



# Responses to Public Comment from the WPDx Working Group

---

*Note: Numbers in blue represent the corresponding section of standard as presented in public comment at <https://collaborase.com/wpdx>. Indented comments represent replies to original comments by other reviewers. Blue text is the WPDx Response.*

## 0

- Water Quality is missing in this database - would like to see basic TDS / Conductivity & Bacteriological
  - Thanks so much Jamie! This is question we have gotten fairly often and something I would love to integrate. In the desktop review that we completed we didn't find a common attribute (50%) that captured water quality. Unfortunately given the incredibly wide range of how organizations capture water quality there appears to be no single indicator that we could include that would work for the majority of folks. I am hoping that as the sector continues to become more coordinated we will see an agreed-upon framework for measuring quality emerge that could be captured in future versions of WPDx. It would be great to hear any other ideas that you or others may have on how best to capture this!- Brian
  - Total Dissolved Solids is probably the most common type of water-quality information (also measured as conductivity or salinity).
  - **Unfortunately, only 18% of data sets captured either conductivity or TDS, and salinity wasn't recorded. The most common water quality attribute was turbidity, with 22% data frameworks including this. Unfortunately, this is still far short of the threshold of "majority," so water quality will not be included in this version of WPDx. Should the sector move towards a common water quality metric, there would be opportunities to incorporate this in the future.**
- How to find the right balance between integrating existing datasets providing analytical possibilities and start providing a standard for future collections?: I would say that several requirements weight on the definition of the fields: 1/ some field in as many as possible existing datasets can be used to fill it 2/ the values can be analysed in themselves and between different datasets 3/ the definition of fields are used in the design of future surveys. The coding of fields is a solution for objectives 2 and 3 but it will require some work on existing datasets and exclude some "uncodable" values which

would be lost behind a “other” code. Keeping the imported fields as such provide little analytical value. So would it then be relevant to have 2 fields one coded one raw? And to what extent would the dataset providers be willing to do the work to code their data to fit in the codes chosen in the standard ? On a more general note could you also tell us how the WPDx team approaches this challenge?

- **WPDx is focused on sharing what people are already collecting at this point, rather than shaping the collection process. While recommended categories could be added in the future, the emphasis now remains focused on sharing existing data. To implement coding requires both determining a list of categories and defining each category. While this may seem straightforward, the data we’ve seen suggests too much variation to make this effective at this point. As the sector continues to evolve, there may be opportunities to begin providing recommended categories through WPDx. In the meantime, we encourage stakeholders to independently develop knowledge-products that map values provided to categorical fields to aid analysis.**

## 1

- Section 1 description minor typo: Minor comment but says 1.1-1.5 but there are 6 indicators here.
  - Thanks Rob. You're absolutely correct. I will suggest that we make this change in the final standard for the working group.
  - **The numbering system has been updated to remove the categories.**
- Decimal places: Is there a recommended or minimum number of decimal places?
  - Good question Rob. Do others have experience on this? Are folks familiar with a standard number of decimal points for WGS 1984?
  - Four decimals 0.0001 brings you to 5-10 m accuracy depending on your location on the globe; consequently 5 decimals is 0.5-1 m accuracy. Latest off-the-shelf GPS chips in phones and tablets can offer 3-4 m accuracy and considering the fact that hardly any WP location is recorded with sub-meter accuracy the fifth decimal is all you need. In short: IF you want to be able to trace a location within 10 m accurate you need 5 decimals: 0.00001. So 5 decimals should be the minimum requirement and at the same time anything more is over kill.
  - **Given the emphasis on sharing as much data as possible, the standard has been updated to require at least 4 decimals.**
- require nearest named location in addition to gps points: Spatial analysis requires high certainty of the location of the data. For quality control purposes having a second location identifier such as the village or smallest administrative district. in addition to the GPS latitude/longitude helps identify errors in GPS point collection--- such as satellite errors as well as incorrectly identified GPS locations by the person in the field. For example-- I've seen one instance where the contractor in the field collected household data and some remote households were called from the capital city. The GPS coordinates were identified as the capital city while the household data corresponded in actuality to a remote village many km away.

- Thanks so much Elizabeth! This is a great point. Given your good suggestion I believe that Indicators 1.3 (Village) and 1.4 (District) should work to support data validation. Please let me know if you are not sure that these will address the issue you've described. –Brian
- Thanks for your reply Brian. I believe 1.3 and 1.4 would help with the data validation yes. I was just unsure whether these indicators would be required components of the data you aggregated together or whether the user will need to filter the database to use only the selected points that have 1.3 or 1.4.
- **As “village” was captured in 70% of data frameworks reviewed, it will not be required, as that would bar some 30% of data from inclusion. That said, it is expected that village data would be quite common. If needed, users could filter data by cases where district or village was provided.**
- Add option called “Name of Water Scheme”: I recommend adding a field called Name of Water Scheme. There can be multiple water schemes in a community and multiple communities can also rely on one water scheme.
  - **This unique identifier, captured in 1.5 (Water Point ID) will allow for the name of the water scheme to be captured.**
- Mark the map DATUM!: WGS84 is an obvious datum requirement but also a tricky one. Commonly this is the default setting for many GPSes so no problem there. However if people start to wonder what a DATUM is and how to find out what they have used they might run into trouble if they lack the knowledge for it. Besides this there is the issue that most countries do not use WGS84 as a national standard. usually countries use national grids. therefore location data which is gathered through a national project under national guidelines is unlikely to be collected with a WGS84 datum. As with degree notation there are easy conversions possible as long as the used map datum is known. The important issue therefore is that the location DATUM must be known. If the grid system used on the GPS is not recorded in the metadata there is no possibility to match the GPS coordinates. If you demand WGS84 data to be uploaded you still do not know whether this is truly done or whether the conversion is done right. Recommendation: “map datum” should be recorded as a required field. preferably including a comment on whether any conversion between grids or projections has taken place.
  - Seconding this. Shifts between datums can cause points to be misplaced. WGS84 is rather widely adopted and seems a suitable candidate to serve as a standard. It is also used in the framework of IATI (see links). I think it would be best if WPDx and IATI would overlap on such specifications unless specific objections arise. <http://iatistandard.org/201/activity-standard/iati-activities/iati-activity/location/point/>  
<http://www.opengis.net/def/crs/EPSSG/0/4326>
  - **The raw data set will be linked, and will provide datum if it is provided. For compliance with the standard, all data will be converted to WGS84 by the WPDx team, if it is not already in that format. Given this suggestion, an additional field will be created to capture what has been converted from the original data set.**

- Add altitude and precision ?: It is my impression that many if not all GPS systems provide values for “altitude” and “precision”. Both would provide additional factual information for little additional work.
  - **Only 10% data frameworks included altitude, and precision/accuracy was captured in only 6% of data frameworks. As these do not meet the threshold of “attributes collected by the majority”, they will not be included in WPDx at this time.**

## 1.1

- GPS as Identifier: I'm wondering whether the GPS position could be used to identify datapoints across datasets.
  - This would be approximate since GPS measurements are likely to vary slightly from user to user.
  - Thanks David! This is exactly why we have proposed Unique ID as well as GPS.
  - A 3 word address could also be considered as complementary to GPS coordinates. It makes it easier to communicate location using more memorable words and the words are available in multiple languages. [www.what3words.com/about](http://www.what3words.com/about) The words are also less prone to error when sharing information about the location (an autosuggest function automatically corrects errors and makes suggestions about nearest locations).
  - **As mentioned above, GPS measurements are not precise enough to replace Water Point ID. As none of the evaluated frameworks captured a 3 word address, this will not be included in the current version of the standard.**
- More precise wording and give example: More precise to say Longitude in decimal degrees ie - 10.12345
  - Thanks David. Do others agree that we should switch “Geo-Coordinate X” with “Longitude”? Will this be confusing to folks that collect data in terms of Northings and Eastings? As far as the example would a format example work (instead of a specific example)? This would be something like -##.#####. Or would a random Latitude/Longitude be more helpful?
  - It would be possible to accept both depending on the person collecting data and their preference -there are easy ways to automatically standardise the data collected back to a decimalised system within the database. Need then to provide a sense check that the data point is in the right country / not in the sea etc. best done on site.
  - In this era of GPS I think latitude and longitude in decimal degrees (WGS84) should be required. If data are loaded in bulk from digital files then the person loading the data can do the conversion.
  - **“Geo-Coordinate X” and “Geo-Coordinate Y” have been changed to Longitude and Latitude. An example is now provided with the definition.**
- Accuracy and clarity needed. Also altitude/elevation!: I think using the terminology X-Y is misleading. It could be interpreted as UTM grid northings and eastings which require a third number (UTM zone). We should use Lat-Long in our terminology if that's what we're asking for. WGS 84 isn't a format: it's the datum used for these geodetic coordinates. It should be made clear that it isn't DMS Or DM.xxx but decimal degrees. The info shouldn't contain more than 5

decimal points (1 meter accuracy is overly ambitious already). A note to specify north and east are positive is pertinent. Also agree with David W. Please give an example. Finally I believe altitude can be a useful data point and it's already captured by the GPS. It could help for instance determine if two boreholes are tapping the same aquifer. Hope this helps. Written from: 1.9782 32.5168 :)

- **“Geo-Coordinate X” and “Geo-Coordinate Y” have been changed to Longitude and Latitude. The definition has been updated to clarify that the standard is requiring decimal degrees, and a minimum of four decimals are required. A note has also been added to clarify that north and east are positive.**

### 1.3

- Problems: Unique spelling at village level: Need to use a standard to name villages - In Timor there is a government list of spellings - locally there can be 2-5 different ways to spell a certain village name - coding the village from largest administrative unit to smallest is really helpful. In Timor each 13 district is a 2 digit number, sub-district 2 digit number, 422 suco (village) 2 digit number, 2228 Aldeia (hamlet) 2 digit number. Every hamlet has a unique 8 digit number and official spelling - using this and repopulated drop down style forms helps enormously with consistency. Keith
  - Indeed local spelling and the mistakes made while recording are a pain. Tanzania used a similar identifier system. problem occurred there with the field recording. this many numbers is prone to errors and the identifier has no link to the local name which users sometimes use with “affection” as the WP might be named after the person who donated it. In TZ 75% of the identifiers were erased from WP's after 3 years due to weathering or vandalism. Villagers can only relate to local names so to retrieve a WP there needs to be some record of the local name as well.
  - Overall question - are all text entries supposed to be in English? Will the database accept Amharic Arabic Asian language characters? Will we have some translation of instructions to Spanish French etc.? Agree that village names sometimes have multiple spellings (e.g. Ethiopia where ) or can be the same as other villages (e.g. Central America) but in combination with geolocators we should still get a decent proxy for unique identifier.
  - Also perhaps in instructions we could encourage data collectors to use standard village names/codes where available and/or to clean their data using those before sharing it?
  - **As not all countries have a similar post-code system, village will still be included as an open text field. That said, given the challenges with relying on village alone, this will be complimented by coordinates and water point ID, helping to address issues of locating the water point and issues of unique identification. At this point, the system will not be limited to English, though if data needs to be converted or mapped to fields, it will need to be in a language the WPDx team is able to access. The decision on foreign language characters will depend on the technological limitations of the online platform. Instructions will request that data is cleaned before sharing.**

- What if the water point is located in a peri-urban area?: Should the name of the city be used? The name of the neighbourhood or commune would be preferable. Actually those 2 fields village/district could be changed to 2 (or more) pairs of fields. For each pair the first field would give the type of jurisdiction (village commune district) and the 2nd its actual name. For instance “Village” “Bosobolo” “District” “Basenga”. Since this issue of variability of the administrative organisation inside a country or between countries must have been tackled by other geographical normalisation efforts one might look at the solutions they found.
- Comment-What about water points shared by more than one village?
  - **The village name provided in the data will be used. This is almost always just one village.**
- The use of the word “village” under both items 3 and 4 is confusing. For us village level will be either a local municipality ward or a community whilst district level is a district municipality.
  - **This has been updated to be “smallest administrative division provided.”**

## 1.4

- Given “district” means different things in different countries presumably this is intended to be the highest administrative unit/division? (e.g. equates to “County” in Kenya ; Liberia “District” in Uganda & Sierra Leone “Region” in Ghana Province in Zambia): Perhaps we could call this “Largest Administrative Unit”
- Because “District” can mean different things in different countries should this be called “Largest Administrative Unit?”
  - This is a good point - there is a large variance in what administrative divisions are called. I believe there have been some efforts to standardize this - see for example: [http://en.wikipedia.org/wiki/ISO\\_3166-2](http://en.wikipedia.org/wiki/ISO_3166-2). Spelling differences (especially at lower administrative units) might also be minimized by adhering to some standard listing on a country-by-country basis - for example in many Bantu languages 'r' is interchangeable with 'l'
  - It might be valuable to add another level (make it optional) as it could be hard to identify some locations only using two administrative units. For example province / district / commune / village is the breakdown in Vietnam and Cambodia. There are a number of duplicate commune names.
  - I like the idea of another level and optional in Nicaragua it is Departamento / Municipalidad / Comarca / Comunidad
  - You either record just the WP ID/name and the coordinates as a location/object identifier or you need to cater for a lot of administrative units. in TZ we have: Region District Local Government Authority or Village Subvillage I don't know how much use it is to know all these layers. You don't need them to map the WP's. Perhaps the largest admin unit would be sufficient to pinpoint an object roughly. However in TZ it would in that case be more useful to know district rather than Region.
  - There are authoritative lists of administrative units at the NGA Geographic Names Database and at Geonames.org. The administrative levels are organized as ADMIN1 ADMIN2 etc. These databases also have village names so you could suggest that people

report for each waterpoint an ADMIN1 name (or an ADMIN2 name) and a village name from one of these authoritative lists.

- I see two purposes for this location name information 1) a basic validation of the GPS data 2) allow easy listing and sorting of the water points based on their location names in a table format rather than the map display only. If purpose #2 isn't relevant then we may not need to add the third administrative level. If purpose 2 is relevant it would be painful and in some cases impossible (in the Vietnamese context in particular) to easily identify the location of these water points if you only listed two administrative levels. I like the Geonames.org database suggestion (report ADMIN level 1 ADMIN level 2 ADMIN level 3)
- **Two fields have been included, based on external standards. These will include “first-order” / “primary” / “First Level” administrative division, as well as “second-order” / “secondary” / “Second Level” administrative divisions. This ties to the structures used for <http://Geonames.org>, <http://www.statoids.com>, and <http://www.gadm.org>.**
- Refer to external standards for administrative divisions: The problem of categorizing the different political or administrative boundaries is complex. One can not assume that district is a universal administrative unit since these vary by country. I suggest that you not attempt to solve this problem in a WP standard but rather refer to an external effort such as GDAM. You could say simply administrative areas should be expressed using GDAM format.
  - Thanks John. This is a great suggestion. As you can see in the other comment several folks have weighed in on this same issue. Another possible option could be to use the secondary administrative division from <http://www.statoids.com/>.
  - **Two fields have been included, based on external standards. These will include “first-order” / “primary” / “First Level” administrative division, as well as “second-order” / “secondary” / “Second Level” administrative divisions. This ties to the structures used for <http://Geonames.org>, <http://www.statoids.com>, and <http://www.gadm.org>.**

## 1.5

- how to ensure uniqueness of the ID?: Is there a unified method for defining such Unique IDs? I can think of the way electronic equipment (network cards mobile phones...) is assigned a worldwide unique ID something like: Unique provider ID (could be one per NGO company etc...) that installs the water point (or assigns the ID when there is none). Coded on 8 digits for instance equipment unique ID which each organisation can assign to their liking provided it is unique within the org's scope. Coded on 10 digits for instance. The resulting Unique ID would look like [providerID][equipmentID] on 18 digits. If there is no agreed upon scheme for numbering water points the unique ID simply won't work.
  - I would love to see a global or national unique identifier system but since we can't make sure all water system installers follow any kind of best practices this is a pipe dream. Again if there is any kind of indicator on a handpump it should be recorded. Along with the geolocators; village name we will likely get close to unique.
  - I have to agree with Susan here. Developing a unique water point ID system is solvable for a single platform or government (mWater has a very simple system that works well);

however agreeing on a universal standard would require buy-in of sovereign government ministries around the world. What is possible: create a translation service that serves up the many different IDs for a water point via an open-access API. We are building just that and encourage others to do the same. For this to work the current standard must have as few restrictions on this field as possible.

- **As noted by comments, standardizing unique identifiers globally is beyond the scope of this effort. When this is achieved in the sector, it may be included into WPDx.**
- ID format: There needs to be more clarity given as to the water point ID format
  - **Recognizing the wide range of formats for water point IDs, this will remain an open text box for this version.**

## 1.6

- Put full name of the country instead of the two-letter ISO code
  - Could be stored as the 2 letter code but selected in a dropdown list of countries to avoid confusion.
  - **The standard will require the two letter code to avoid chances of misspelling at this point. In the future, an online platform can offer the option of selecting a drop-box of full country names.**

## 2

- Water Treatment: Has any thought been given to including a field for treatment method (e.g. settling filtration disinfection etc.)?
  - **Unfortunately water quality is not commonly collected any standardized way. The most common water quality attribute was turbidity, with 22% of data frameworks including this. Unfortunately, this is still far short of the threshold of “majority,” so water quality will not be included in this version of WPDx. Should the sector move towards a common water quality metric, there would be opportunities to incorporate this in the future.**
- Rework entire section to focus on the water point: Water points are usually mapped and surveyed by field workers who have little sector knowledge and limited time to probe for details that they cannot immediately observe. Also the water users often do not have reliable knowledge about the ultimate source or technology used to supply their water particularly if it is piped water. Therefore this standard needs to focus on the water point itself. mWater and WaterAid recently examined this problem in depth and designed an approach that keeps the focus on the water point but allows the addition of optional attributes that might be useful to an implementing organization. The resulting definitions are available in a google doc:[https://docs.google.com/spreadsheets/d/1FFpIpOIdR8yB5v8WlxvZuUjHJsEcfFhIK\\_3qGIsJiQk/edit#gid=697710437](https://docs.google.com/spreadsheets/d/1FFpIpOIdR8yB5v8WlxvZuUjHJsEcfFhIK_3qGIsJiQk/edit#gid=697710437) To summarize: Water points are classified by the JMP types. We follow the definitions used by the UNICEF / WHO Joint Monitoring Programme not because they are perfect but because they are universally recognized by governments and national surveys. These types have been used for the past 15 years and will continue to be used in monitoring progress against the post-2015 Sustainable Development Goals. Water points can have optional



attributes that depend on the type selected. Attributes are additional statements about the water point that are not always observable on a visit but are important in understanding how the system was designed and built. These include: Pump/lifting device (Afridev India MKIII etc) Powered by (applies specifically to pump/lifting device; e.g. manual wind electrical grid) Drilling method. Supply (applies where the supply is upstream of the water point e.g. gravity fed piped water from a spring source or dug well built into a sand dam - the sand dam is the source) Treatment works (any system between the source and the water point used to improve water quality e.g. slow sand filter coagulation etc.) In mWater's implementation of this system specifying a type or attribute is optional. However if you choose to specify a type you must use one of the JMP types. This prevents the common problem we see in water organization's datasets where they have 10 different spellings or phrases for "unprotected shallow dug well." My proposal for this standard is to use JMP types for Water Point Type in Section 2.1 and make Section 2.2 Water Point Attributes.

- **WPDx is focused on sharing what people are already collecting at this point, rather than shaping the collection process. While the JMP types are universally known, data is not universally collected according to these categories. While recommended categories could be added in the future, the emphasis now remains focused on sharing existing data. At this point, this requires an open text box. As the sector continues to evolve, there may be opportunities to begin providing recommended categories through WPDx, as proposed. In the meantime, we encourage stakeholders to independently develop knowledge-products that map values provided to categorical fields to aid analysis. Regarding source, the desktop review found this to be collected by the majority of data frameworks.**

## 2.1

- Coding: Probably should add some guidelines on this or a coding system that people can map to so people aren't using different names for what may be the same water point type. BH for borehole EPS for Electric pumping system or something like that.
  - Cancel EPS that would be covered by Tech item. Same comment on tech some sort of code to unify.
  - I think the codes are a good idea. But also include an 'other' open text field
  - SOURCE is confusing. it should not relate to the WP type.
  - By Water Source do you mean Site Type ie Groundwater River Lake Spring? But if this is strictly a waterpoint database then Water Source would have to do with which aquifer is being tapped which usually is not known or reported.
  - **WPDx is focused on sharing what people are already collecting at this point, rather than shaping the collection process. While recommended categories could be added in the future, the emphasis now remains focused on sharing existing data. At this point, this requires an open text box, as data isn't coded in a consistent manner. As the sector continues to evolve, there may be opportunities to begin providing recommended categories through WPDx, as proposed. In the meantime, we encourage stakeholders to independently develop knowledge-products that map**

values provided to categorical fields to aid analysis. Source refers to the type of source that the water is coming from. Examples include borehole, spring, dug well, and municipal water.

- 
- Missing important parameters: The water source and technology items do not adequately cover the level of supply provided. For example a spring gravity fed system says nothing about the level of service provided. There is also nothing about quality of water source/provided. For all intents and purposes a system that provides non-potable water means lack of service delivery. Water source adequacy is also not addressed. We really need to know about the water point adequacy in terms of demand both now and going into the future.
  - **The standard reflects the attributes that are currently collected by the majority of stakeholders. As service delivery attributes become more commonly collected, they may be incorporated into the WPDx standard.**

## 2.2

- Codelist: In the working group we discussed that this should/could not be standardised and enforced. I wonder if (from your analysis) a list of options could be suggested if people want to use that.
  - Marten this is an interesting idea. Our analysis was limited in terms of values and more focused on the indicators. What would you think of doing this in a future version using data from the initial set of data collected? We'll also have to consider how much more management that will take if the working group needs to be managing and updating these types of category lists as well as the standard itself. That said the analytic benefit could certainly be worth it depending on how rough the initial data set looks.
  - My review is standardised coding is a vital element to data sharing. This should cover almost all standard elements in water supply / service delivery systems. An “other” field to mop up the rest - each time the data sets are analysed coding to include others can keep improving the data sets captured without having to use “other”. This covers water point technology tanks construction material pipes etc.
  - There are two parts here: getting the water out of the well (ie bucket pump) and then transporting the water. Not sure how you simplify a multitude of storage/delivery possibilities unless you simplify (ie None Gravity Water Truck Pressurized pipe).
  - **WPDx is focused on sharing what people are already collecting at this point, rather than shaping the collection process. While recommended categories could be added in the future, the emphasis now remains focused on sharing existing data. At this point, this requires an open text box, as data isn't coded in a consistent manner. As the sector continues to evolve, there may be opportunities to begin providing recommended categories through WPDx, as proposed. In the meantime, we encourage stakeholders to independently develop knowledge-products that map values provided to categorical fields to aid analysis. Regarding transportation of water, this is not included in the current standard given the focus on water points, rather than places where water is consumed.**

### 3

- Savings for CapManEx: Was a field for savings (i.e. amount of funds available to cover CapManEx) considered?
  - **This data was not found to meet the threshold of “collected by the majority”. As such, it will not be included in the standard at this point.**
- Shared water systems/points: Maybe this field is where we could have notes about whether the water system/point is shared.
  - **This data was not found to meet the threshold of “collected by the majority”. As such, it will not be included in the standard at this point.**

### 3.1

- Presumably installation year is date when waterpoint was first installed (e.g borehole drilled) and NOT date when current supply technology was most recently installed.
  - This is a great point. Much of the data we’ve seen doesn’t specify which type of data we received (i.e. date drilled or installation of the specific supply technology). Going on the principle of taking what is already being collected (i.e. the date provided) this may be an opportunity for improvement of the standard in the future.
  - Would be good to clarify which year is requested likely first installation is appropriate - an optional data/text field to add when rehabs were performed might work.
  - Would be logical to have the year correspond to the tech and source. but likely only one year is recorded which is the original installation. in TZ we encountered however breakdown dates that preceded installation dates. Still wondering whether this then referred to a re-installation or a mere typo. Unless you get a separate source date and a tech date you can never be certain.
  - **The data currently collected typically asks for “year of construction” or something similar, without clarity on whether this refers to initial construction or rehabilitation. At this point, with an emphasis on sharing data that is currently being collected, the year provided will be shared, recognizing that this doesn’t provide absolute certainty.**

### 3.2

- “Management structure” is somewhat vague. : By “Management Structure” do you mean waterpoint owner? Are you looking for the type of owner or the name of the owner?
  - **This will be the type of owner, and it will be selected from a dropdown list.**

### 3.4

- In the formal definition you may need to specify how this field applies to waterpoints that have differential fees based on membership (e.g. on south coast of Kenya it is common for members to pay 50 Ksh per month and non-members to pay 2ksh per jerrican). In such cases it might make sense for data collector to input the fee that members pay.

- This is really interesting. Are others finding a difference as well? Should we clarify €œif multiple rates exist please include the most expensive tariff.€² We could also request that this capture the cheapest tariff if multiple rates are included in the data. We would need to decide which is more interesting.
- This is really interesting. I didn't see many records with multiple tariffs so this may not be showing in the data anyways. Are others finding multiple tariffs in your data sets?
- Yes multiple tariff “levels” are a common element of financial plans in the areas where we work.
- The data currently being collected does not tend to show multiple tariffs. For the current version, whatever tariff data is provided will be used.
- currency: Might be easier to have a separate item for currency rather than open text so people are forced to choose from ISO code.
  - Thanks Rob! This is a great idea that we hadn't considered. As you can tell we've tried to capture the cost currency and unit all in one indicator. You raise a great point though on data validation. I believe this is ultimately a technical question. If we can implement data validation on the platform that ensures that an ISO code appears in the open text that seems like it should be fine. If not we may need to add a second field as you've suggested. Ultimately I take your point that data validation is needed to ensure that folks are formatting this field properly including a recognized currency code.
  - Seems somewhat redundant with “Payment Structure” since payment structure is the unit and unit is also included in “Cost per Unit”. For example if Payment Structure is “per container” then “Cost Per Unit” is GHS 15/container.
  - **The reason it was initially proposed to be split is because some data sources only provide the basis of payment (i.e. monthly, or as needed) without any currency. Given this discussion, the information will be provided in a raw format, which may include the cost and basis, or just the basis in some cases.**
- Rising block tariffs: Payment structure is often via a rising block tariff with a concomitant number of unit costs. It is unclear whether or not item 12 allows for such multiple entries.
  - **This will be an open text box. Multiple pieces of data can be concatenated and included in one cell.**

### 3.5

- clarify exactly which type of implementer you want: Need to make this clearer. Currently it implies that e.g. if the borehole was installed by a drilling company then this name would be entered. Perhaps a better definition (assuming this is what you want) would be “enter the name of the entity who directly managed the programme which led to this WP being installed” Some examples (i) put “Government of Tanzania” and not “DrillCo pty” (ii) put “World Vision Malawi” not “DFID” or “Partner NGO X”. Or would you want the local NGO named rather than the INGO? (Ian Ross - Oxford Policy Management)
  - Might be easier to split this into two categories: funder(s) and installer? This is how the Tanzanian data does it - in some cases it is the same but in many cases and NGO pays a

contractor to do the actual installation. Both pieces of information are potentially valuable.

- suggest to split. These fields are however also the most prone to error as they are entered in the field as open text. in the TZ data “German government” as a donor is recorded with 10 different spellings ranging from “G” to “german” “germen” “Germany Gov” etc. It's two very interesting fields to have but for analysis they are consequently either very useless or cumbersome.
- “Installer” might be a better name for this field. Are you interested in who installed the well or who installed the water delivery system?
- Source of funding (i.e. Donor) was not collected by the majority of stakeholders, and has not been included. Implementer was described as “Installer”, “Contractor”, and “Implementing Agency” among others. The term has been changed to “installer” as this is the most common notation among data collection frameworks.

#### 4

- Add field for Water Quality: The question type can be a multiple checkbox field with options for “not known” and also a text box for comments.
  - This field would indicate whether water quality data is available or not? It appears that water quality data itself is not part of this database.
  - **Unfortunately water quality is not commonly collected any standardized way. The most common water quality attribute was turbidity, with 22% of data frameworks including this. Unfortunately, this is still far short of the threshold of “majority,” so water quality will not be included in this version of WPDx. Should the sector move towards a common water quality metric, there would be opportunities to incorporate this in the future.**
- “Water Point History” is vague.: The name “Water Point History” implies a series of information taken at multiple dates describing some change with time. “Water Point Inventory” might be a better name since it appears that each waterpoint will have a single record which represents data recorded on a single inventory visit.
  - Agree this needs clarification or re-naming. Is the idea to capture when it was rehabilitated? When it last broke down or was damaged?
  - **These section names have been removed.**
- Function to update water point functionality status in real time
  - **The online platform is expected to allow for real time data uploads.**
- Missing parameter: There is really no appropriate field to enter this comment under so I putting it in here. There is no parameter for the number of people or households served by each water point. I would have thought this was a most important statistic.
  - **This has not been included because the way there is currently no consistent way this figure is calculated across stakeholders. As such, it has been deemed subjective and excluded from the standard. If a common definition of this is determined, it may be included in a future version of the standard.**

## 4.1

- Availability of definitions/codebooks used to train data collectors: Different organisations/governments have used different definitions of key attributes in their WPM exercises with functionality (often taken as a binary indicator) a key case in point. This indicator (“Presence of Water when Assessed”) is the most absolute cross-sectional definition of functionality but not everyone will have defined it in this way. Therefore it is crucial that not only the data from each organisation is available but also the codebook and \*definitions\* used to train the data collectors. Ideally this would be attached alongside the full dataset itself (as the webinar discussion indicated would happen which is great news). If not researchers may end up comparing apples with oranges. Foster’s ground-breaking econometric work on cross-country WPM datasets was mentioned “ it is an excellent paper but the analysis surely has an element of error introduced by this definitions issue. We should make sure it is easy for future researchers to know what they are comparing.”
  - Thanks so much Ian. This is as I'm sure you know well one of the more discussed indicators. The goal of a binary “presence” vs. “absence” approach is that nearly all of the data sets we reviewed would provide a clear answer for this regardless of their specific definition. While “functionality” can mean many things it was often evident whether any water was coming out at all. For example though one organization may say “limited flow” this can still easily be coded as water was present. Though it wasn't the majority we did come across some terms that were ambiguous and didn't definitely answer the question. In these cases we would use the categorical option of “unknown” and put whatever was known in the “Condition” indicator. While this is imperfect and has its drawback it allows us to collect as much information as possible through the standard. A more restrictive definition would almost surely severely limit what could be shared. Aside from collecting these “codebooks” which may be a challenge to do consistently are there other approaches you or others would suggest? As the sector gets further aligned on functionality and we begin to see the majority of organizations using a specific structure this is definitely an area that would be terrific to see the standard evolve.
  - If there is a definition of functionality for the national water point inventory it would be nice to encourage data collectors to use that.
  - **Both the raw data and the standard compliant data will be available. Additional information would need to be gathered by contacting the data source directly.**
- limited water?: Limited might be a valid and interesting answer. or Rationing.
  - **This has not been included because the way there is currently no consistent way this figure is calculated across stakeholders. The challenge with "limited" is that this can be defined in many ways that are not well agreed. While one organization may see it as "functional, but with problems", another organization may see it as "functional." As such, it is unlikely to be objective and has been excluded from the standard. If a common definition of this is determined, it may be included in a future version of the standard.**

- I would add more categories like high flow limited flow no flow:
  - **This has not been included because the way there is currently no consistent way this figure is calculated across stakeholders. The challenge with "limited" is that this can be defined in many ways that are not well agreed. While one organization may see it as "functional, but with problems", another organization may see it as "functional." As such, it is unlikely to be objective and has been excluded from the standard. If a common definition of this is determined, it may be included in a future version of the standard.**
- Water available but pump is broken: What if water is present but the waterpump is dysfunctional. The answer would be yes but there is no practical access. In our surveys generally our partners ask: is the facility functional? My proposal is to change this question to 'Access to water' or 'IS water available on the day of visit'. Note that this takes into account that it may be dry season and that is the reason that there is not water available. From the timestamp one can deduct how to interpret the data.
  - I'm pretty sure you are agreeing with the standard. See below it means "Identify if any water is available at the time of the visit recognizing that it may be a limited flow". i.e. the question the enumerator asks themselves is "can I draw water from this pump/tap/spring right now?".
  - The main issue is to avoid the use of the word "functional". I do agree that it should refer to the availability of water during the day of recording. Many WP's have water only certain hours of the day.
  - **The standard now asks "Is water available on the day of the visit" to include cases where the visit takes place when the water point is closed.**
- Simple Presence/Absence?: By the name of the field it appears that this is a binary absent/present observation. While this is usually simple to determine other useful information would be Water Level and Well Yield.
  - **This data was not commonly collected, and accordingly, has not been included in the standard at this point.**
- What if a standpost is functional but simply closed when assessed? : How to understand the "presence of water" in case a standpost is simply closed when assessed. It might be because the standpost agent is absent or it's this time of the day when the reservoir is filled or simply it's sunday. Strictly speaking water is not "present" but it's because of perfectly normal operations and a visit some hours later might have resulted in a different value. Presence of Water when Assessed
  - **The standard now asks "Is water available on the day of the visit" to include cases where the visit takes place when the water point is closed.**
- If the answer is "no" this should be further validated by question "how long has this WP stopped working"
  - **This data was not commonly collected, and accordingly, has not been included in the standard at this point.**

- Parameter shortcomings: presence of water when assessed may not be that meaningful a parameter as it is so dependent upon chance. Should for example a system be out of operation for 1 day in the year and that day falls on the day of inspection the system would wrongly be rated as non-functional. On the other hand if it is functional only 1 day and that days falls on the day of inspection it will be rated as functional. A more meaningful assurance of supply parameter would be the number of days over a period of time (say the past year) that the system did not provide people with water. This could then be compared against basic norms to determine functionality or assurance of supply.
  - **This data was not commonly collected, and accordingly, has not been included in the standard at this point. The availability of water on the day of a visit does have the shortcomings you have noted. However, this question allows for the greatest amount of data sharing.**
- Change this to Functionality and use a simple definition: The discussion in this section has been excellent with many good points raised. No matter what standard you choose any new data will have to be mapped to a set of definitions. For that reason I propose using the most simple definition possible: Functional- the water point usually provides water everyday during opening hours. Needs repair - the water point usually provides water but at a reduced capacity because of a maintenance or supply issue. Non-functional - the water point no longer provides water on a reliable basis. No longer exists - the water point was closed or damaged beyond repair. These definitions are intentionally vague about the details in order to accommodate a wide variety of other local definitions. It is important not to confuse functionality with other attributes such as reliability continuity of service or accessibility. These should be tracked separately.
  - **Significant discussion is taking place throughout the water sector regarding definitions for functionality. At this point, there is currently no consistent way this definition is captured across stakeholders. The challenge with "needs repair" is that this can be defined in many ways that are not well agreed. While one organization may see it as "needs repair", another organization may see it as "non-functional." As such, it is unlikely to be objective and has been excluded from the standard. If a common definition of this is determined, it may be included in a future version of the standard.**

## 4.2

- Selection vs open text: Maybe a simple list of options and an open text would be interesting. Functional Broken Source Dried up .... Simple list but an option to provide more data in open text.
  - **WPDx is focused on sharing what people are already collecting at this point, rather than shaping the collection process. While recommended categories could be added in the future, the emphasis now remains focused on sharing existing data. To implement coding requires both determining a list of categories and defining each category. While this may seem straightforward, the data we've seen suggests too much variation to make this effective at this point. As the sector continues to evolve, there may be opportunities to begin providing recommended categories through WPDx. In**



the meantime, we encourage stakeholders to independently develop knowledge-products that map values provided to categorical fields to aid analysis.

- Here I would also add categories. Then is the waterquality an issue or not?
  - Condition of the WP would be technical and on visual inspection (broken handle missing stolen etc.). I would not mix this with water quality
  - **WPDx is focused on sharing what people are already collecting at this point, rather than shaping the collection process. While recommended categories could be added in the future, the emphasis now remains focused on sharing existing data. To implement coding requires both determining a list of categories and defining each category. While this may seem straightforward, the data we've seen suggests too much variation to make this effective at this point. As the sector continues to evolve, there may be opportunities to begin providing recommended categories through WPDx. In the meantime, we encourage stakeholders to independently develop knowledge-products that map values provided to categorical fields to aid analysis.**
- "Condition" is too general.: "Condition" is a general term and if the field is open text then you are going to get many different kinds of answers. If you mean waterpoint status this could be "Abandoned" "Filled" or "In use". The condition of any existing pump would be a separate bit of information as would any info about the status of the water distribution system.
  - **WPDx is focused on sharing what people are already collecting at this point, rather than shaping the collection process. While recommended categories could be added in the future, the emphasis now remains focused on sharing existing data. To implement coding requires both determining a list of categories and defining each category. While this may seem straightforward, the data we've seen suggests too much variation to make this effective at this point. As the sector continues to evolve, there may be opportunities to begin providing recommended categories through WPDx. In the meantime, we encourage stakeholders to independently develop knowledge-products that map values provided to categorical fields to aid analysis.**

## 5.1

- Time of the day?: I question the wisdom of asking for the time of the day. It sounds very superfluous.
  - Thanks Philippe! This was initially suggested to allow the database to understand continuous data. For example if a sensor is sending data every 15 minutes the time of day will be required to sequence the data. Do folks think this is worth keeping? Are there other options for sequencing data collected over short intervals? Thanks Brian
  - Time is optional so doesn't hurt to keep it but I tend to agree with Philippe.
  - When WP is recorded either time is recorded manually and written down or it is automatically recorded from GPS. It is interesting to know the time of day only if "presence of water at time of recording" is a required field. If it is recording just whether water was "present during the day of recording" the time is not relevant.
  - **Time of day will be allowed, but not required. This will allow sequencing of continuous data sources.**

- Date of Inventory?: I think “Date of inventory” is more precise than “Date of Data Collection”. Time is not so important for the inventory itself although if a water level were being measured then a time might be useful.
  - **The name of this field has been changed to “Date of Inventory” as suggested.**

## 5.2

- Engaging governments with the capacity and mandate for national-scale databases: Maybe this is not the right place for this comment but anyway... It feels like there are a lot of NGOs involved in this initiative which is great. However it is ultimately governments who have the mandate to develop and maintain inventories of WPs in their countries. They also have the resources (either themselves or through external funding) to collect data on a national scale which is infinitely more enlightening than a single-district exercise. How does WPDx plan to engage with governments which have done or are planning to do WP inventories and persuade them to (i) align their systems with the standards (ii) share their data?
  - Thanks so much Ian. Engagement with local/national governments has been a cornerstone of this work. We’re doing this by engaging governments directly as well as through partners that work closely with governments (i.e. UNICEF) and networks that government stakeholders are active in (i.e. AMCOW). I would definitely welcome any other suggestions that you have on this! Regarding persuading governments to align their systems we have worked the other way around. We are working to reduce the work that we need to persuade them to do. By including government monitoring frameworks in the desktop review the proposed standard closely aligns with what governments are already measuring. We expect to be able to integrate most government data sets (at least the required indicators) with little to no additional work needed on their part. Over time we will continue to engage government partners to get them even more fully aligned with WPDx. Regarding persuading governments to share their data it would be great to hear what ideas you or others might have. I do believe that as more data becomes transparent pressure will increase for others to get on the bandwagon. That said there are sure to be some governments that choose not to share data. Ultimately that's their choice. Even if 50% of governments choose not to share data we'll still have a 50% clearer picture of what is happening. In short continued engagement with governments will help to ensure that WPDx is being developed in a way that is as helpful to governments as possible and hopefully will also ensure their comfort in sharing data. Please feel free to share any other thoughts or suggestions. This remains one of the largest priorities for this effort.
  - **As noted above, continuous engagement with governments is a key priority for WPDx.**

## 5.3

- hosting: would this not always be hosted on WPDx itself to avoid broken links 1-2 years down the line?
  - Really good question Ian. What do other folks think? While the data in the exchange (i.e. the core attributes in this standard required and not required) would certainly be

hosted on our own servers what do folks think about the original data sources? How does this impact questions of ownership? One option could be that WPDx could host data if it didn't exist elsewhere. Or perhaps we could look at a local archive of all original data sets but still direct traffic to the original source unless the link is flagged as broken? I think this question also needs to be answered when considering linkages with other databases such as Akvo or mWater. Their "original" data will certainly remain on their servers and a copy of the relevant data (i.e. core indicators) will be copied to WPDx servers. Should this have any bearing on how we handle the occasional excel sheet from a specific WPM effort? This is a really good question that would be great to get folks input on! We will certainly continue to explore this further as well when we begin developing the scope for the online platform itself.

- I think a local archive sounds like the best option just so files don't get lost. I just think it would be a missed opportunity to not have the full dataset (e.g. in XLS or CSV) for that source being hosted on WPDx so people can easily see the additional variables collected which aren't in the standard. If I think of the 5-6 datasets we're using for current cross-country WPM analysis they were all shared as XLS files none available online! Ownership of the data (and right to request its removal) would still rest with the uploading agency?
  - I fully agree with Ian. WPDx should have a copy of everything in its system.
  - I agree with Ian and Jeroen. You could use web archiving software to keep an archive of the original site the data came from as well as any spreadsheet or database files that you download. In the case where the source is mWater we will have already done some data cleaning and processing to get it into our format and definitions so it would be ideal if we include a property linking to the original source. You could include both sources in your repository.
  - **A local archive of all data sources will be generated. If the raw data is publically available, that link will be provided. If the public link no longer works, or if there is no publically available copy of the raw data, a link to the WPDx archive of the file will be provided.**
- What is this?: Sorry I don't understand this. What if the dataset is not on the web? Most of ours is not...
    - Thanks Philippe. This is really good information. Would you (or others) be willing to have the whole dataset hosted if that was something WPDx could offer? Or would you prefer that just the relevant data be shared through the platform? One option could be to make this optional rather than required for those datasets that are not on the web. Let me know if that would address this issue!-Brian
    - Would it not be better if all the data was loaded on a site from which all could interact with? We would certainly be happy to send our data... Philippe
    - **A local archive of all data sources will be generated. If the raw data is publically available, that link will be provided. If the public link no longer works, or if there is no publically available copy of the raw data, a link to the WPDx archive of the file will be provided.**

## 6

- Allow for several pictures: Pictures from different angles or of different elements of the water point can be collected and be useful. Maybe just allow several URL to be written (and separated by commas) in the field.
  - **The standard now allows for multiple URLs for photos, separated by commas.**

### 6.1

- Bandwidth?: Photos take a lot of storage and bandwidth to transmit. We do not use them.
  - Thanks for this Philippe. I certainly appreciate that transmitting photos can be a massive drain on bandwidth and many organizations do not collect photos. It is for these reasons that photos are not a required indicator. For those organizations that do collect photos we have requested the URL rather than the file to help address the bandwidth constraints you have identified. Please let me know if this seems to address the point you have raised or if there are other related challenges that need to be addressed.
  - We are happy to host photos; it is part of our organizational mission. <http://api.mwater.co>
  - **Photos will not be required, and can be shared if they are available.**