WPdx Data Upload Tutorial

For this tutorial, you will need a copy of the WPdx_Upload Tutorial Data.csv file. The steps below will walk you through the upload process including mapping the data to the [WPdx Data Standard](#).

Login to WPdx Upload Engine

1. Navigate to http://upload.waterpointdata.org
2. Login using a Google account. This includes Gmail or [Google Accounts associated with other (i.e., work) email addresses.](#)
3. **For training only,** we will use a training version of the upload engine, which is found here: [https://upload-demo.waterpointdata.org.](#) Data and processing tasks are used only for training and not stored permanently.

Upload data file

1. Select Source Data from top navigation bar and then select Upload Data File from right hand corner of main screen.
2. Browse until you find the file for upload. Make sure that the name helps to identify the information included in the file. The preferred file name format is Organization Name_Geography_DateRange, i.e. Global Water Challenge_Ethiopia_2022.xlsx. For the test data, please name using the format: "Your Org Name_Upload_Tutorial_Data_2020.xlsx"

Start a New Upload Task

1. Select Processing Tasks from the top navigation bar and then select New Processing Task from the top right-hand corner of the main screen.
2. Provide summary information about the dataset. Hints for suggested responses are provided below each text box. For the WPdx Test Dataset, the following answers should be entered:
   a. Task Name: Your Org Name_Ethiopia_2022
   b. Description: This dataset was collected for a baseline inventory of water points in the East and West Regions.
   c. Kind – leave as 'WPDx' (note: only visible in regular upload engine)
   d. Data Source: Your Org Name
   e. Geographical scope: East and West regions of Ethiopia
   f. Types of water sources included: All water points in East and West Regions – improved and unimproved sources
   g. Sampling strategy: Comprehensive census
   h. Frequency of data collection: One time
   i. Point of Contact: Name of appropriate contact
   j. Email address: info@orgname.org (*suggest using organizational level vs. individual email for continuity if personnel changes*)
3. Review and Agree to [data sharing policy](#)
4. Select 'Disable execution of this task' (*this ensures that the test data is not uploaded to the WPdx dataset. This option can also be used when you are setting up an upload but it is not ready for review by the WPdx administrator. When uploading an actual dataset to WPdx, make sure this option is unchecked*)
5. Leave Schedule and Visibility at their default settings.
6. Select Save & Go to Workbench (*you also have option here just to 'Save Task' and come back later to finish*)

Map Your Data to the WPdx Data Standard on the Data Import Workbench

7. Select 'WPdx_Upload_Training_Data.csv' from the Source File dropdown menu.
8. Wait for 'Data to Process' – entries will appear automatically for Direct URL, Format and Sheet. Check to make sure the right Tab from your Excel file is listed under 'Sheet'.
9. Under Data Structure: set Skip Rows to 2 and leave Skip Columns at 0. This will ignore the first two rows of the spreadsheet which includes a title that is not needed for upload.
10. Under Ignored Values: The entry "unknown,Unknown" appears automatically for Empty Values. Entries containing lowercase unknown or uppercase Unknown will appear as blank cells in the WPdx dataset. Based on the training dataset, add "N/A" with no spaces between the comma and the terms.

Data Import Workbench: Column Mapping

Note: Some mapping entries will appear automatically. The test dataset uses commonly used headers, but other files may differ in how columns are named. This is where the ingestion notes Excel file may be useful.

1. For **Latitude (#lat_deg)**, select 'Latitude' from the dropdown menu
2. For **Longitude (#lon_deg)**, select 'Longitude from the dropdown menu
3. For **Presence of Water when Assessed (#status_id)**, select 'Is Water Available?' from the dropdown menu
   a. Click on more settings… to reveal text boxes for "True Values" and "False Values," if they do not appear automatically.
   b. For True Values, enter 'Functional,Partially Functional' into the text box
   c. For False Values, enter 'Not Functional' into the text box. You might also notice "Abandoned" listed as a bad value. We will fix this later in the tutorial.
4. For **Date of Data Inventory (#report_Date),** select 'Date of Visit' from the dropdown menu.
   a. Click on more settings...
   b. For Date Format, select 'Automatic' (may already be selected as default)
5. For **Data Source (#source),** select 'Constant..' from the top of dropdown menu.
   a. Enter "Organization Name" in the Constant Value text box
6. For **Water Source (#water_source),** select 'Source Type' from the top of dropdown menu.
7. For **Water Tech (#water_tech),** select 'Technology Type' from the top of dropdown menu.
8. For **Country (#source),** select 'Country' from the top of dropdown menu.
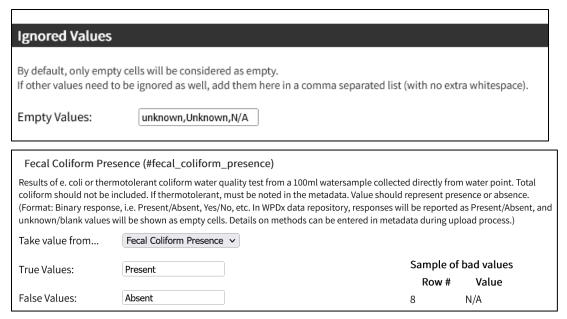
a. This will result in an error which shows up as 'Sample of bad value' as the WPdx Standard required two-letter country codes and not country names (to avoid issues with spelling, language, etc.)

Country (#country_id)

Select the ISO two letter country classification code, selected from a list of all ISO country codes. (Format: Two letter code (e.g., ET)

Take value from...   Country   ⌄

Sample of bad values

Row #	Value
0	Ethiopia

b. To fix the error, select 'Constant..' from the dropdown menu and enter 'ET' in the text box. (Visit this link for a full list of two-letter country codes)

c. Note that if add that if you encounter an error that you cannot fix through the upload engine, it is possible to make the necessary changes directly in the spreadsheet and then update the spreadsheet on the platform. After you have updated the spreadsheet, save your progress on the workbench and navigate to the "Source Data" tab on the Upload Engine. Click "Update Contents" to re-upload the revised file but not lose any progress in the workbench stage. You may need to hit refresh for the changes to be reflected in the processing task window.

9. For **Primary Administrative Division (#adm1),** select 'Region from the dropdown menu.

10. For **Secondary Administrative Division (#adm2),** select 'Zone' from the dropdown menu.

11. For **Tertiary Administrative Division (#adm3),** select 'Woreda' from the dropdown menu.

12. For **Water Point ID (#activity_id),** select 'Unique point ID' from the dropdown menu.
    a. Note that this should be a constant ID associated with the point and not a survey ID.

13. For **Scheme Identification (#scheme_id),** select 'Scheme ID' from the dropdown menu.
    a. Note that this should be a constant ID associated with the water scheme. A scheme is a small water system with several connected points.

14. For **Installation Year (#install_year),** select 'Year of Installation from the dropdown menu.
    a. Click on more settings..
    b. For Date Format, select 'Automatic' (may already be selected as default)

15. For **Installer (#installer),** select 'Installed By' from the dropdown menu.

16. For **Rehabilitation Year (#rehab_year),** select 'Rehab Date' from the dropdown menu.
    a. Click on more settings..
    b. For Date Format, select 'Automatic' (may already be selected as default)

17. For **Rehabilitator (#rehabilitator),** select 'Rehabilitated By' from the dropdown menu.

18. For **Management (#management),** select 'Management Structure' from the dropdown menu.

19. For **Payment (#pay),** select 'Payment' from the dropdown menu.

20. For **Condition (#status),** select 'Status detail' from the dropdown menu.

21. For **Public Data Source URL (#orig_lnk),** select 'Dataset link (fake URLs)' from the dropdown menu

22. For **Photograph (#photo_lnk),** select 'Photo (fake URLs)' from the dropdown menu

23. For **Fecal Coliform Presence (#fecal_coliform_presence),** select 'Fecal Coliform Presence' from the dropdown menu.
   a. For True values check to make sure that "Present" is included in the list.
   b. For False values check to make sure that "Absent" is included in the list.
   c. Note if there is an error message displayed on the right-hand side. This message is indicating that an unacceptable value of N/A is included in the Fecal Coliform Presence column. N/A should have been included in the 'Ignored Values' text box towards the top of the page.
   d. Add N/A to the Ignored Values list.

24. **For Fecal Coliform Value (#fecal_coliform_value),** select 'Fecal Coliform CFU' from the dropdown menu.
25. For **#Subjective Quality (#subjective_quality),** select 'Overall Water Quality' from the dropdown menu.
26. For **Notes (#notes), select 'Seasonality'** from the dropdown menu.
27. For **Converted Fields (#converted)**, leave as 'No Value for this field'. *This parameter is for internal WPdx use only.*
28. For **Water Quality**
   a. For Measurement Unit: Select 'CFU' from the dropdown menu
   b. For Sampling Method: Select 'Idexx quanti-tray' from the dropdown menu
   c. For Coliform Type: Select 'e. coli' from the dropdown menu
   d. For Coliform Incubation Method: Enter 'Incubated at 125 F for 24 hours'
29. For **#notes Metadata**, enter 'Seasonality' in the Notes field contents text box
30. Notice the red error message above the data grid: Field "status_id" can't cast value "Partially Functional" for type "boolean" with format "default"
   a. This means you need to go back and add 'Partially Functional' to the 'True Values' under the #status_id parameter.

b. When you scroll back up to #status_id, you see an error message indicating that 'Abandoned' is also listed as a 'Bad Value.' Add 'Abandoned' under 'False Values' to clear this error.

Presence of Water when Assessed (#status_id)

Description: Identify if any water is available on the day of the visit, recognizing that it may be a limited flow. (Format: Binary response, i.e. Yes/No, Functional/Not-Functional, etc. In WPDx data repository, responses will be reported as Yes/No, and unknown/blank values will be shown as empty cells.)

Take value from...   Is Water Available?   ⌄

Sample of bad values	
Row #	Value
1	Partially Functional
8	Abandoned

True Values: onal,Partially Functional

False Values: t Functional,Abandoned

less settings...

c. Scroll down to the bottom of the screen. You should now see a message that the "Mapping completed successfully, click save to proceed.

d. You can scroll through the 'Original Data', 'Mapped Data' and 'Transformed Data' tabs to see how your data has been mapped to the WPdx Data Standard. The 'Transformed' tab shows the added '_clean' columns which include the standardized terms used by WPdx.

💾 Save   Mapping completed successfully, click save to proceed

Data Sample

✅ Original Data 13 rows   ✅ Mapped Data 13 rows   ✅ Transformed Data 13 rows

#	Country	Latitude	Longitude	Is Water Available?	Date of Visit	Source Type	Technology Type	District	Subdistrict	Unique point ID
1	Ethiopia	11.3494	37.9785	Functional	2020-01-01T14:00:00	Borehole	Afridev handpump	North	Northwest	B123
2	Ethiopia	11.3485	37.9786	Partially Functional	2020-01-01T00:00:00	Protected spring	Tapstand	South	Southwest	PS321
3	Ethiopia	11.3596	37.9752	Not Functional	2020-01-01T00:00:00	Protected spring	Tap stand	South	Southwest	T456
4	Ethiopia	11.2999	37.9632	Functional	2020-01-02T00:00:00	Borehole	Handpump - India Mark II	West	Southwest	B234
5	Ethiopia	11.3412	37.9425	Not Functional	2020-01-02T00:00:00	Shallow well	Handpump	East	Northeast	SW123
6	Ethiopia	11.3598	37.8456	Functional	2022-01-04T00:00:00	Borehole	Mechanized Pump	North	Northeast	B345
7	Ethiopia	11.3698	37.3256	Functional	2020-01-02T00:00:00	Borehole	Tap Stand	North	Northeast	T123
8	Ethiopia	11.2963	37.1259	Functional	null	Borehole	Tap Stand	North	Northeast	T124
9	Ethiopia	11.3544	37.5423	Abandoned	2020-01-03T00:00:00	Well	Handpump	South	Southwest	W123
10	Ethiopia	11.3544	37.5423	Partially Functional	2020-01-03T00:00:00	Protected Hand Dug Well	Afridev handpump	South	Southwest	WX986
11	Ethiopia	11.3544	37.9423	Not Functional	2020-01-03T00:00:00	Unprotected Hand Dug Well	Rope and Bucket	South	Southwest	WH1986
12	Ethiopia	11.3596	37.9752	Functional	2020-01-01T00:00:00	Unrotected spring	null	South	Southwest	TS4456
13	Ethiopia	11.3598	37.9752	Functional	2020-01-05T00:00:00	Surface water	Bucket	North	Southwest	SW1987

31. Click on 'Save and Submit for Approval.' You will then be taken back to the Processing Task screen and see that your task is now listed as 'Pending'. In the regular upload engine, these tasks are reviewed by a WPdx Admin and then approved. You should receive an email notification once the file has been uploaded.

32. Data is available on WPdx-Basic within 24 hours of upload and on WPdx+ (for relevant countries) by the following Monday after upload.

Questions? Please reach out to info@waterpointdata.org