

WPdx Cleaning Terms for:

- #water_source
- #water_tech
- #management
- #status

And

- facility_type categorization

A first pass at cleaning is conducted using the terms below and some fuzzy matching using natural language processing. Additional manual cleaning is done on a case-by-case basis.

#water_tech to water_source_clean mapping

```
WATER_TECH_MAPPING = {
  'Rope and Bucket': [
    'bucket',
  ],
  'Hand Pump - Afridev': [
    'afri',
    'dev',
    'aferdive',
  ],
  'Hand Pump - PB Mark II': [
    'pb',
  ],
  'Hand Pump - India Mark': [
    'indi',
    'mk',
    'mar',
  ],
  'Hand Pump - Nira': [
    'nira',
  ],
  'Hand Pump - Vergnet': [
    'verg',
  ],
  'Hand Pump - Kardia': [
    'kardia',
  ],
  'Hand Pump - Indus': [
    'indus',
  ],
  'Hand Pump - Consallen': [
    'consallen',
  ],
  'Hand Pump - Inkar': [
    'inkar',
  ],
}
```

],
'Hand Pump - Canzee': [
 'canzee',
],
'Hand Pump - Pamir': [
 'pamir',
],
'Hand Pump - Kabul': [
 'kabul',
],
'Hand Pump - SWN 80': [
 'swn 80',
],
'Hand Pump': [
 '!not a manual',
 '!hand dug',
 'manu',
 'handpump',
 'bush',
 'hand',
 'elephant',
 'manuelle',
 'rope pump',
 'rope washer pump',
 'ropepump',
 'blue Pump',
 'emas',
],
'Stand Post': [
 'tap',
 'Stand',
'standpipe'
],
'Kiosk': [
 'kiosk',
],
'Hydram': [
 'hydram',
],
'Mechanized Pump': [
 '!non mechanized',
 'mecha',
 'elec',
 'submersible',
 'submasive',
 'motori',
 'surface pump',
],
'Mechanized Pump - Diesel': [
 'deisel',

```
'diesel',  
'gener',  
],  
'Mechanized Pump - Solar': [  
  'solar',  
],  
'Tapstand': [  
  'tap',  
  'Stand',  
'standpipe'  
],  
'Kiosk': [  
  'kiosk',  
],
```

*Note that we are working to further refine hand pump types and will continue to update as possible. (ie, Hand Pump – Bush Type, Hand Pump – EMAS, etc.)

#water_source to water_source_clean mapping

```
WATER_SOURCE_MAPPING = {
  'Borehole': [
    'tube', 'hole', 'deep', 'drill', 'forage', 'bore',
  ],
  'Rainwater Harvesting': [
    'rain',
    'roof',
    # 'harvest',
    # 'storage',
    # 'catchment',
  ],
  'Sand or Sub-surface Dam': [
    'sand',
    'sub-surface',
  ],
  'Undefined Shallow Well': [
    '!unprotected',
    '!protected',
    'dug well',
    'shallow well',
  ],
  'Protected Shallow Well': [
    '!unprotected',
    '!un protected',
    'protected dug well',
    'protected shallow well',
    'protected tradional hand dug well'
    'protected well',
  ],
  'Unprotected Shallow Well': [
    'unprotected hand dug well',
    'unprotected dug well'
    'unprotected shallow well',
    'unprotected constructed hand dug well'
    'unprotected well',
    'unprotected well/',
  ],
  'Undefined Spring': [
    '!unprotected spring',
    '!protected spring',
    'Spring',
  ],
  'Protected Spring': [
    '!unprotected spring',
    'protected spring',
    'proteced spring catchment',
    'improved spring',
  ],
}
```

```
],
'Unprotected Spring': [
  'unprotected spring',
  'unimproved spring',
],
'Surface Water (River/Stream/Lake/Pond/Dam)': [
  'surface pump',
  'river',
  'stream',
  'lake',
  'pond',
  'surface',
  'dam',
],
'Piped Water': [
  'piped'
],
'Delivered Water': [
  'truck'
],
'Packaged water': [
  'bottled',
  'sachet'
```

#management to management_clean mapping

Community

- Community
- VWC
- Small community WSMT (WATSAN)
- Small community WSMT (WSDB)
- WUG
- WUA
- water committee
- Water Board
- Small town WSMT (WSDB)
- Small Town WSMT (WSDB)
- Reservation Quarters
- watsan
- member
- Elders
- Leaders
- Town Chief
- Traditional rulers
- Area
- Youth group
- Headman
- Opinion
- WASHCOM
- Residents
- Water user association
- Mechanics
- Village
- People
- Assemblyman

Direct Government Operation

- DWD
- Government
- MOHS
- WSD
- DHMT
- GOSL
- Office of the Prime Minister
- Govt workers
- Ministry
- District
- Gov
- Govt
- salwaco
- city
- assembly

- MDC
- LGA
- Ward
- Federal
- Parish

Private Operator/Delegated Management

- Operator
- Delegated
- Delegee
- Délégué
- Entrepreneur
- SME
- Enterprise
- company
- Holdings
- Salini construction
- Vendor

School

- School
- headmaster
- students
- college
- teacher
- Education
- smc
- Primary
- Secondary
- Sch.
- PTA
- Headmistress
- Student
- Principal

Health Care Facility

- Hospital
- health
- PHU
- clinic
- mch
- nurse
- doctor
- The Incharge

- c.h.c
- cawec
- Maternity
- Healthcare

Religious Institution

- Religious
- LWF
- Mosque
- ADRA
- EU - ADRA
- Catholic
- Church
- Muslim
- Mission
- father kibira
- Moslem

Other Institution

- Institutional
- military
- police
- prison
- SLRA
- RSLAF
- NACSA
- NASSIT
- MASAF

Private Person

- Private owner
- Private Person
- Priv?e
- private individual

Other

- C & D
- UNICEF
- GOAL
- Oxfam
- Lion Heart Foundation
- Mercy Corps
- Medair
- Red Cross
- ACT-U
- World Vision
- GIZ

- ICRC
- Caritas
- PLAN
- Care
- CDC
- Donor
- USAID
- AfDB
- Philanthropist

No Management structure

- No management
- No manager
- none
- No one
- Nobody
- null
- No maintenance
- Aucune gestion

#status to status_clean mapping

Fully Functional

- Functional
- Always
- No problem
- Operative
- No leak
- Operational
- Fonctionne normalement
- Normal
- No_pump_problem
- Working
- No downtime

Functional but needs repair

- Minor
- Partially
- Partial
- Leak
- Not enough
- Not effective
- Poor
- Fonctionne mal
- Servicing
- Bad state
- Bad shape
- Partly damaged
- Sub-optimally
- But with problems
- Low yield
- Functional but in need

Not Functional due to dry season

- Dry
- Not enough
- Seasonal
- It is dry
- Not enough water
- Natural
- Low water
- Water shortage

Not functional and needs repair

- Major
- Broken

- Not working
- Not Function
- Non-Functional
- Not functional
- Not functioning
- Non functional
- Non-func
- Non-Operational
- En panne
- Non équipé
- Repair
- Not Working
- Rehabilitation
- Construction
- Breaking
- Spare parts
- Mechanic
- Break
- Damaged
- Dead

Abandoned

- Abandoned
- Not in use
- Collapsed
- No longer
- Abandoned
- Stolen
- Permanently
- Abandonné
- Deactivated
- Disused

Other

- Funds
- Management / financial problem

Facility Type Categories (facility_type):

Improved:

- Borehole
- Protected shallow well
- Protected well
- Protected spring
- Undefined shallow well with hand pump
- Undefined spring with tapstand
- Undefined well with mechanical pump
- Rainwater harvesting
- Piped water*
- Sand or Sub-surface dam (with hand pump and/or described as protected)
- Packaged water**
- Delivered water**
- Tapstand (when no source is provided)
- Kiosk (when no source is provided)
- Mechanized pump (when no source is provided)

Unimproved:

- Unprotected shallow well
- Unprotected spring

No Facilities:

- Surface water (river/stream/lake/pond/dam)

Unknown:

- Undefined shallow well with no #water_tech information provided (ie, handpump, tap, etc.)
- Undefined spring with no #water_tech information provided (ie, handpump, tap, etc.)

* Assumed to be standpipe/tapstand as a specific GPS location has been provided

** These are included in the WPdx+ dataset, but not in analyses to determine population with basic access as the number of people using these sources is undefined.