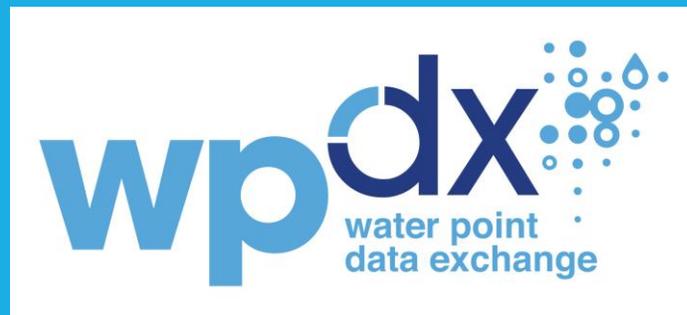


The Water Point Data Exchange Platform



*User
Guide*

This working version of the WPdx User Guide was released in March 2022. Regular updates to the WPdx User Guide are anticipated and will be posted to the [WPdx website](#).

The Water Point Data Exchange (WPdx) is an initiative of the Global Water Challenge.



Generous funding and support to the WPdx platform provided by:



Water Point Data Exchange (WPdx) User Guide

Table of Contents

I. What is the Water Point Data Exchange?.....	1
II. How WPdx Supports Rural Water Decisions	3
III. Background.....	4
IV. WPdx-Basic and WPdx+	8
V. How to Prepare Data for Upload	10
VI. WPdx Data Standard.....	12
VIII. How to Upload Data to WPdx.....	16
VIII. Visual Guide: Uploading Data to WPdx.....	17
V. WPdx Decision Support Tools.....	21
Getting Started	23
View Water Points Tool	24
Administrative Region Analysis Tool.....	25
Rehabilitation Priority Analysis Tool	29
New Construction Priority Analysis.....	37
Data Staleness Analysis.....	40
Functionality Status Prediction Analysis	42
Appendix A. WPdx+ External Datasets.....	43
Appendix B. WPdx Upload Preparation Tips.....	44
Appendix C. Sharing Data from Mobile Applications.....	46

I. What is the Water Point Data Exchange?

The mission of the Water Point Data Exchange (WPdx) platform is to unlock the potential of water point data to improve rural water services through evidence-based decision-making. WPdx combines water point data from governments, NGOs, academic researchers and other stakeholders with population demographics, climate, and conflict datasets to provide key insights. The WPdx platform has three main components:

Share: WPdx Data Standard

The WPdx Data Standard was collaboratively designed for data collection from rural areas at the water point or small water scheme level. The core parameters included in the WPdx Data Standard are parameters which are commonly measured by governments, non-governmental organizations, and researchers to enable easy sharing without changing the types of data typically collected. The WPdx Data Standard is managed and updated on an as-needed basis by a [Global Working Group](#). Please check out the entire [WPdx Data Standard](#).

Access: Global Data Repositories

The WPdx Data Repository is a cloud-based data library that enables sharing of global data that is compliant with the WPdx Data Standard. Data is fully open and free to access. Data is machine readable via an API. The repository includes an online data playground for analysis and visualization. To visit the WPdx Global Data Repository, please see the [WPdx Basic dataset](#). An enhanced subset of the data, [WPdx+](#) is also available. Please visit [this blog post](#) to learn more about two datasets. The entire WPdx+ dataset can be visualized on the [WPdx Decision Support Tools web-app](#).

Use: Decision Support Tools

WPdx currently offers a suite of [decision support tools to improve evidence-based decision making](#). The tools were designed in response to the most frequently asked questions in a survey of government water experts. Check out the [WPdx Decision Support Tools web-app](#). The tools include:

1. **Administrative Region Analysis.** How many people lack basic access to basic water services per administrative region? In which administrative regions should investments be focused?
2. **Rehabilitation Priority Analysis.** Which rehabilitation would reach the most people? Where are people currently without access to basic services due to a broken water point?
3. **New Construction Priority Analysis.** Which new locations would reach the most people? Where are people currently without access to basic services by an existing water point?
4. **Data Staleness Analysis.** How fresh or stale is the data from a specific water point and/or administrative region?
5. **Functionality Status Prediction Analysis.** Which water points are at highest risk of failure? Why? Where should preventative maintenance be focused? (coming soon)

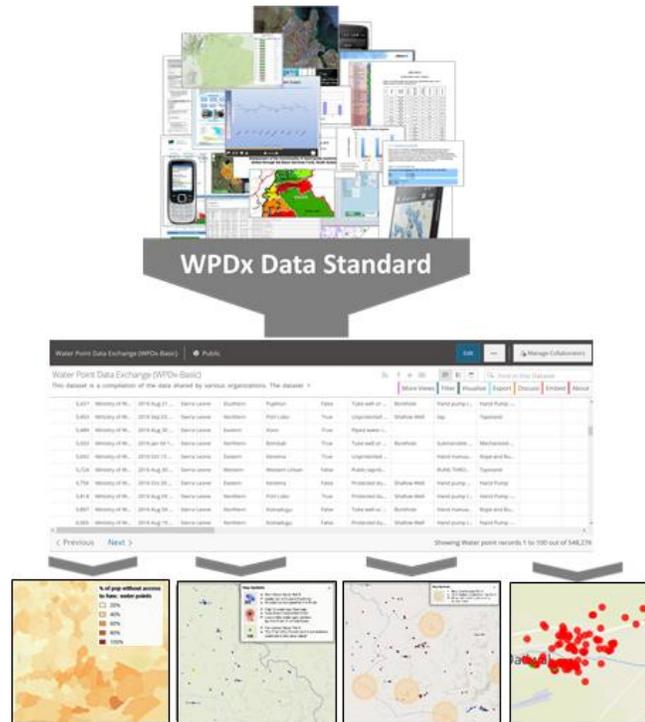


WPdx is a platform to share, access and use water point data to support evidence-based decision-making to optimize resources and maximize impact.

II. How WPdx Supports Rural Water Decisions

1. Governments and their partners are collecting substantial amounts of data. However, data is often too fragmented to use, with different approaches for data collection and storage. Given the challenges in compiling, accessing, and using data, it is often only used for reporting and not decision-making.

2. All stakeholders can easily use the [WPdx Data Standard](#) to harmonize their existing data into the global dataset. Formatting and uploading data to the platform typically [takes less than 30 minutes](#). The WPdx data standard fits most water point data, even if the dataset was collected with no prior knowledge of the standard.



3. Uploaded data is available on our [online data repositories](#). An automated process cleans and categorizes data including the removal of duplicate entries and linking of multiple updates to the same water point through the assignment of a WPdx_id. The platform has two datasets: **WPdx-Basic**, which includes all data shared to the platform and **WPdx+**, a refined analysis-ready dataset.
4. The [WPdx Decision Support Tools web-app](#) accesses data from the WPdx+ dataset and utilizes advanced geospatial and machine learning analyses to provide answers to key questions. Designed in partnership with governments and leading data scientists, these tools provide concrete insights including priority lists of locations for where water point rehabilitations and new construction which would have the greatest impact. Estimates for basic water access are provided for available administrative divisions to help prioritize investments and resources and benchmark progress. A data staleness analysis provides a qualitative metric for the age of available data at the administrative region scale. A status prediction analysis (coming soon) gives insights for which water points have the highest risk of failure which may benefit from preventative maintenance and/or increased monitoring.

III. Background

CHALLENGE

According to the United Nations, at the current pace the world will not meet the Sustainable Development Goal (SDG) on water. This means that in the year 2030, hundreds of millions of people will still be spending hours to collect unsafe drinking water, especially in rural areas where most of the unserved population currently lives. Achieving the SDGs and making sure that nobody is left behind will require dramatic acceleration of water access.

One of the biggest challenges limiting progress to date is the lack of information about rural water services and infrastructure. Historically, detailed global data about rural water points has not been available. Global actors looking to develop evidence-based strategies have had to either rely on national summary estimates, reference small-scale studies, or spend massive amounts of resources to bring together disparate data sets. At a local level, the barriers to using evidence are even greater. Many developing countries have no national inventory of rural water services, and many others have no way to regularly update the data that has been collected. Further, critical data that has been collected is often difficult to access, even by government officials. In cases where data is available, the capacity needed to turn raw data into insights and improved decisions is often limited.

Critical decisions about water services at all levels, from global to local, are being made without enough information. Ultimately, barriers to accessing and using evidence are holding back the achievement of the SDGs.

WPdx SOLUTION

The data needed to accelerate progress on the SDGs already exists. In fact, data about rural water services is being collected more frequently than ever before. Smartphones, tablets, and sensors are generating the data needed to increase the impact of scarce resources, target marginalized communities, and help to accelerate the achievement of SDG 6.

However, these disparate data sources are often fragmented and difficult to harmonize. To address this, Global Water Challenge launched the Water Point Data Exchange (WPdx) in 2015 to bring together diverse rural water data. Governed by an [expert working group](#) and built upon data types that most stakeholders are already collecting, the Water Point Data Exchange is the global framework for sharing, accessing, and using water point data.

DATA EXCHANGE STANDARD

WPdx provides a simple format for sharing data that allows all actors to share information about rural water services. This format is flexible, allowing users to develop their own contextually relevant monitoring framework while still contributing standardized data to the global community. The [WPdx Data Standard](#) allows people to easily share information about location, type, management structure, cost recovery model, water quality, and more for rural water services.

GLOBAL DATA REPOSITORY

In addition to a simple standard, WPdx provides accessible [online global repositories](#) (WPdx-Basic and WPdx+) for rural water point data. These repositories are easy to access, require no login, and present all the data in a simple table. Users can share new data by uploading their data file in a variety of formats or download relevant data in just a few clicks. WPdx automatically joins different updates on the same water point which allows for a dynamic reflection of the reality on the ground, even if different sources provide the information.

By enabling all stakeholders that are collecting water point data to contribute information, WPdx stitches together a more complete and timely understanding of rural water services than any one entity could do alone. In cases where robust national information systems already exist and bring together different data sources, WPdx can simply connect to the national data inventory.

In just over 5 years, WPdx has provided global access to information on over 600,000 water points across 50 countries. This data has been contributed by dozens of stakeholders and continues to grow.

DECISION SUPPORT TOOLS

The WPdx+ dataset provides the information needed to accelerate access to water and ensure that nobody is left behind. However, varying levels of data literacy can limit the potential impact of this data. To increase the use of the data, WPdx brings together governments and data scientists to develop tools that help with routine decisions that are being made all over the world.

From identifying high impact locations for investment to predicting water point failure before it happens, decision-support tools developed by WPdx are changing the way that water services are delivered. WPdx is puts advanced analytics, including machine learning and geographic information systems (GIS), directly into the hands of those who are working to accelerate water access. The suite of tools includes:

- Administrative Region Analysis
- Rehabilitation Priority Administrative Region Analysis
- New Construction Priority Administrative Region Analysis
- Data Staleness Analysis
- Functionality Status Prediction Analysis (coming soon)

SCALING THE USE OF EVIDENCE

WPdx has engaged with national governments in Eswatini, Ethiopia, Liberia, Sierra Leone, and Uganda to support the use of evidence in improving services at national and local levels. This work includes supporting a national framework for data sharing, identifying specific use cases for data, and providing the support needed to turn data into insights. Early achievements include [a policy from the Ministry of Finance in Sierra Leone requiring the use of WPdx decision support tools](#) in annual planning processes and the development of [the first national data sharing standard in Liberia](#). In the coming years, WPdx aims to continue to provide tailored support to increase the use of evidence and accelerate the progress towards the SDGs in additional countries around the world.

Recognizing the large role international NGOs play in collaborating with governments to deliver water services and shape strategies, WPdx is collaborating with the world's largest water NGOs to help provide access to the data they have collected and integrate evidence-based decision making into their own programs. WPdx is committed to building a community of water sector leaders that advance the use of data to achieve SDG 6.

SUPPORTING GOOD GOVERNANCE

The USAID Water Security, Sanitation, and Hygiene Governance Technical Brief ([2020](#)) defines “good governance” as when institutional roles and responsibilities are clear, people are trained and empowered to fulfill their roles, entities have incentives to do what needs to be done, regulations exist and are enforced, and there are consequences for poor performance. The brief goes on to define the goal of improving WASH governance is to ensure all people have access to the high-quality services and resources they need, and that they know who to turn to if there are issues. USAID states that better WASH governance contributes to increased trust in the government’s ability to manage—and be accountable for—critical and highly visible basic services and resources, with positive ripple effects across other development priorities.

The USAID Rural Water Service Technical Brief ([2020](#)) emphasizes that the WASH sector recognizes the need to view water supply as an ongoing service that needs to be effectively managed over the life of the physical infrastructure. This emerging emphasis on a service delivery approach recognizes the importance of wider systems of governance and the enabling environment, political economy, life cycle costs, and the role of local institutions.

The Sanitation and Water for All (SWA) 2021 Summary Progress Update on SDG6 states that “credible and timely water and sanitation data provide numerous social, economic, and environmental benefits in both public and private sectors, such as stronger political accountability and commitment, as well as public and private investments. [High quality data] also enables evidence-based policy-making, regulations, planning and investments at all levels, to ensure the most effective deployment of resources.” However, SWA identifies specific data challenges including gaps in data collection and sharing and suggests that improved data generation, validation, standardization, and information exchange can build trust so leaders can make informed decisions and increase accountability. These ideas are also expressed in the U.S. Global Water Strategy which identifies the lack of data for decision-making as a limiting factor in service provision sustainability and includes a strategic approach to promote common data exchange formats and access to data for decision-making to help support reaching Strategy objectives.

WPdx supports good governance by providing a platform which allows for transparency and accountability through the consolidation, sharing and analysis of regularly collected functionality data from a variety of actors across a target geography. This information, along with the decision-support tools can be used to benchmark performance and identify areas with and without access to sustainable services. The insights gained through effective data collection, analysis and use contribute to an improved understanding for decision-makers to better prioritize limited resources and ensure that the most vulnerable populations are not left behind.

Why Share Data with WPdx?

WPdx is an efficient way to strengthen local and national systems by providing a common platform for data sharing and analysis. Sharing data through WPdx provides significant benefits to both the data provider and the decision-makers. These benefits include:

- **Improve governance through increased data availability and contribute to evidence-based decisions**
 - WPdx decision support tools allow governments and other organizations to make informed choices on water sector budgets and priority locations for preventative maintenance, rehabilitation and construction.
- **Benchmark progress**
 - WPdx offers a standardized framework for sharing data allowing similar information to be directly compared and benchmarked across different geographies.
- **Demonstrate a commitment to data transparency**
 - The WPdx platform provides free and open access to global water point data for all users.
- **Increase impact of data collection efforts**
 - The harmonized WPdx dataset provides a more comprehensive view of the local water sector than any one organization has on their own.

IV. WPdx-Basic and WPdx+

The WPdx platform hosts two datasets: [WPdx-Basic](#) and [WPdx+](#). The WPdx+ dataset is focused on a subset of countries for which WPdx has enough data either nationally or for specific administrative regions for the decision support tools to be activated. The tools and dataset enhancements can be made available for any country if a representative dataset can be shared with WPdx. The current country list includes: Bangladesh, Central African Republic, Eswatini, Ethiopia, Ghana, Haiti, India, Kenya, Liberia, Malawi, Nigeria, Sierra Leone, Tanzania, Uganda, and Zimbabwe.

The WPdx+ dataset is the input for the new suite of decision-support tools which have been released in beta form and are available on the [WPdx Decision Support Tools web-app](#). Additional updates to the remaining tools will be released in the coming months.

For more information on how to add a new country to WPdx+, email info@waterpointdata.org with “New Country Interest” in the subject line.

Please see below for a brief summary of the two datasets:

WPdx-Basic	WPdx+
<ul style="list-style-type: none"> • All data shared to WPdx is included in WPdx-Basic Global Data Repository. • There are five validation/data-cleaning steps which occur during the ingestion process: <ul style="list-style-type: none"> • Ensure that all records contain the required parameters. Records that do not contain required parameters are not uploaded. A summary of records included in the upload can be found in the WPdx Data Catalog. • Check to ensure that points are located within a country boundary per GADM boundaries. Points which do not fall within country boundaries (i.e., in the middle of the ocean) are not uploaded. A summary of records included in the upload can be found in the WPdx Data Catalog. • Formatting of entries for consistency. For example, for the Presence of Water When Assessed (#status_id) parameter, the repository will shown as “Yes” or “No”. 	<ul style="list-style-type: none"> • Focused dataset on countries for which WPdx has enough data for decision-support tools to be activated. • Further data processing steps including: • Removal of records identified as having location mismatches (i.e., data provided states that record is from Country X, but GPS location is in Country Y) • De-duplication for any records mistakenly uploaded twice (exact matches only). • Removal of records classified as ‘Unimproved’, ‘No Facilities’ or ‘Unknown’ based on facility_type categorization. • Assignment of a WPdx_id which matches water point records shared by different organizations and on different dates, based on GPS location. • Addition of external relevant data sources which are used in the water point status predictions models, including (for full details see Appendix A):

<u>WPdx-Basic</u>	<u>WPdx+</u>
<ul style="list-style-type: none"> • Addition of ‘clean’ version of country name, #adm1, #adm2, #adm3 based on provided GPS coordinates and GADM boundaries. ‘Clean’ values are appended to the record, leaving all original data intact. • Addition of “water_source_clean”, “water_tech_clean”, “management_clean” and “status_clean” columns. These new columns are created using fuzzy matching to organize entries into consistent categories. For more information on the cleaning process, please see here. • Categorization of a facility_type into improved and unimproved based on JMP definitions. 	<ul style="list-style-type: none"> • Distance between water point and nearest road (primary, secondary and tertiary), town and city using OpenStreetMap data. • Population density and demographic information from Facebook. • Tabular access to results from advance decision support tools: <ul style="list-style-type: none"> ○ Rehabilitation Priority ○ Data Staleness ○ Status Predictions (coming soon) • These results and additional decision-support data can also be downloaded directly from the WPdx Decision Support Tools web-app: <ul style="list-style-type: none"> ○ Adm Region Analysis ○ New Construction Priority

V. How to Prepare Data for Upload

For additional tips on how to prepare your data for upload, please see Appendix B.

STEP #1: VERIFY THAT YOUR DATA IS **COMPLETE**

- WPdx requires a minimum of vital information before you upload your data to the platform:
 1. **LOCATION OF THE WATER POINT:** Provide the decimal value of the latitude and longitude in WGS 1984. At least four decimals should be included, but more are welcome. Do not round or truncate the values.
 2. **PRESENCE OF WATER WHEN ASSESSED:** Identify if any water is available on the day of the visit, recognizing that it may be a limited flow.
 3. **DATE OF DATA INVENTORY:** Provide the date when the data was collected.
 4. **DATA SOURCE:** Name of organization sharing data.
 5. **WATER SOURCE:** Describe the source where the water comes from **AND/OR**
 6. **WATER POINT TECHNOLOGY:** Describe the system being used to transport the water from the source to the point of collection.

STEP #2: VERIFY **PERMISSIONS** FOR SHARING DATA

- To provide unpublished data to the WPdx repository, the Global Water Challenge requires each data provider to agree to the WPdx Provider Agreement shown below.

WPdx Provider Agreement

As the provider of data that is not currently in the public domain, and for the purposes of sharing through the Water Point Data Exchange (WPdx), please confirm the following:

- You have secured all rights necessary to authorize sharing of this data and to determine how it is shared in the future.
- You are licensing this data to the Water Point Data Exchange under the Creative Commons Attribution 4.0 license, available at <https://creativecommons.org/licenses/by/4.0/>.
- WPdx reserves the right to change or modify the provider agreement at our sole discretion at any time. We will notify you about significant changes in the provider agreement by sending a notice to this email address so that you can choose whether to continue sharing data with WPdx. Significant changes will go into effect no less than 30 days after we notify you.
- Any modifications to the provider agreement will not change the license referenced above unless written permission is given by the provider.
- Your email address will be the main avenue for informing you of any updates to the provider agreement.

STEP #3: COMPARE YOUR DATA WITH THE **WPdx STANDARD**

- Compare your dataset with the WPdx standard prior to data upload. There is no longer a need to rename columns, but it is important to ensure that formats align and to take note of which columns from your data match to a specific standard parameter. The WPdx Data Mapping Table in Appendix B may be a useful tool for this step.

STEP #4: REMOVE ALL PERSONALLY IDENTIFYING INFORMATION

- Make sure to remove all personal information (i.e. address, name, phone number, etc.) from your data before saving and uploading the file.

STEP #5: SAVE THE DOCUMENT OR ONLINE DATASET

- The WPdx Ingestion Engine can accept a variety of file formats including: .xls, .xlsx, .csv, .json, Google Sheets, and DropBox. For direct uploads, save your file in an easy to access location. For connections to online data repositories, please make note of the appropriate URL.
- For data stored in another online platform, please ensure the URL includes permissions for public access. See Appendix C for further information.

VI. WPdx Data Standard

ALL OF THESE FIELDS ARE REQUIRED

FIELD	HASHTAG	DESCRIPTION	FORMAT
LATITUDE	#lat_deg	Provide the decimal value of the latitude in WGS 1984. At least four decimals (but more are encouraged) should be included. North and East should be noted as positive numbers.	Numbers only. <i>Example: 44.475821.</i>
LONGITUDE	#lon_deg	Provide the decimal value of the longitude in WGS 1984. At least four decimals (but more are encouraged) should be included. North and East should be noted as positive numbers.	Numbers only. <i>Example: 73.211954.</i>
PRESENCE OF WATER WHEN ASSESSED	#status_id	Identify if any water is available on the day of the visit, recognizing that it may be a limited flow.	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.
DATE OF DATA INVENTORY	#report_date	Provide the date that the data was collected on using <u>ISO 8601</u> . Time and time zone designator are optional.	YYYY-MM-DD (2015-08-23) or HH:MM:SS+Timezone offset (2015-11 04T13:51:00-04)
DATA SOURCE	#source	Provide the name of the organization collecting the data record.	Open text. <i>Example: WPdx</i>

AT LEAST ONE OF THESE FIELDS ARE REQUIRED

FIELD	HASHTAG	DESCRIPTION	FORMAT
WATER SOURCE	#water_source	Describe the water source (e.g. shallow well, spring, borehole, river, pond, etc.).	Open text.
WATER POINT TECHNOLOGY	#water_tech	Describe the system being used to transport the water from the source to the point of collection (e.g. Handpump (include manufacturer, i.e. Afridev, India Mark II, Malda, etc.), Kiosk, Tapstand, etc.).	Open text.

THESE FIELDS ARE OPTIONAL

FIELD	HASHTAG	DESCRIPTION	FORMAT
COUNTRY	#country_id	Select the ISO two letter country classification code .	Two letters. <i>Example:</i> LR
PRIMARY ADMIN. DIVISION	#adm1	Provide the name of the tertiary administrative division. See GADM.org .	Open text. <i>Example:</i> Amhara
SECONDARY ADMIN. DIVISION	#adm2	Provide the name of the tertiary administrative division. See GADM.org .	Open text. <i>Example:</i> Debub Gondar
TERTIARY ADMIN. DIVISION	#adm3	Provide the name of the tertiary administrative division. See GADM.org .	Open text. <i>Example:</i> Dera
WATER POINT ID	#activity_id	Provide the Unique ID for the water point infrastructure. This could be a physical ID on the water point or an internal system ID.	Open text. <i>Example:</i> Borehole46812_ZM
SCHEME ID	#scheme_id	Provide the unique identifier for a small, piped scheme that connects multiple water points.	Open text. <i>Example:</i> SchemeA34
INSTALLATION YEAR	#install_year	Provide the 4-digit install year.	Four number year. <i>Example:</i> 1995
INSTALLER	#installer	Provide the name of the entities that installed the water system. This should be the entities that completed or were directly responsible for the construction, rather than a donor or other involved stakeholder.	Open Text, with multiple entities separated with a “,”. <i>Example:</i> WPdx, Global Water Challenge
REHAB YEAR	#rehab_year	Provide the 4-digit year of most recent major rehabilitation (not just maintenance).	Four number year. <i>Example:</i> 1995
REHABILITATOR	#rehabilitator	Provide the name of the entities that completed the most recent rehabilitation of the water system. This should be the entities that complete or were directly responsible for the construction, rather than a donor or other involved stakeholder.	Open Text, with multiple entities separated with a “,”. <i>Example:</i> WPdx, Global Water Challenge

FIELD	HASHTAG	DESCRIPTION	FORMAT
MANAGEMENT	#management	Select the classification of the entity that directly manages the water point.	Open text, suggest: “Direct Government Operation”, “Private Operator/Delegated Management”, “Community Management”, “Institutional Management”, “Other”
PAYMENT FOR WATER	#pay	Provide the payment amount and basis. If no amount is provided, the basis can be provided alone. An amount without a payment basis cannot be included.	Open Text. If an amount is included, it should include the currency using <u>ISO 4217 code</u> . <i>Example:</i> monthly, per jerry can, when broken, etc.
CONDITION	#status	Provide a status of the physical/mechanical condition of the water point.	Open text. <i>Example:</i> The water point was leaking.
PUBLIC DATA SOURCE URL	#orig_lnk	Provide the public link to the data record for a specific water point or full data set, including any non-standard compliant data.	URL of dataset or record, starting with http://
PHOTOGRAPH	#photo_lnk	Provide the URL of a photograph of the water system. Multiple URLs can be included, with each URL separated by a semi-colon (;).	URL of Photograph, starting with “http://”
FECAL COLIFORM PRESENCE	#fecal_coliform_presence	Results of e. coli or thermotolerant coliform water quality test from a 100ml water sample collected directly from water point. Total coliform should not be included. If thermotolerant, must be noted in the metadata. Value should represent presence or absence.	Binary response, i.e. Present/Absent, Yes/No. In WPdx data repository, responses reported as Present/Absent, and unknown values shown as empty cells. Details on methods can be entered in metadata during upload process.
FECAL COLIFORM VALUE	#fecal_coliform_value	Results of e. coli or thermotolerant coliform water quality test from a 100ml water sample collected directly from water point. Total coliform should not be included. If thermotolerant, must be noted in the metadata. Value should represent the most probable number or colony forming units in 100ml.	Must be a floating-point value. Details on methods should be entered in metadata.

FIELD	HASHTAG	DESCRIPTION	FORMAT
SUBJECTIVE QUALITY	#subjective_ quality	Information regarding the perceived quality of the water including taste, appearance, and/or odor.	Open text. Example: Water was clear, but it tasted like soil.
NOTES	#notes	This field can be used to incorporate any additional information not already part of the WPdx standard that is useful to the data provider.	Open text.

VIII. How to Upload Data to WPdx

STEP #1: **LOGIN** TO YOUR WPdx UPLOAD ACCOUNT

- To upload your data to the WPdx platform, please visit:
 - <https://upload.waterpointdata.org/> to login with a Gmail account.

STEP #2: **UPLOAD** YOUR DATA TO WPdx

- After reviewing data for the required parameters and [verifying that you have permissions to share the data publicly](#), you are ready to upload data.

To upload your data to the WPdx platform, please follow these steps:

- Go to <https://upload.waterpointdata.org>
- Click on the "**Login to the System**" option located in the blue button.
- **Enter your Gmail address.** (Gmail address is required for login)
- On the top menu, click on the option that says "**Source Data**" (if you are using an online dataset, skip to Step#3).
- Select **+ Upload Data File**.
- **Select** your data file.

STEP #3: **MAP** YOUR DATA TO THE WPdx Data Standard

- On the top menu, click on the option that says "**Processing Tasks**"
- Select **+ New Processing Task**
- Provide **task name, basic description of dataset** and **point of contact**
- **Review and Agree** to [data sharing terms](#).
- Select **Save & go to Workbench**
- Select **source file from dropdown** or **enter direct URL**
- Map your data to WPdx standard parameters using the dropdown menus
- Finally, click "**Save and Submit for Approval**"
- Once an administrator has reviewed and approved your upload you will be notified via email.

Your data has been uploaded to WPdx.

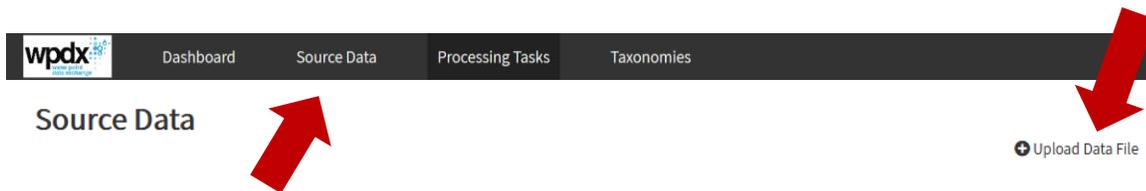
VIII. Visual Guide: Uploading Data to WPdx

This visual guide provides a graphical view of uploading data to the WPdx platform. For an even more detailed step-by-step guide, please visit the [WPdx website How To Guide](#).

1. Open your internet browser and go to <https://upload.waterpointdata.org>.
2. Click on the **"Login to System"** option located on blue button and enter a **Gmail account address** to access the system.



3. Click on **"Source Data"** and then **"+ Upload Data File"** (skip this step if using data from online database). Your file will appear in the Source Data list and be available for your use.



4. Click on **"Processing Task"** and then **"+ New Processing Task"**



5. Enter a task name, basic description of the dataset and contact details on the metadata page. This information will be name will be publicly visible on the [WPdx Data Catalog](#). Please use the following naming convention: “OrgName_Geographic focus_Year of Data Included” (for example: “Global Water Challenge_Sierra Leone_Spring2020”). Agree to terms and click “Save & Go to Workbench”

Edit Task

Task Name

Please provide a name for the processing task using the following format, Organization Name_Country/Region_Month/Year of data included. e.g., Global Water Challenge_Ethiopia_2018

Description

Describe the main purpose for the collected data; examples may include: routine monitoring, project evaluation, regional inventory, national WASH inventory, water quality testing, etc.

Kind

About the Data

Data Source

Name of the organization that owns the data. This entry will be applied as the `Source` parameter for all uploaded records. If your dataset includes data from multiple data sources, please map the `Source` parameter to the appropriate column in the Data Workbench.

Geographical scope

or ‘sampling frame’ of data collection (e.g. particular region, whole country etc.)

Types of water sources included

e.g. improved water sources only, handpumps only, water systems of a certain age, functional water points only, etc.

Sampling strategy

e.g. comprehensive census, simple random sampling, stratified random sampling, cluster random sampling, purposive sampling, convenience sampling, quota sampling etc.

Frequency of data collection

e.g., annually, monthly, one-time, etc.

Point of Contact

Point-of-contact information to be displayed on the dataset page
We recommend using organizational level contact information and not personal.

Full Name

Email Address

Agree to data sharing [terms](#)

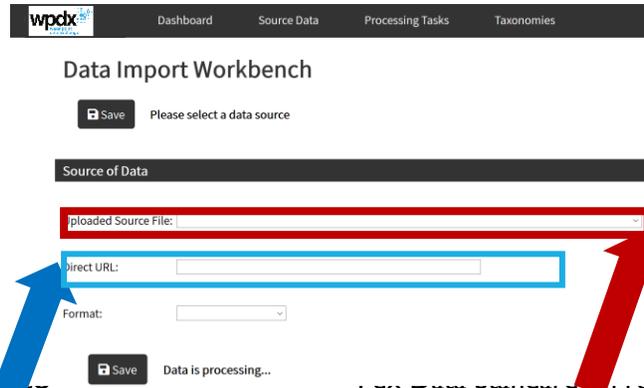
Schedule

Visibility

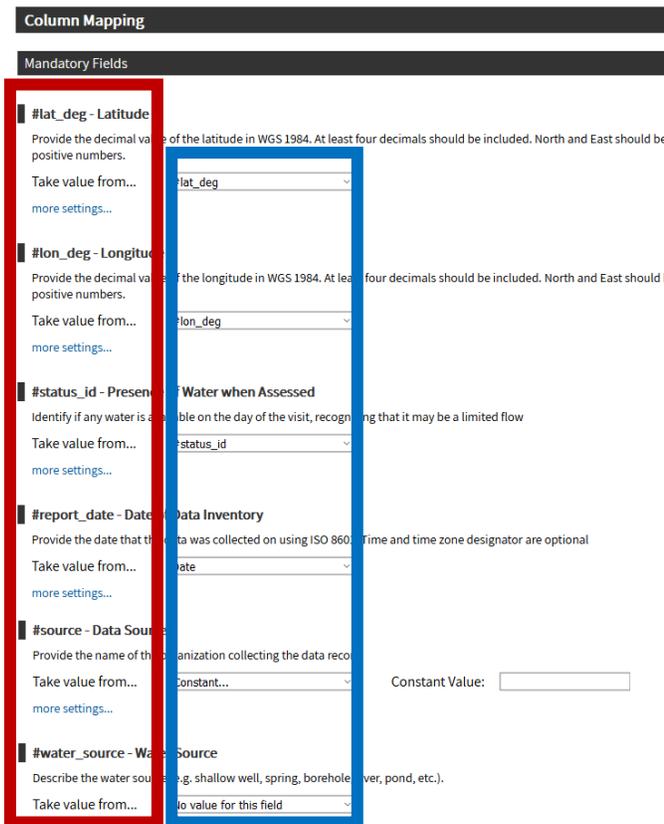
 Save & Go to Workbench

 Save Task

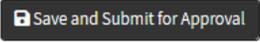
6. In the workbench, select your **source file** from dropdown or enter a **Direct URL** (if using an online database).



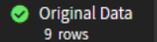
7. Map your data (using the dropdown menus) using the dropdown menus.

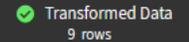


7. Review your data and when ready, click on **Save & Submit for Approval**. On average data will be approved and available online within 24 hours. Once data has been uploaded, a summary of the upload can be found by selecting the 'eye' icon to access its WPdx Data Catalog entry.



 Mapping completed successfully, click to submit data for approval

Data Sample





#	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


 Dashboard Source Data Processing Tasks Taxonomies

Processing Tasks



Name	Status	Kind	Schedule	Actions
WPDx_test	Success	WPDx	Manual	View Status Edit 

9. On the [WPdx Data Catalog](#), a summary of the metadata and overview of the number of records uploaded and explanations for why any records were considered invalid.

Statistics

- i. **Total rows processed:** total number of rows included in file
- ii. **Missing Lat/Lon/Report Date:** number of rows not uploaded due to missing required parameters
- iii. **Exif Fixed Coordinates:** GPS from photographs are used in place of provided GPS coordinates when available
- iv. **Out of Range Latitude of Longitude:** number of rows where the provided GPS coords are outside of global country boundaries.
- v. **Missing water tech/source:** number of rows not uploaded due to missing required parameters
- vi. **Validation errors:** number of entries which do not meet formatting requirements. These entries will appear as blanks in the dataset, but the row will still be uploaded as long as the required parameters are included.
- vii. **Valid Rows:** total number of rows uploaded
- viii. **Dropped Bad Coords:** number of rows dropped because the provided GPS coords do not exist in any country boundaries.
- ix. **Dropped No coords:** number of rows dropped because GPS coords are missing

V. WPdx Decision Support Tools

The data available through the Water Point Data Exchange provides a powerful foundation for advanced analytics. By harmonizing the data in WPdx and employing geospatial information systems and machine learning, WPdx has created powerful decision-support tools to help governments and their partners make data-based decisions on a routine basis. These tools were designed to be simple and user-friendly.

You can learn more about the tools by visiting <https://www.waterpointdata.org/use-data> or access them directly at <https://tools.waterpointdata.org>**

** Please note that the tools are in beta release and all results should be validated by local stakeholders. Feedback is always welcome. Please provide inputs via [this form](#) or by emailing info@waterpointdata.org.

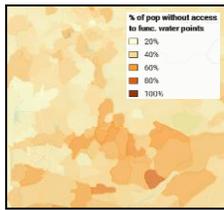
CONSIDERATIONS and LIMITATIONS

All WPdx decision-support analyses are based on the latest available data from the [WPdx+ dataset](#). WPdx is an open-source repository for data and contributors include governments, NGOs, academic researchers, and others. WPdx has not directly collected this data or verified the accuracy of any of the data on the platform. WPdx does perform basic validation checks to ensure that required parameters are included and that GPS locations match with the country boundaries of the provided datasets. Additional cleaning and categorizing steps are taken to prepare the dataset for analysis. Full details of these processes can be found [on our website](#). Questions and feedback are more than welcome.

Population calculations utilize the [Facebook High Resolution Population Datasets](#). Administrative boundaries are taken from the [GADM dataset](#) for all countries except [Eswatini](#), [Ghana](#), and [Uganda](#) which utilize the [OCHA Common Operational Datasets](#) available on HDX. Populations in urban areas, as defined by the [EU Global Human Settlement Database](#) are removed from the total administrative region population to provide an estimate of the rural population. While included on the maps for visualization purposes, packaged and delivered water facilities are not included in the analyses to estimate rural populations with access to basic services.

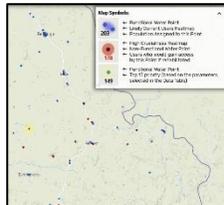
Any water points that have been broken, rehabilitated, or constructed since data was uploaded are not included in this analysis. Additionally, household connections are not captured in WPdx. While large urban areas are removed from the analysis if they reach a specified population density, smaller piped schemes with household connections may be missed. As a result, peri-urban areas may be identified as optimal locations for rehabilitation or new construction, even if those communities are already served by household connections.

WPdx DECISION SUPPORT TOOLS



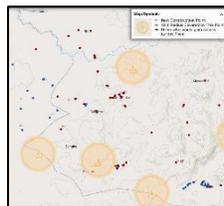
ADMINISTRATIVE REGION ANALYSIS TOOL

This tool provides an estimate on rural basic water access, providing the percent of the rural population in each administrative region living near a functional water point. **Example use case: Understanding which districts present the greatest need for rural water access.**



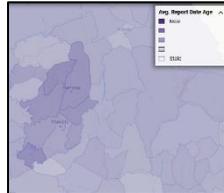
REHABILITATION PRIORITY ANALYSIS TOOL

This tool evaluates all non-functional water points in an administrative region and determines the rural population which could regain basic access if the water point was repaired. **Example use case: Prioritizing regional maintenance and repair plans for district or sub-district.**



NEW CONSTRUCTION PRIORITY ANALYSIS TOOL

This tool examines all possible locations where a water point could be constructed in an administrative region, taking into consideration existing functional and non-functional water points. The tool identifies the rural population which could gain basic access for each proposed location. **Example use case: Initial siting layouts for potential new water points.**



DATA STALENESS ANALYSIS TOOL

This tool provides a qualitative overview of the average age of data available from the WPdx dataset. **Example use case: Identifying areas in need of updated data sharing and/or data collection.**

FUNCTIONALITY STATUS PREDICTION TOOL (coming soon)

This tool harnesses the power of machine learning to make predictions about the status of water points based on the past performance of similar water points in the country and relevant demographic, population, conflict and environmental data which may impact water point functionality. **Example use case: Prioritizing preventative maintenance schedules for a district and/or service provider.**

Getting Started

The entire WPdx+ dataset and the complete suite of decision support tools can now be accessed in this single web-based app. Users can scroll through the entire available dataset or filter to focus on a specific geographic area of interest.

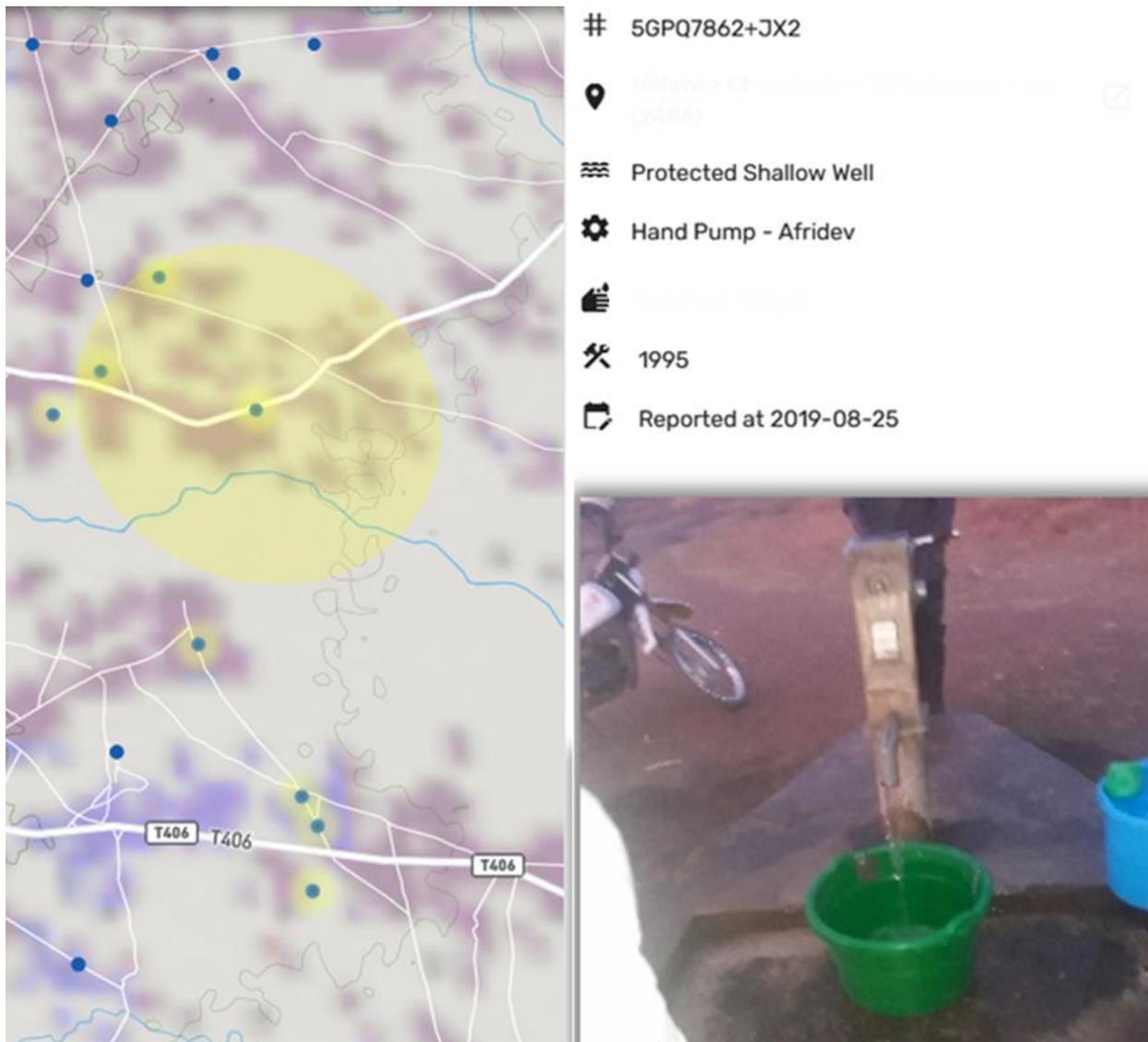
To get started:

1. Navigate to <https://tools.waterpointdata.org>. The landing page will display the entire WPdx+ dataset on an interactive map.
2. Select area of interest by clicking on the gear icon  in the lower left corner and selecting 'Filter by Region.' Select country and administrative divisions of interest from dropdown menus.
3. Select an analysis from the Decision Support Tool dropdown menu in the header bar.
4. Explore results from country level down to the most detailed administrative division available (typically Adm2 – Adm4 depending on the country).
5. Click on individual water points or administrative regions to find more detailed information.
6. For the Rehabilitation Priority Analysis and New Construction Analysis, click on the table icon  to view an overview of the top 15 water points (additional details below).
7. If desired, download the results by clicking on the gear icon in the lower left corner, selecting Download Data and choosing the file of interest.
8. To view data contributors for the selected view, click on the data-sharing icon  in the lower left corner.

View Water Points Tool

Select **View Water Points** from the Decision Support Tool dropdown box in the header bar. Functional and non-functional water points are included on the map.

Click on an individual water point to open a pop-up box with an overview of key information from the point's WPdx record. The map will show the 1km radius around that water point to highlight nearby water points.



Example of pop-up box and 1km radius marker which appears when you select an individual water point.

Administrative Region Analysis Tool

Overview

The Administrative Region Analysis Tool provides an overview of the rural population with access to basic services, without access to basic services and uncharted (i.e., data is not available in WPdx to determine access for these populations). for each available administrative level.

- **Rural Population with Basic Access:** Population within 1km of a functional water point
- **Rural Population without Basic Access:** Population within 1km of a non-functional water point (but not within 1km of a functional water point)
- **Uncharted Rural Population:** Population for which no data on water services is available in WPdx. These populations may be without basic access or basic services may exist, but data has not been shared with WPdx.

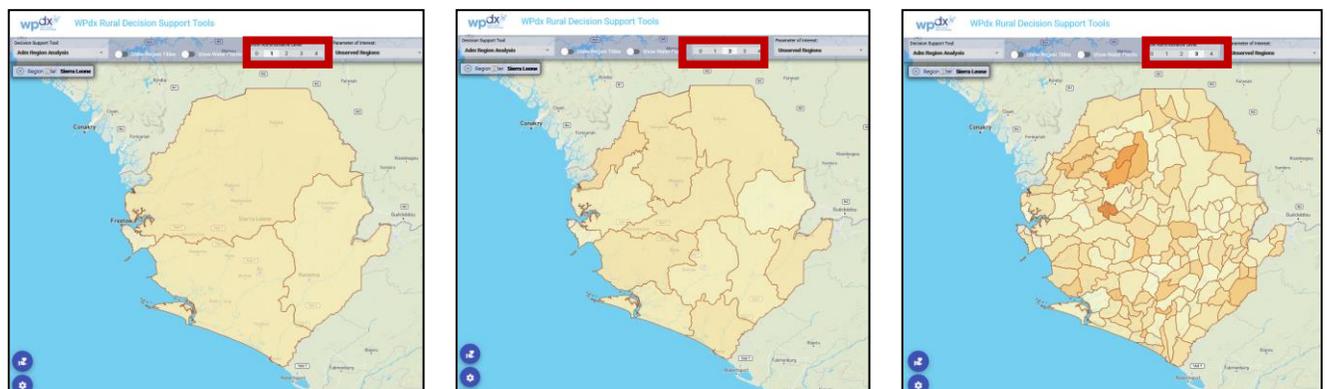
Users can view choropleth maps, which provide administrative regional analysis for the percentage of the rural population With Basic Access, Without Basic Access and which are Uncharted by selecting an option from the 'Parameter of Interest' dropdown menu on the header bar and by selecting their preferred administrative zone (Adm0, Adm1, Adm2, Adm3, Adm4) from the top navigation bar. Please note that different countries have varying numbers of administrative divisions.

Illustrative Uses

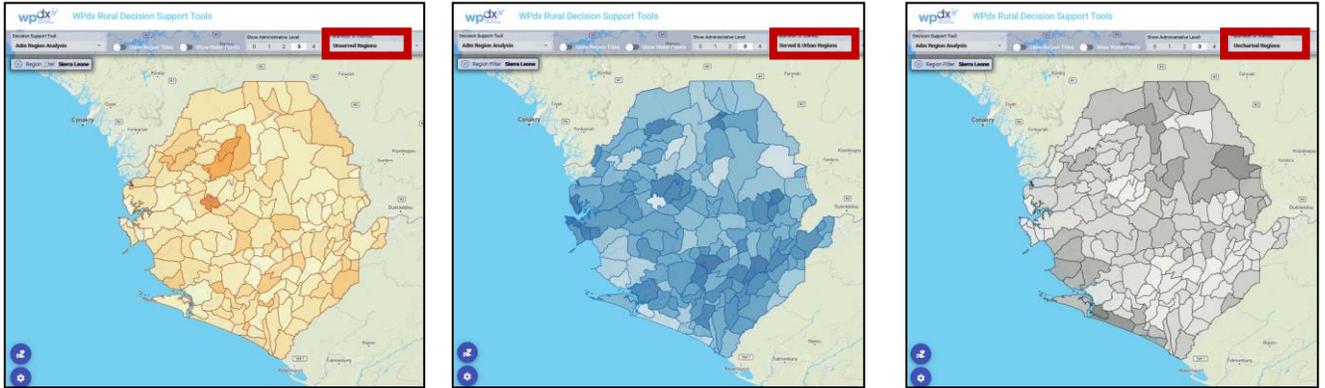
- Prioritizing administrative divisions for budget and resource allocations
- Identifying target administrative divisions for interventions
- Evaluating equity

Examples

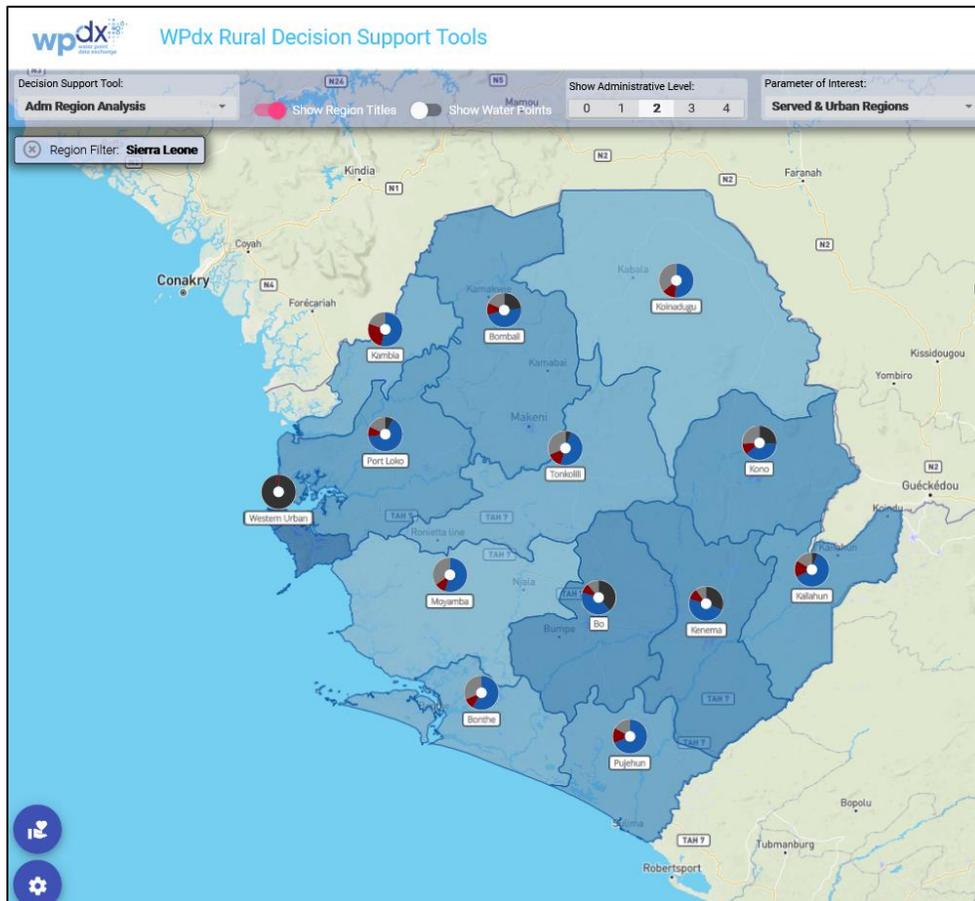
1. Explore data at different administrative divisions



2. Compare populations with basic access, populations without basic access and uncharted populations for the same administrative divisions.



3. View summary charts showing population percentages for rural populations with basic access, rural populations without basic access, and rural uncharted population per administrative district.



Instructions

1. Select area of interest by clicking on the gear icon  in the lower left corner and selecting 'Filter by Region.' Select country and administrative divisions of interest from dropdown menus.



2. Select **Administrative Region Analysis** from the Decision-Support Tool dropdown menu in the header bar.
3. Select the *administrative division level* (Adm0, Adm1, Adm2, Adm3, Adm4) from the header bar.
4. Select *Rural Pop with Basic Access, Rural Pop without Basic Access or Rural Pop Uncharted* from the *Parameter of Interest* dropdown menu in the header bar.
5. Options include:
 - a. Toggle *Coverage Summary* in the header bar to see/hide summary pie charts showing the percentage of the rural population with basic access, without basic access and which is uncharted for the administrative division of interest.
 - b. Toggle *Region Titles* to see/hide names of the administrative divisions.
 - c. Toggle *Show Water Points* to see/hide functional and non-functional water points.
6. Click on an administrative area or *Coverage Summary* pie chart to open a pop-up box which includes an overview of the rural population with basic access, without basic access and which is uncharted and the total functional and non-functional water points for each nested administrative division.
7. To view the data sources included in the filtered view, select the data sharing icon  in the lower left corner.
8. If desired, data can be downloaded as an XLSX file by clicking on the gear icon  in the lower left corner and selecting *Download Data* and then *Adm Region Analysis*.

Methodology

For each administrative division, the population is determined using the [Facebook High Resolution Population Data](#) and administrative boundaries from [GADM](#) or HDX ([Eswatini](#), [Ghana](#) and [Uganda](#)). The population from urban areas, based on the definitions provided by [EU Global Human Settlement Database](#) is then removed, leaving an estimate of the total rural population for that administrative region.

Using data from WPdx+, the total population within a 1km radius of functional water points is identified (local_access_pop). The local_access_pop is removed from the population raster, and then the total population within a 1km radius of non-functional water points is determined (local_no_access_pop). Next, the amount of population assigned to functional water points who exceed the maximum capacity of those water points is determined (overcap_pop). The overcap_pop is subtracted from the local_access_pop to find the estimated Rural Population with Basic Access. The overcap_pop is added to the local_no_access_pop to find the estimated Rural Population without Basic Access. The remaining

population which are more than 1 km from an existing water point are designated as the Uncharted Rural Population. Uncharted could mean that populations in these areas lack basic access or that there is not data in WPdx+ about the water services for populations in these areas.

The following maximum capacity values (extended from [Sphere Guidelines](#) and [Yu et al. \(2017\)](#)) are currently in use:

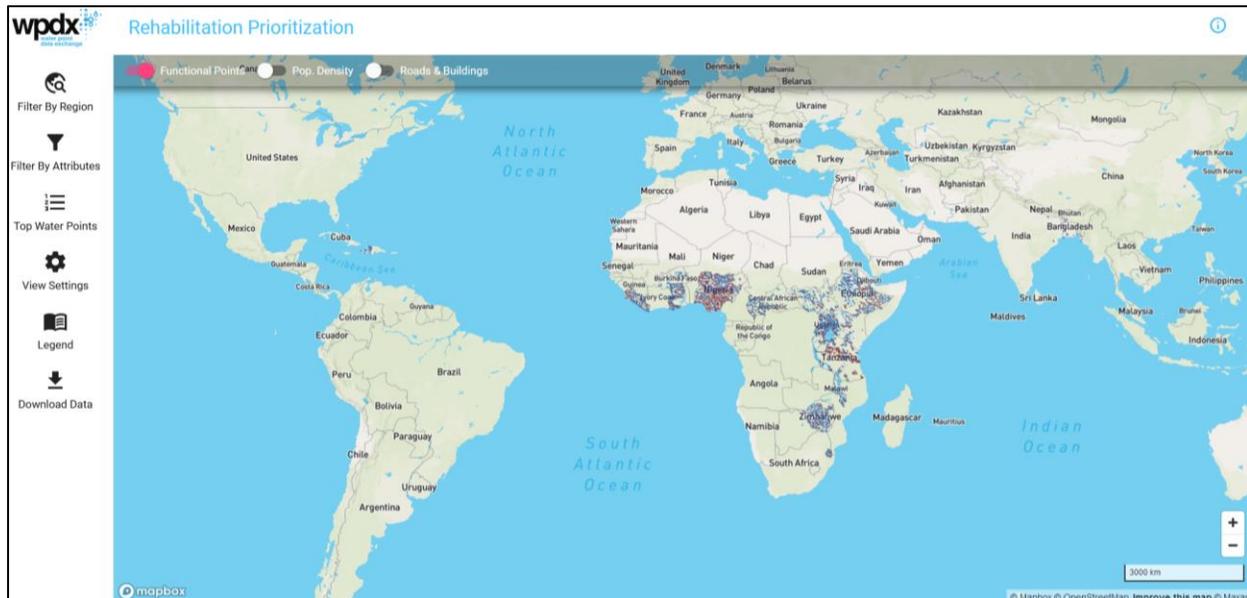
- 250 people per tap [tapstand, kiosk, rainwater catchment]
- 300 people per protected spring
- 400 people per open hand well [rope and bucket]
- 500 people per hand pump [all hand pumps]
- 1000 people per mechanized (powered) well

We recognize that these are not perfect figures and that different countries/regions/districts define maximum users in different ways. We also recognize that the number of users may vary during the year due to seasonality, pastoralism, conflict/displacement, etc. Our aim is to provide two consistent 'best estimates' of potential water point users.:

- Local pop – total population within 1 km of the water point
- Water point pop – total population divided up to specific water points to avoid double counting (population is assigned based on a relative distance calculation if population centers have more than one water point within a 1 km distance)

In the future, we hope to explore approaches to customize these max capacities based on different national/regional/district guidelines as well as changes related to seasonality.

Rehabilitation Priority Analysis Tool



Overview

The **Rehabilitation Priority Tool** provides recommendations for which non-functional water points should be considered for rehabilitation and repair. The tool also provides insights on which water points are critical in that there are limited nearby alternatives and which water points are being overutilized. Results can be viewed and filtered based on:

- Potential population that would regain access if point was repaired (default)
- Total population within 1km of water point
- Crucialness of water point (i.e., are there alternative water points nearby)
- Pressure on the water point (i.e., is the water point over or under-utilized)

Illustrative Uses

- **Prioritizing** which water points to rehabilitate
- **Highlighting** areas where there are limited alternative water points available
- **Understanding** which water points are over- or under-utilized
- **Benchmarking** rehabilitation needs to inform district budgets and workplans

Notes

The Rehabilitation Priority tool excludes delivered and packaged water services from the analysis.

Examples

1. Prioritize which water points to rehabilitate

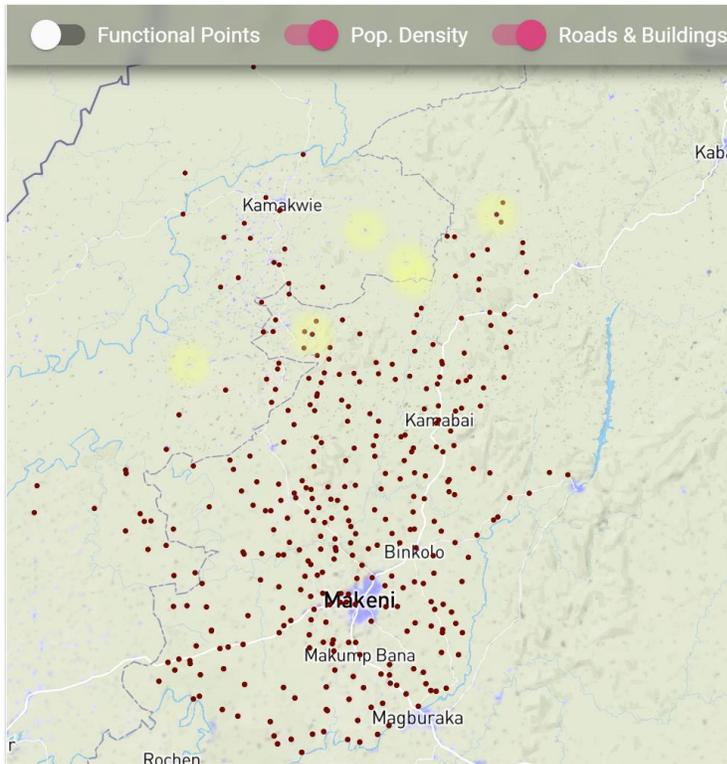
The 'Top Water Points' Table summarizes which currently non-functional water points would reach the greatest number of people if they were to be repaired. This list can be reviewed and verified by district officials.

The table shows four relevant results. The WPdx Rehabilitation Priority Recommendation is made based on the 'Water Point Pop' column. Full results can be downloaded as a CSV by selecting the 'Download Data' button. The list can be re-ordered based on each of these parameters depending on user needs.

- **Local Pop.** – total number of people living within 1 km of water point.
- **Water Point Pop.** – number of people estimated who use (for functional points) or who would use (for non-functional points) based on availability of other working water points.
- **Crucialness** – calculated as the ratio of 'water point pop' over 'local' pop to provide a metric of the local population's reliance on a specific water point.
- **Pressure** – calculated as the water point pop/max recommended capacity usage to provide a metric of over or under utilization of the water point.

#	Functional?	Source	Tech	Local Pop.	Served Pop.	Crucialness	Pressure
1	Yes	Borehole	Rope and Bucket	1,568	1,568	100%	313.6%
2	No	Undefined Shallow Well	Unknown	1,425	1,425	100%	285%
3	Unknown	Undefined Shallow Well	Unknown	1,425	1,425	100%	285%
4	No	Undefined Shallow Well	Unknown	1,425	1,425	100%	285%
5	Yes	Undefined Shallow Well	Unknown	1,328	1,328	100%	265.6%
6	Yes	Protected Shallow Well	Hand Pump - India Mark II	1,114	1,114	100%	222.8%
7	Yes	Protected Shallow Well	Hand Pump - India Mark II	1,061	1,061	100%	212.2%
8	No	Protected Shallow Well	Hand Pump - India Mark II	787	787	100%	157.4%
9	No	Protected Shallow Well	Hand Pump - India Mark II	1,568	772	49.2%	154.4%
10	No	Protected Shallow Well	Hand Pump - India Mark II	760	760	100%	152%
11	Unknown	Borehole	Unknown	760	760	100%	152%
12	Unknown	Undefined Shallow Well	Unknown	746	746	100%	149.2%
13	No	Undefined Shallow Well	Unknown	743	743	100%	148.6%
14	No	Protected Shallow Well	Hand Pump	743	743	100%	148.6%
15	Yes	Protected Shallow Well	Hand Pump - India Mark II	713	713	100%	142.6%

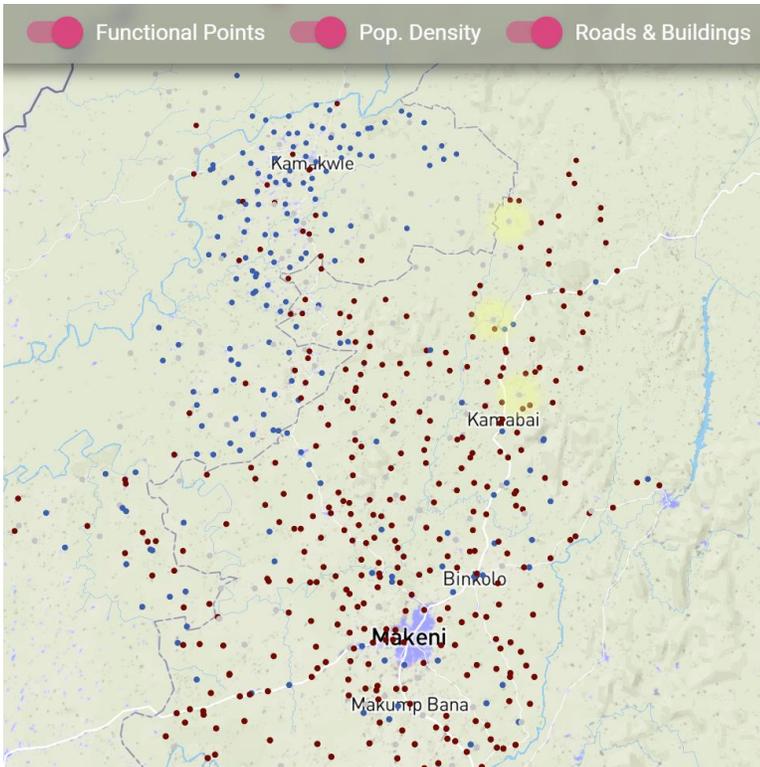
Example Top Water Points table for identifying rehabilitation priorities



The highlighted points in the northern part of the district show that rehabilitating these points would reach more people currently without basic access compared to other potential rehabilitations.

2. Highlight areas where there are limited alternative water point available

By ranking the 'Top Water Points' table by Crucialness, users can explore which water points are the only working water point within 1km. A crucialness score of 100% means that the population appears to be completely reliant on that water point and indicates there are no alternative available if this water point were to breakdown. To build resilience, planners may want to examine these points as inputs to rehabilitation and new construction priorities. This score is calculated for both functional and non-functional water points.



The map highlights the top 15 water points with the highest crucialness scores.

This screenshot shows a detailed view of a specific water point. On the left is a satellite map with a green circular overlay indicating a 1km radius around the water point. On the right is a metadata panel with the following information:

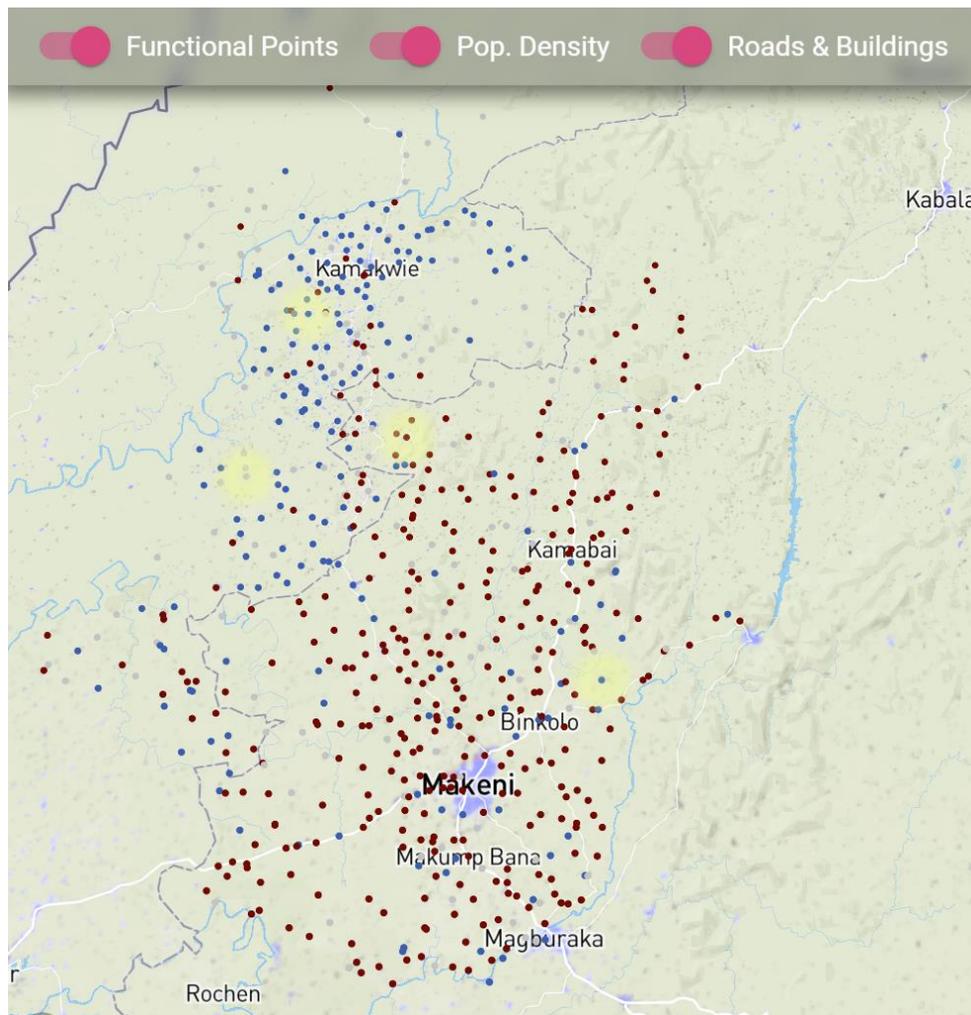
- # 6CW99GX8+7WV
- Sierra Leone > Northern > Tonkolili > Yoni
- Undefined Shallow Well
- Hand Pump
- Population in 1km radius: 321 ppl
- Crucialness: 100%
- Pressure: 64%
- Likely Current Users: 321 ppl
- Usage Capacity for this Point: 500 ppl
- UNICEF
- Community Management
- 10
- Reported at 2019-03-11

Below the metadata panel is a photograph of the water point, which is a hand pump structure with a concrete base and a metal frame. A person is visible in the background near the pump.

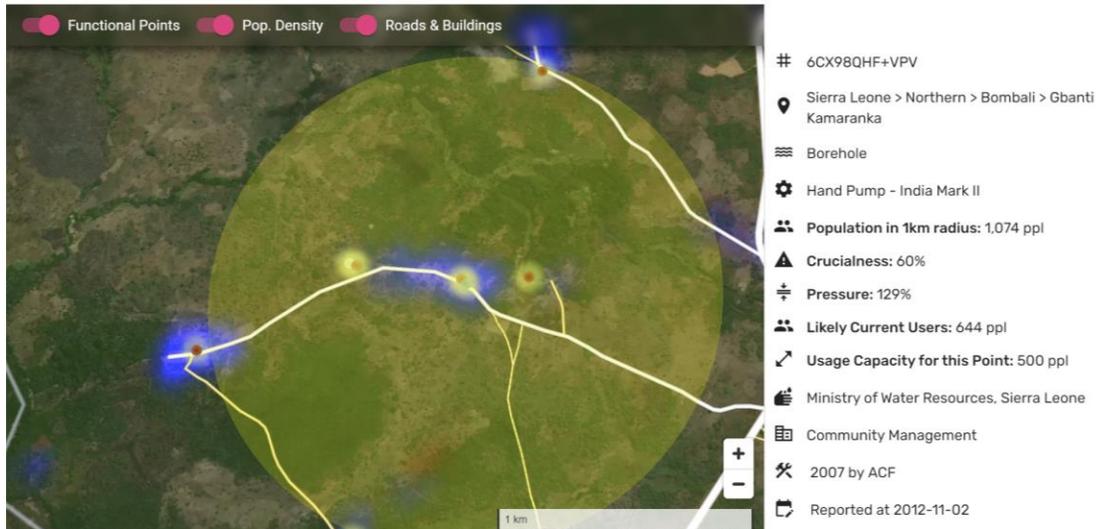
Example of a functional water point which has a 100% crucialness score, meaning there are no other working water points within 1km. If the non-functional water point was repaired, the crucialness score of each point would be approx. 50%

3. Understand which water points are over- or under-utilized

By ordering the 'Top Water Points' table by 'Pressure', users can view which points have the highest 'Pressure' score, including those which may be over 100%. Points with pressure scores near or over 100% indicate the water point is at or above maximum capacity for the water point technology. The top 15 water points appears on the map with a yellow halo, and full results can be further analyzed by downloading the table.



Overview of water points with highest pressure score (with yellow halo) for an area of interest



This water point has a pressure score of 129% showing that it is over capacity, likely because all nearby water points are non-functional, suggesting that the population is over-utilizing the water point.

4. Benchmark rehabilitation needs to inform district budgets and workplans

The results from the analyses described above can be synthesized with local knowledge to develop a list of priority water points for repair and rehabilitation.

Instructions

1. Select area of interest by clicking on the gear icon in the lower left corner  and selecting 'Filter by Region.' Select country and administrative divisions from dropdown menus.



2. Select **Rehabilitation Priority Analysis** from the Decision-Support Tool dropdown menu.
3. If desired, **Filter by Attributes** through the dropdown menu available by clicking on the gear icon . Filter by Attributes allows users to view specific water point sources, technologies, or management types. Note that the analysis still considers all water points, not just those included in a filtered view.
4. Click on the table icon  to view the **Top Water Points** shows the top fifteen water points for rehabilitation. You can choose to sort this table by four parameters: *total local population within 1km of water point*, *water point population*, *crucialness* or *pressure* (see definitions below for more detail). The top 15 points will be shown on the map with a yellow halo and can be zoomed to by selecting the row in the table.
5. Users can customize the results shown in the table and the map visualizations. Options include:
 - a. **Show functional water points:** If turned on, both functional and non-functional water points will be visible on map and included in the Top Water Points table. To focus only on non-functional points, toggle this option to be off.
 - b. **Show Population Density:** Population density can be toggled on/off to view where population clusters exist with respect to existing water points. Note: the app currently displays the population from WorldPop, though Facebook data is used in the analyses.
 - c. **Show Roads & Buildings:** The roads and buildings layer from OpenStreetMap can be toggled on/off for context on surroundings.
6. For additional settings, click on the gear icon  and select **View Settings:**
 - a. **Show points in urban areas:** Most data in WPdx is for rural areas. Turning this option on will display the points located in urban areas, however, points in urban areas are not included in the decision-support analyses.
 - b. **Show Water Point Pop figure:** This option displays the number of people who could gain access (non-functional points) or those with current access (functional points) on the map for easier comparisons.
7. The results from the analysis can be downloaded by clicking on the gear icon  and selecting **Download Data**. Select "Rehab Priority Analysis." The ranked priority column in the download file is based on the Water Point Pop. column for non-functional points only. This is the number of users who would gain access if the water point was returned to a functional state.

Methodology

This tool uses the latest available data from the [WPdx+ dataset](#) combined with a population estimate layer. Please note that water points with an unknown status are analyzed as if the status is non-functional.

The tool calculates the **local population**, which is the number of people living within a 1km radius of each water point.

The tool also calculates the **water point population** for each water point. The water point population is determined based on the relative distance between the population and water points within a 1km radius. Water points which are closer to the population are assigned a higher proportion of the population compared to water points which are farther away. For functional water points, this figure is referred to as **Likely current users**, or the number of people who currently have access to a specific working water service. For non-functional water points, this figure represents the **Users who would gain access** to a water service if the point were to be rehabilitated.

The **crucialness score** (0-100%) is the ratio of potential users to the total local population within a 1km radius of the water point. Crucialness provides a measure of water system redundancy. For example, if there is only 1 water point within a 1km radius, the water point crucialness score is 100%, meaning that there are no nearby alternatives. If there are two functional water points within 1km, the crucialness score for each point will be ~50% indicating there is some redundancy in the system, so if one water point is broken down, users have an alternative water point available. For non-functional water points, the crucialness score shows how important the water point would be if it were to be rehabilitated. [See example here.](#)

The **pressure score** (0-100%) is calculated based on the ratio of the number of people assigned to that water point over the theoretical maximum population which can be served based on the technology. If a point is serving less than the recommended maximum, the pressure score will be less than 100% (i.e., $250/500 = 0.5$). If a point is serving more than the recommended maximum, the pressure score will be over 100% (i.e., $750/500 = 150\%$). The following recommended maximum values (extended from [Sphere Guidelines](#) and [Yu et al. \(2017\)](#)) are currently in use. Additional work may be done to refine and improve these estimates and where possible take into account local regulations:

- 100 people per rainwater catchment system
- 250 people per tap [tapstand, kiosk, rainwater catchment]
- 300 people per protected spring
- 400 people per hand pump [all hand pumps]
- 1000 people per mechanized (powered) well

New Construction Priority Analysis

Overview

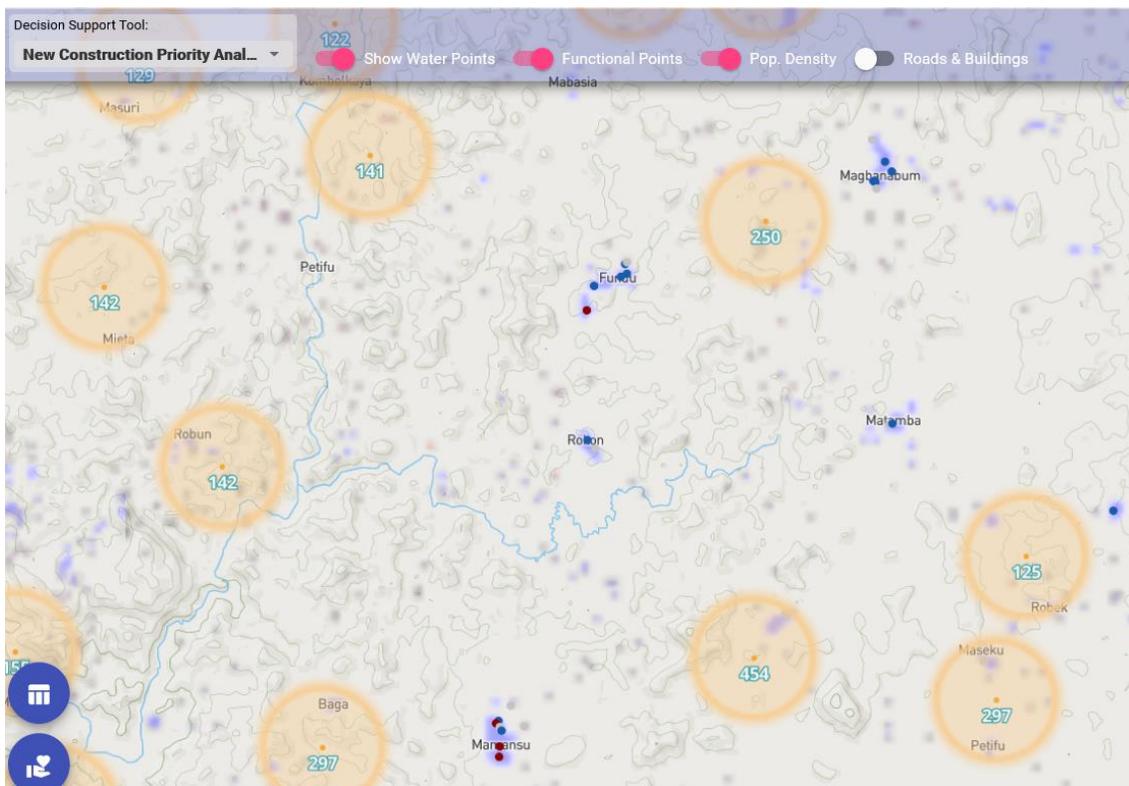
The **New Construction Priority Analysis Tool** evaluates all possible locations where a water point could be constructed in a given administrative area and evaluates how many people that are not near an existing water point (regardless of functionality) could gain access if a water point was constructed in that location.

Illustrative Uses

- Identify locations to construct new water points
- Evaluate the relative benefit of new construction compared to rehabilitating existing water points
- Provide insights on potential data gaps which could be filled by uploading data to WPdx

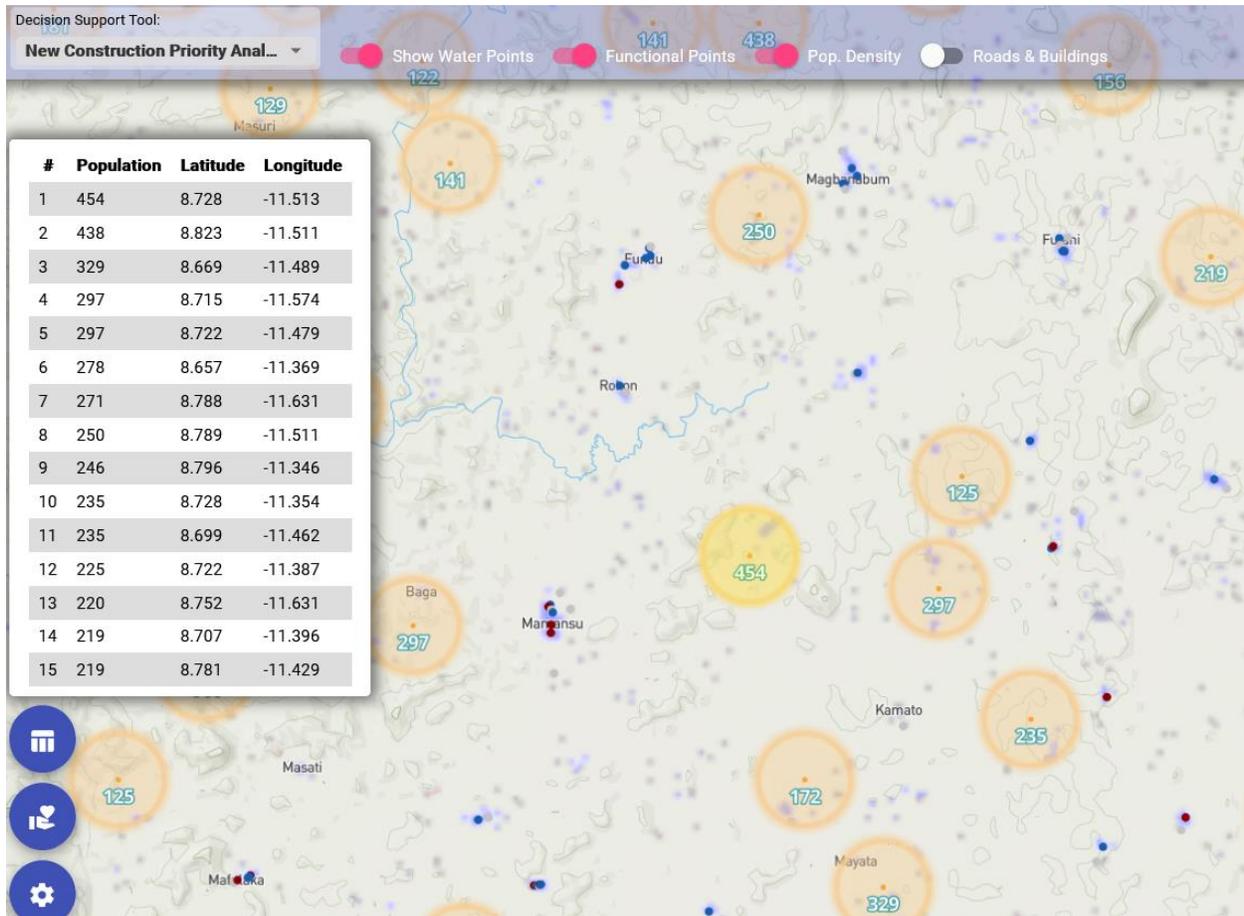
Examples

- 1. Explore potential locations for new construction relative to existing functional and non-functional water points.*



Orange circles show locations for consideration for new water point construction. Figures inside orange circles are the number of people living with a 1km radius of the center point. Blue dots are functional existing water points and red dots are non-functional existing water points

2. Sort list of potential new construction locations by population within 1km of proposed locations.



Click on the table icon  to see an overview of the GPS coordinates and local population within 1km of each proposed new construction location.

Instructions

1. Select area of interest by clicking on the gear icon in the lower left corner  and selecting 'Filter by Region.' Select country and administrative divisions from dropdown menus.



2. Select **New Construction Priority Analysis** from the Decision-Support Tool dropdown menu.
3. Click on the table icon  in the lower left corner to view the Top Water Points table which shows the GPS coordinates and local population within 1km for the top fifteen proposed points for new construction based on potential population reached.
4. The results from the analysis can be downloaded by clicking on the gear icon  and selecting **Download Data**. Select *New Construction Analysis*. The downloaded file includes the country, administrative divisions, latitude, longitude and population living within 1km for the proposed point.

Methodology

For each administrative region, we inspect the 'uncharted' areas to find concentrations of population. Uncharted areas are those in which the population is outside of a 1km radius of an existing (functional or non-functional) water point in the WPdx+ dataset.

- First, we find the point location that has the most uncharted population in a 1km radius and save its coordinates and population size.
- Then we repeat the process, while ignoring population that was already "found" in previous iterations.
- We end the process when we can't find any point in the administrative region that has more than 100 people in less than 1km distance from it.

Data Staleness Analysis

Overview

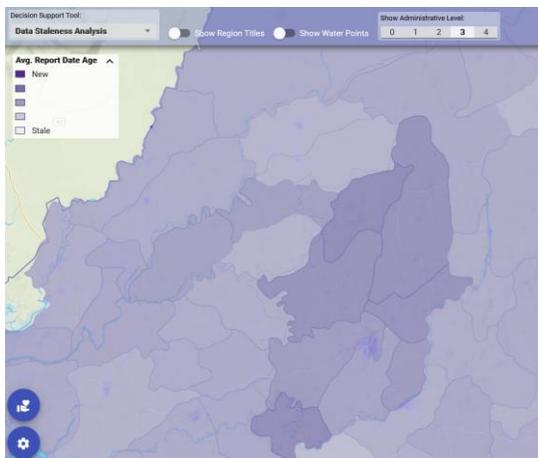
The Data Staleness Analysis provides a relative measure of the average age of data available from the WPdx+ dataset.

Illustrative Uses

- Identifying areas for targeted data sharing outreach
- Selecting areas for focused data collection
- Ensuring a clear understanding of the age of data available for other analyses

Examples

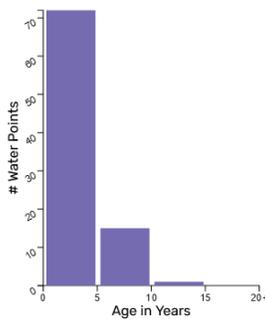
1. **Identify areas with fresher/staler data encourage data sharing and/or identify areas for data collection.**



2. **Click on a region to see a pop-up with a histogram on the age of available data for that administrative division.**

📏 Average Report Age: 2.8 years

Water Point Report Age Distribution



Instructions

1. Select area of interest by clicking on the gear icon in the lower left corner  and selecting 'Filter by Region.' Select country and administrative divisions from dropdown menus.



2. Select **Data Staleness Analysis** from the Decision-Support Tool dropdown menu.
3. Click on a region to see a pop-up with a histogram on the age of available data for that administrative division.

Methodology

For each point, we calculate a 'staleness' score which depends solely on its report date.

The score represents the depreciation of the point's relevance - older updates are less valuable than fresh data.

The main parameter that controls this score is the 'target value' - for example, in the current state we define that a 20-year-old report is worth only 10% in comparison to a 'day-old' report. All other reports are then given values based on their report age using a geometric decay model.

These value assignments are qualitative and are only meant to give a rough estimate of data 'freshness'.

Functionality Status Prediction Analysis

RELEASE EXPECTED 2022

OVERVIEW

The **Functional Status Prediction Analysis Tool**

harnesses the power of machine learning to make predictions about the status of water points based on the past performance of similar water points in the country.

ILLUSTRATIVE USES

- **Predict** which water points are at a higher risk of failure in order to carry out preventative maintenance.
- **Identify** high-risk water points in order to increase monitoring where it is most needed.
- **Determine** which districts have relatively more high-risk water points to more effectively match maintenance budgets with likely need.

LIMITATIONS

Like all predictions, these predictions are based on probabilities and may not reflect the reality of the status of water points at a given point in time.

Appendix A. WPdx+ External Datasets

With appreciation, we utilize the following data sources within the WPdx Decision Support Tools web-app:

Administrative Boundaries

- [GADM](#)
- HDX: [Eswatini](#), [Ghana](#), and [Uganda](#)

Population data

- Facebook Connectivity Lab and Center for International Earth Science Information Network - CIESIN - Columbia University. 2016. High Resolution Settlement Layer (HRSL). Source imagery for HRSL © 2016 DigitalGlobe. Accessed DAY MONTH YEAR." Available online at: [HDX: Facebook High Resolution Population Data](#)

Roads and buildings

- [Open Street Map](#)

Urban areas (defined by)

- [EU Global Human Settlement Database](#)

Water point data

- [Water Point Data Exchange Plus \(WPdx+\)](#)

Appendix B. WPdx Upload Preparation Tips

The WPdx ingestion engine cannot map a single column to multiple parameters in the WPdx data standard. If you would like to map a single column to more than one parameter, please duplicate the column in your dataset. For example, your “functionality” column may include information that is relevant both to #status_id and #status parameters. In the example below you could map “Functional Status” to #status_id and “Functional Status_copy” to #status during the upload process.

FIND AND REPLACE TOOL: This tool in Excel (other spreadsheet programs have similar functionality) allows you to make numerous changes in a few clicks. WPdx does have data cleaning processes in place to make entries more consistent, but it is always helpful to minimize the number of acronyms and short-hand terms used to ensure your entries are easy to understand, especially if you are using any non-standard acronyms. For example, when possible, “India Mark II” and “Hand-Dug Well” are preferred over “IM2” and “HDW”.

CONCATENATE COLUMN FUNCTION: This function in Excel (other spreadsheet programs have similar functionality) allows you to merge two (or more) columns into a single column. This process may be helpful to create singular columns for #payment or #subjective_quality.

Upload Preparation: Data Mapping Table		
WPdx Standard	Dataset Column Match	Upload Notes
Latitude (#lat_deg)		Must be in decimal degrees
Longitude (#lon_deg)		Must be in decimal degrees
Presence of Water when Assessed (#status_id)		Default is Yes/No. Can enter custom terms during ingestion process.
Date of Data Inventory (#report_date)		
Data Source (#source)		If not included in dataset, enter as Constant during upload.
Water Source (#water_source)		
Water Point Technology (#water_tech)		
Country (#country_id)		If not included in dataset, enter as Constant during upload.
Primary Admin. Division (#adm1)		
Secondary Admin. Division (#adm2)		
Tertiary Admin. Division (#adm3)		
Water Point ID (#activity_id)		
Scheme Identification (#scheme_id)		
Installation Year (#install_year)		
Installer (#installer)		
Rehabilitation Year (#rehab_year)		
Rehabilitator (#rehabilitator)		
Management Structure (#management)		
Payment for Water (#pay)		
Condition (#status)		
Public Data Source URL (#orig_ink)		
Photograph (#photo_ink)		
Fecal Coliform – Presence or Absence (#fecal_coliform_presence)		Add information about methods used during upload process.
Fecal Coliform – Value (#fecal_coliform_value)		Add information about methods used during upload process.
Subjective Quality (#subjective_quality)		May need to concatenate columns
Notes (#notes)		Add details on information included for #notes during upload process

*Bold parameters are required. At least one of #water source or water_tech must be included.

**Table available as Excel download at: https://www.waterpointdata.org/wp-content/uploads/2020/11/Ingestion-Notes_for-website.xlsx

Appendix C. Sharing Data from Mobile Applications

If you are currently using a mobile application to collect water point data, such as Akvo Flow or mWater, uploading to WPdx is a simple process! Below is a brief guide to how to connect from commonly used platforms. As a reminder, **please be sure to remove any personally identifying information before uploading data.**

Please note that if you have trouble with the automatic connections, you can always upload the XLSX, XLS or CSV version of the file directly.

- For an xlsx or csv file in **Dropbox**, copy the download link and not the sharing link to use in “Direct URL” box in the workbench during upload.
- For **Google Sheets**, ensure that the document is shared publicly.
- For **custom data platforms**, please contact us to determine how we can best connect.

For **Akvo Flow**:

1. Determine which columns from your dataset match with the parameters in the WPdx data standard.
2. Contact your Akvo account manager with a request for a WPdx direct link to an Akvo Flow form, specifying which Akvo Flow form you want to export.
3. The account manager will provide a URL with an API endpoint.
4. Copy/paste the URL of the API endpoint in the Direct URL field (Source of Data section) in your WPdx ingestion engine Processing Task.
5. Map the column headers from your dataset to the WPdx data standard using the dropdown menus.
6. Uploads can be scheduled on a monthly, quarterly, etc. basis. Please provide a recommended update schedule in the Description section of the Processing Task. (one time, monthly, quarterly, annually, etc)

For **mWater**:

Create a datagrid with your data formatted to the WPdx data standard.

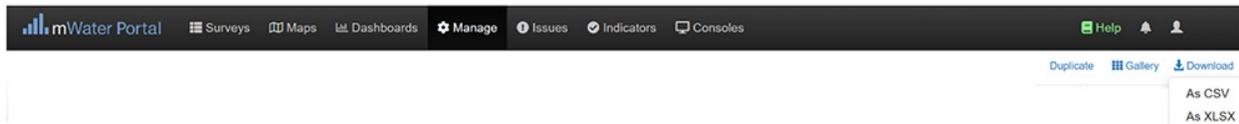
Click on “Download” and select “As XLSX.” Browse to find the download link, which is a permanent URL which can be used to upload and update your data into WPdx. Each browser handles download links slightly differently. Please see below for options using Firefox, Google Chrome, and Microsoft Edge.

A few helpful tips for setting up your datagrid:

(1) Consider naming the datagrid columns to match the field codes used by WPdx. This is not necessary (you can map individual column names to the WPdx standard parameters later in the process) but may help to make things more straightforward during the mapping process.

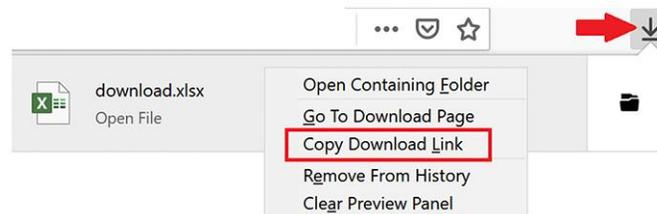
(2) If categorical (or binary) data does not match the categories used by WPdx, users can use If/Then logic when programming the datagrid to transform their categories into the categories WPdx uses. This could extend to open text fields, if users want to change “local” wording for more internationally accepted terminology (for water source types, for example). This could also be used to translate from local languages to English without having to change the wording in the surveys/responses themselves.

(3) If users collect relevant water point data across multiple different surveys, they can pull the relevant data into a single datagrid by using the Water Point as the data source, then pulling in data from other surveys by selecting “Related Surveys” → “+Add Related Survey” → then selecting the survey and data field of interest.



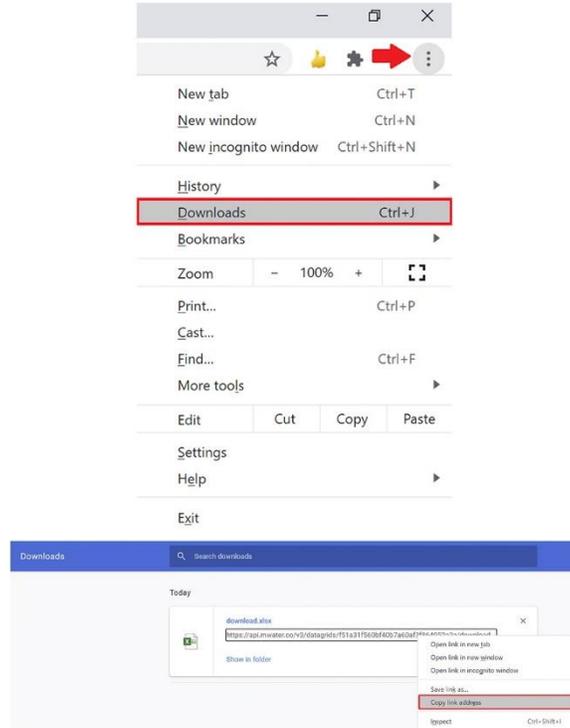
In Firefox:

1. Left click on the download arrow in upper right-hand the Firefox window.
2. Right click on your mWater file
3. Select Copy Download Link
4. Save this link or copy directly into the Direct URL field (Source of Data section) in your WPdx ingestion engine processing task
5. Map the column headers from your dataset to the WPdx data standard using the dropdown menus.
6. Uploads can be scheduled on a monthly, quarterly, etc. basis. Please provide a recommended update schedule in the Description section of the Processing Task. (one time, monthly, quarterly, annually, etc)



In Chrome:

1. Left click on the three vertical dots in the upper right-hand corner of the Chrome window.
2. Select Downloads
3. Right click on the download link and select ‘Copy link address’
4. Save this link or copy directly into the Direct URL field (Source of Data section) in your WPdx ingestion engine processing task
5. Map the column headers from your dataset to the WPdx data standard using the dropdown menus.
6. Uploads can be scheduled on a monthly, quarterly, etc. basis. Please provide a recommended update schedule in the Description section of the Processing Task. (one time, monthly, quarterly, annually, etc)



In Microsoft Edge:

1. Left click on the three horizontal dots in the upper right-hand corner of the Edge window.
2. Select Downloads
3. Right click on the download link and select 'Copy link address'
4. Save this link or copy directly into the Direct URL field (Source of Data section) in your WPdx ingestion engine processing task
5. Map the column headers from your dataset to the WPdx data standard using the dropdown menus.
6. Uploads can be scheduled on a monthly, quarterly, etc. basis. Please provide a recommended update schedule in the Description section of the Processing Task. (one time, monthly, quarterly, annually, etc.)

