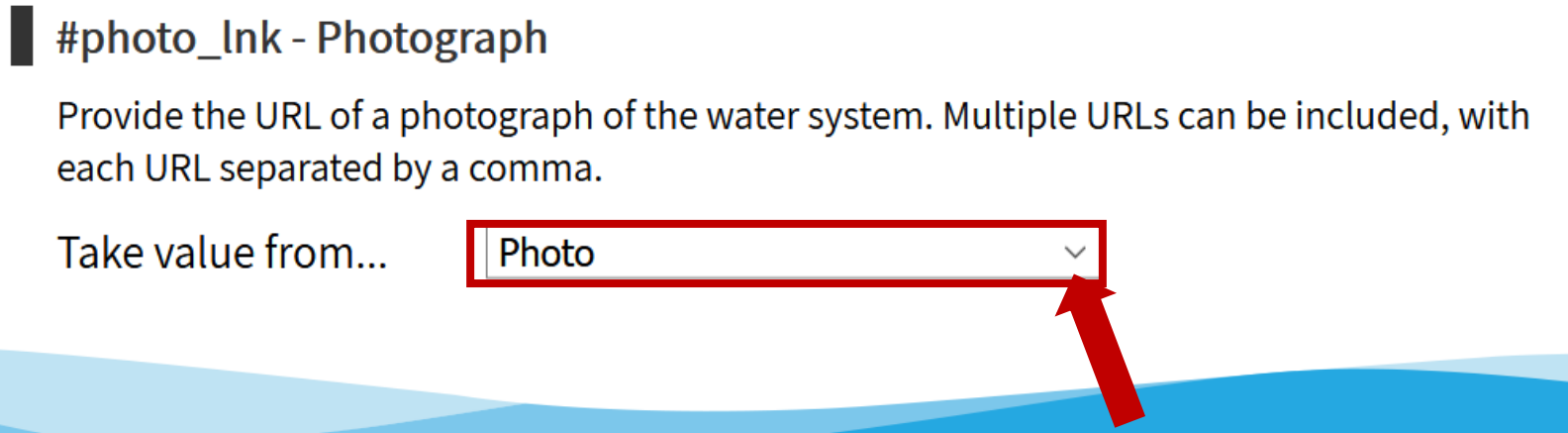
Data Mapping: #orig_lnk

- If the data is available via a public link, select ‘Constant’ from the dropdown and enter it so that it can be applied to all rows.
- If there is to a public link, leave as ‘No value for this field’



Data Mapping: #photo_lnk

- Select appropriate column header from dropdown



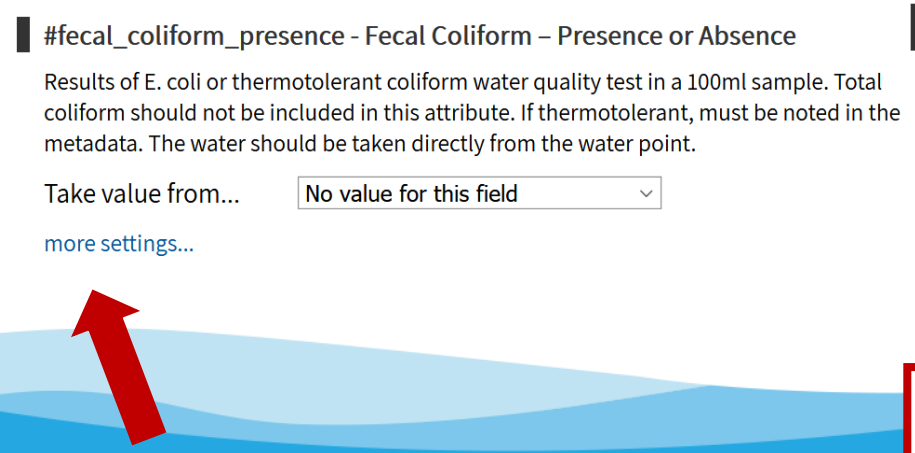
wpdx water point data exchange

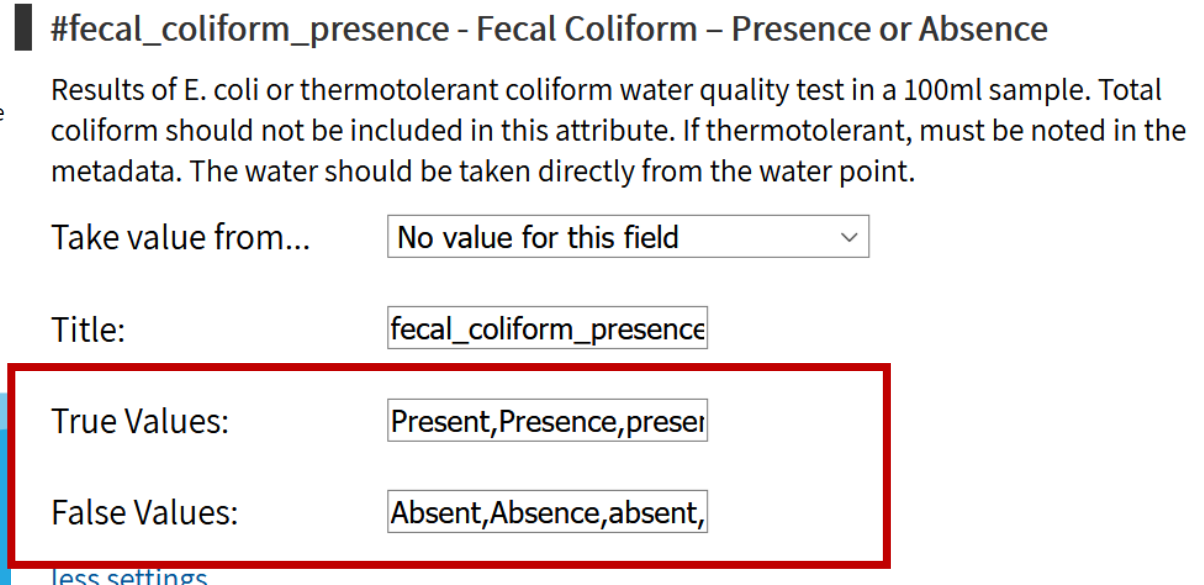
Data Mapping: #converted

- The #converted parameter is for internal use by WPDx to indicate when data has been interpreted by WPDx staff to fit within the standard.
- This only applies when WPDx staff have uploaded data on the behalf of another organization.

Data Mapping: #fecal_coliform_presence

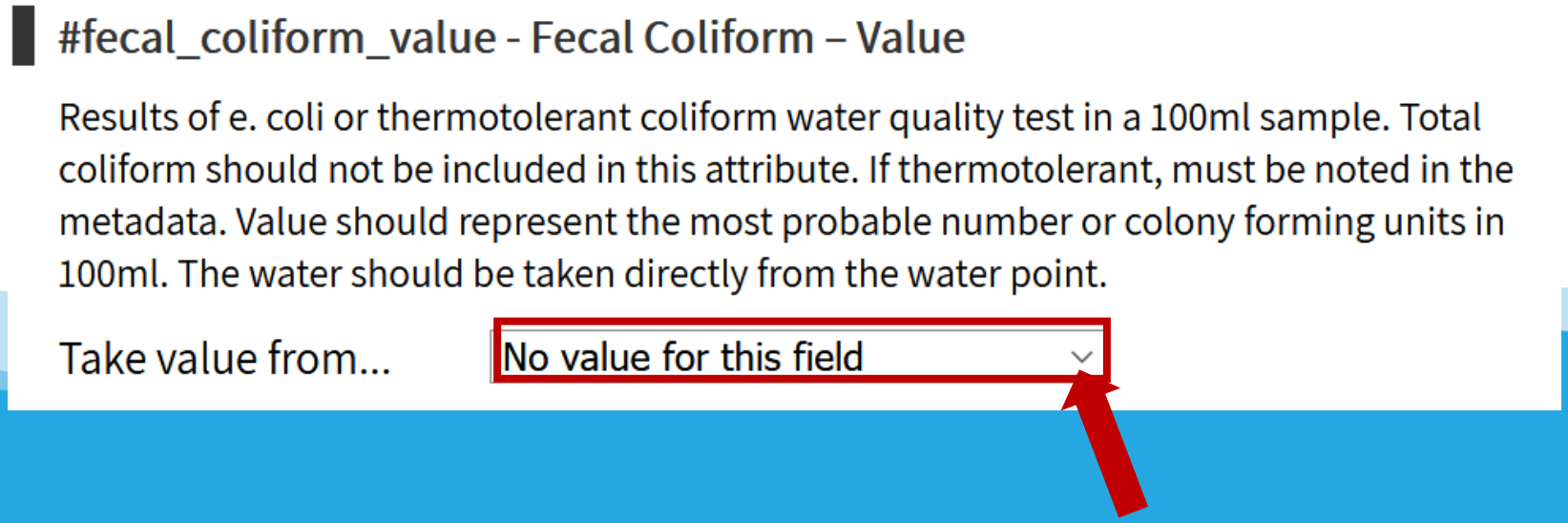
- Select appropriate column header from the dropdown
- Default values include Present/Presence and Absent/Absence. If your dataset include other terms, select ‘more settings…’ and enter the terms into the True Value and False Value text boxes.
- Separate terms with a comma but do not include any spaces.
- Complete associated metadata questions





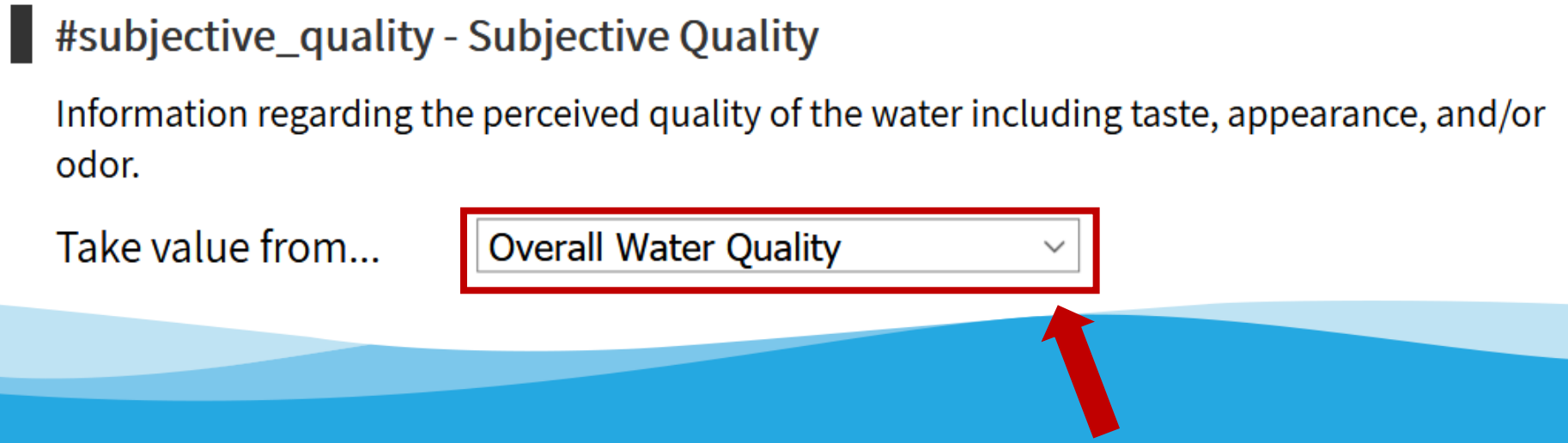
Data Mapping: #fecal_coliform_value

- Select appropriate column header from dropdown
- Complete associated metadata questions



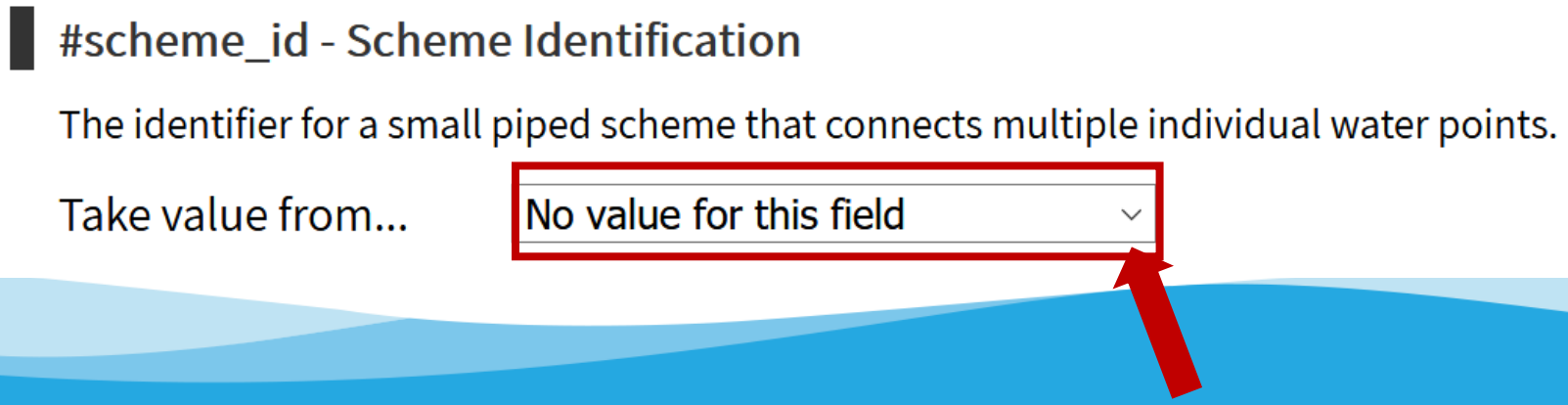
Data Mapping: #subjective_quality

- Select the appropriate column header from dropdown



Data Mapping: #scheme_id

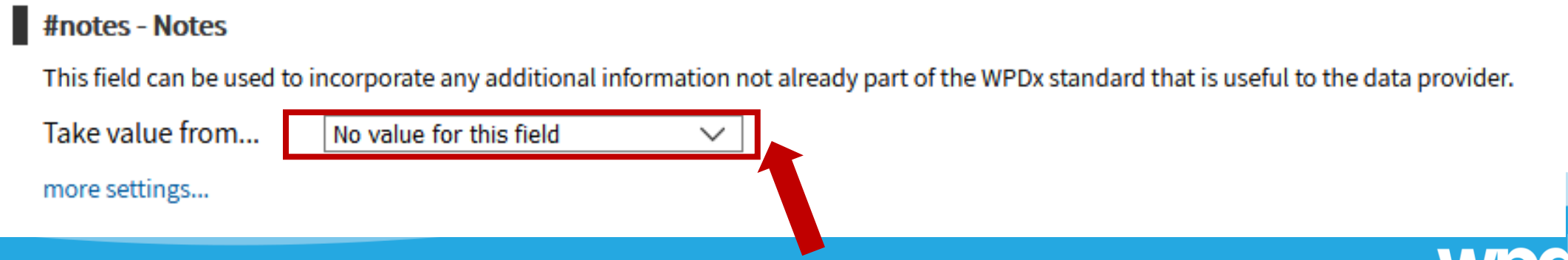
- Select the appropriate column header from dropdown



wpdx water point data exchange

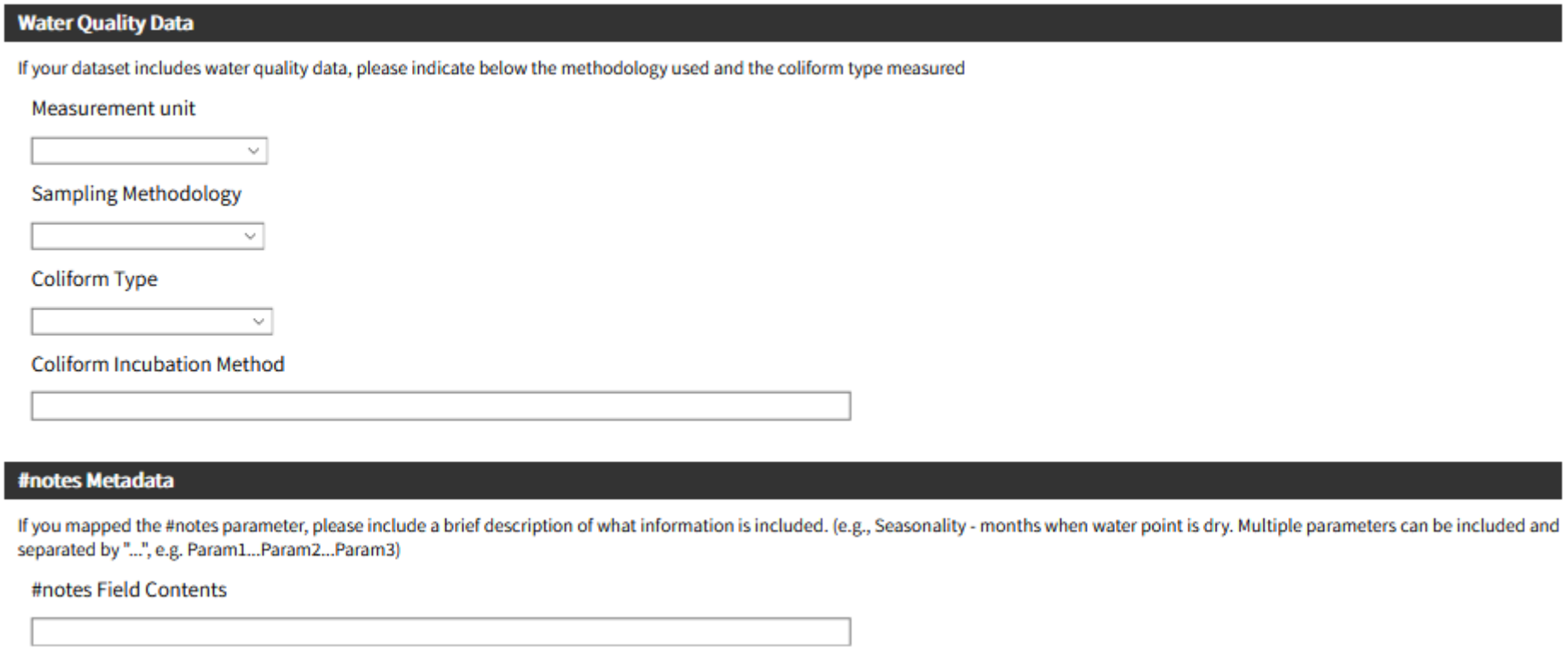
Data Mapping: #notes

- Select appropriate column from header or apply Constant value is appropriate.
- The #notes parameter can be used to enter custom data which the host country government or organization has selected.
  - For example, some organizations want to track seasonality, additional administrative districts, or some combination.
  - Multiple parameters can be included by creating a column that includes the parameters of interest, separated by a “;” or “…” delimeter.



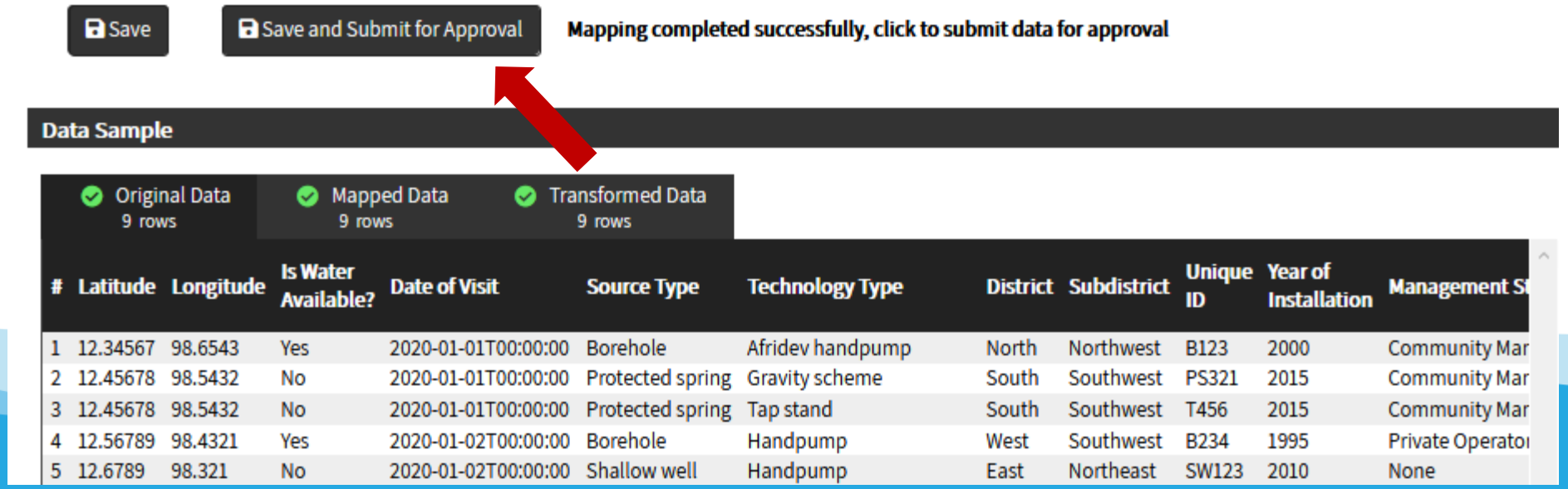
Water Quality and Notes Metadata

- If you mapped the #fecal_coliform_presence, #fecal_coliform_value or #notes columns, please complete the additional metadata question section.



Once mapping is complete

- Select “Save” or “Save and Submit for Approval”
  - Select Save and Submit for Approval when your data has been fully mapped and is ready for upload
- The status in the Processing Tasks tab will now show as “Pending”
- An administrator will be notified and will complete the uploading process
- Once approved, an email will be sent to the uploader’s email address
- If the mapping was not successful, you will see an error message indicating which parameter was not mapped and explanation of why. Once the error has been fixed, you can submit the processing task for approval.

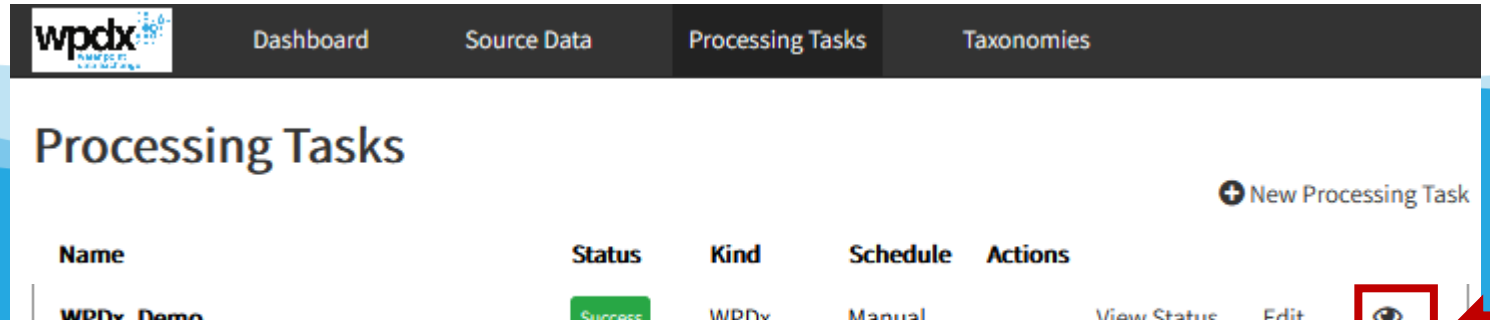


#	Latitude	Longitude	Is Water Available?	Date of Visit	Source Type	Technology Type	District	Subdistrict	Unique ID	Year of Installation	Management St
1	12.34567	98.6543	Yes	2020-01-01T00:00:00	Borehole	Afridev handpump	North	Northwest	B123	2000	Community Mar
2	12.45678	98.5432	No	2020-01-01T00:00:00	Protected spring	Gravity scheme	South	Southwest	PS321	2015	Community Mar
3	12.45678	98.5432	No	2020-01-01T00:00:00	Protected spring	Tap stand	South	Southwest	T456	2015	Community Mar
4	12.56789	98.4321	Yes	2020-01-02T00:00:00	Borehole	Handpump	West	Southwest	B234	1995	Private Operato
5	12.6789	98.321	No	2020-01-02T00:00:00	Shallow well	Handpump	East	Northeast	SW123	2010	None

wpdx water point data exchange

Successful upload

- Once the data upload has been completed by an administrator, the status in the Processing Task will be marked as “Success”
- You can view an overview of the dataset by clicking on the eye icon
- The overview page includes:
  - Metadata and contact details
  - Ingestion report - summary statistics of the number of rows uploaded and any errors encountered.
  - Link to download source file
- Data will be visible on the WPDx data repository within 24 hours



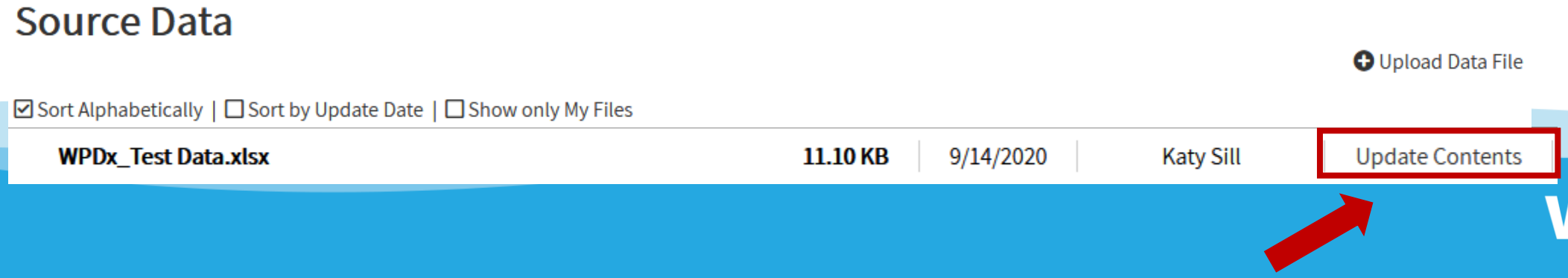
Need to make changes?

- Users can edit their datasets and processing tasks to correct errors or make other additions (i.e., add a new column that was not previously mapped).
- To remove data from WPDx, please contact the administrator at info@waterpointdata.org with “Request to remove data from WPDx” in the subject headline. Include the name of the source file and the reason for the removal request.

wpdx water point data exchange

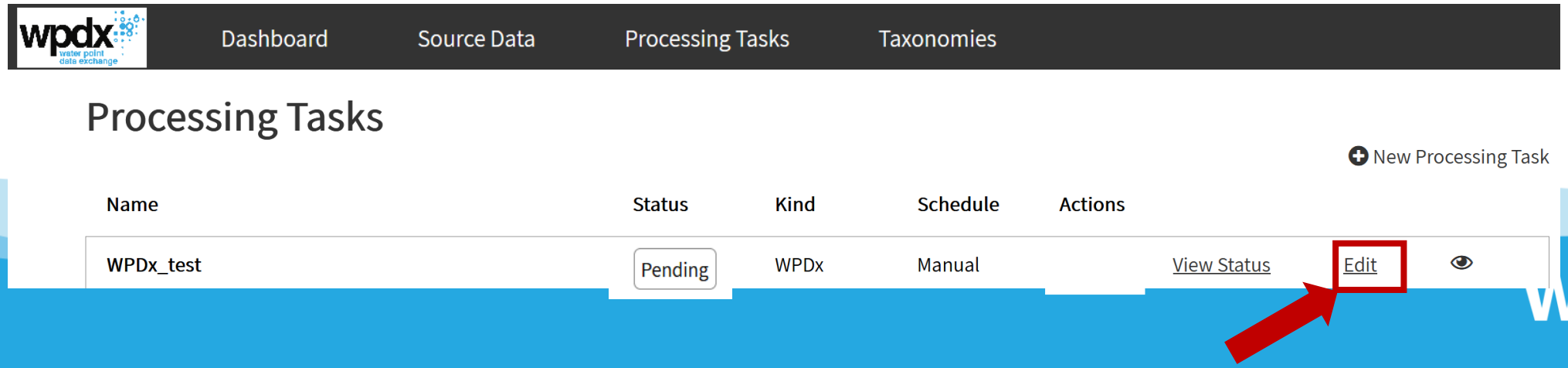
Source Data: Update Contents or Delete

- If you realize you have made an error and need to edit or amend an existing dataset, select ‘Update Contents’ and upload a revised file.
- Once the file has been updated, go back to the associated Processing Task and check/edit the Processing Task content and data mapping and hit “Save and Submit” at the bottom of the Data Import Workbench page.
- Do not use ‘Update Contents’ to initiate a new dataset upload. Instead upload a new file and start a new Processing Task.



Editing a Processing Task

- If you want to add/edit the metadata for your dataset or make changes to the way that the data is mapped to the standard, select “Edit” from the Processing Task tab.
- Make any changes and hit “Save and Submit” at the bottom of the Data Import Workbench page.
- An admin will be alerted of your update and will process the upload.



wpdx
water point
data exchange

Questions?

Please contact
info@waterpointdata.org