Data for Water Security: Improving Water Data Access and Use

August 2020

This publication was produced for review by the United States Agency for International Development. It was prepared by Winrock International, in partnership with Tetra Tech, International Union for Conservation of Nature, Stockholm Environment Institute, and World Resources Institute

TABLE OF CONTENTS

1. INTRODUCTION

Basic data is critical to solving any water security problem as critical information can lead to practical solutions. Solutions to water security issues are not simple and require a thorough understanding of the past, current, and future situation. These solutions can be aided by data analytics where raw data is analyzed using tools such as statistical models, hydrologic models, or data visualization techniques to draw conclusions on status, data trends, or apply metrics for decision-making. Data analytics can help planners locate hotspots of dangerous pollutants, understand how structures can be positioned within or near a flood plain, how water can more equitably be allocated among diverse stakeholders in a water-scarce area. Furthermore to understand potential scenarios due to changes in climate, demographic characteristics, and land use. Certainly, data is highly useful for ensuring water security, but it requires that decision-makers also have the systems in place to collect, analyze, and set policy. This report introduces water-related data for solving different types of water security problems, considerations for data collection initiatives and an overview of the data value chain. The data value chain begins by identifying the information needs, followed by planning and data collection, transmission of data from the field, verification for consistency and errors, management of data for storage, and formatting for analysis and reporting. The data must be managed during this process and eventually communicated to key stakeholders. Throughout the data value chain, it is important to consider transparency and accessibility of data to a wide range of users. Data sets that are accessible via the internet are summarized to point stakeholders in the right direction to begin their process of gathering useful data to address their specific needs. This report also discusses why data is critical for water security, what data should be collected and how. Finally, global water data repositories are listed to help trigger the search for relevant data and a template for a data management plan is provided as well as examples of data collection done by SWP for water allocation planning. With these resources, stakeholders should be able to better understand how to successfully utilize data to help solve their water security issues.

Water data can be grouped into two main categories, quantity and quality, to address issues related to a sustainable supply, floods, droughts, or contamination of the water (Table 1). Water quantity is measured as streamflow rates (cubic meters per second), water abstraction rates (cubic meters per hour), precipitation amount (millimeters) and water levels (meters) in water bodies such as rivers, lakes, and aquifers. This type of data can be collected via devices such as rainfall and stream gauges, flow meters, and lake-, river-, or well-depth rulers. Water quality is generally measured as concentrations of pollutants or sediment (milligrams per liter or parts per million) in water bodies or other indicative measures of water body health relative to standards for human consumption or other uses. Such data is generally collected manually through water sampling or using specialized probes. Other auxiliary data supporting the two types of data categories are basin or river topography, location of water bodies and groundwater wells, river networks, soil properties, vegetation, and land use among others.

Open data encourages constructive scrutiny and encourages data use in innovative ways that benefits everyone. The USAID Open Data Policy (Annex 1) is included at the end of this report, along with a template (Annex 2) for a data management plan.

Table 1. Causes and effects of the three major water issues: Sustainable Water Supply, Floods, Droughts, Contamination.

Topic	Water quantity		Water quality
Basic issue	**Floods**	**Droughts**	**Contamination**
Causes	• Climate variability and change • Poor urban planning • Upstream deforestation • Land use change • Improper land use zoning • Ecosystem degradation	• Climate variability and change • Land use change/desertification • Upstream deforestation • Upstream over-extraction by different economic sectors • Poor urban planning • Ecosystem degradation	• Point-source pollution: wastewater disposal, mining runoff, industry, etc. • Non-point source pollution: agricultural runoff, runoff from urban, industrial and transportation areas • Intrusion of saltwater in coastal aquifers • Other groundwater contamination: infiltration and subsurface flow from polluted sites • Erosion and sedimentation from deforestation, agriculture, construction and mining sites, or land use changes • Ecosystem degradation
Effects	• Loss of life • Damages to property/infrastructure • Loss of agricultural lands and food scarcity • Ecosystem degradation • Water borne diseases • Loss of livelihoods • Human migration	• Lack of drinking water • Lack of water for agriculture • Lack of water for industry, including electricity generation • Lack of water for wildlife and living resources in water bodies • Loss of biodiversity • Dehydration and related diseases • Loss of human life • Loss of livelihoods • Loss of livestock • Human migration • Environmental flows not sustained	• Increase in human and animal diseases related to contaminated water • Human migration • Loss of livestock • Contamination of agricultural produce • Contamination of drinking water supply • Oxygen depletion and eutrophication • Fish die-offs • Loss of biodiversity

Topic	Water quantity		Water quality
Basic issue	**Floods**	**Droughts**	**Contamination**
Useful indicators and types of data to address issue	• Rainfall amounts: hourly, daily, seasonal • River and floodplain geometrical characteristics • Land use change trends • River flows water levels and flow rates: minute, hourly, daily • Types of settlements and population numbers • Flood plain areas	• Amount of water demand in watershed • Water abstraction rates from surface and groundwater • Historical rainfall amounts • Historical weather indicators: temperature, relative humidity, evapotranspiration, solar radiation, wind, etc. • Land use change trends • Aquifer levels and presence of groundwater • Types of settlements and population numbers • Industrial water uses and allowances • New upstream water storage infrastructure	• Type, location, and pollutant discharges from point sources in watershed • Fertilizer and pesticide applications • Land use change trends • Human settlements • Types and location of industry • Water quality indicators in water bodies: temperature, sediment concentration, water conductivity, salinity, turbidity, pH, ammonia, nitrite, nitrate, phosphorus, , dissolved oxygen , fecal coliform amounts, etc. • Aquifer and/or river levels of site-specific pollution threats, such as saltwater intrusion or chemicals leaked from nearby facilities

2. THE IMPORTANCE OF DATA FOR WATER SECURITY

When rivers, reservoirs and wells go dry, or flood decimates the infrastructure on the floodplains of a river, a water-borne illness sweeps through a municipal drinking supply, or a drought causes mass fish or livestock die-offs, and loss of life and livelihoods, calls for action are often heard to ensure such problems do not happen again. However, solutions to these issues are neither simple nor quick. They always begin with a thorough understanding of the problem and the water resources in the surrounding or upstream areas.

Gaining this understanding requires data. Data allows stakeholders to not only understand why a water security issue occurred, but how it could potentially be mitigated, solved, and prevented. Good and long periods of records of meteorological and hydrometric data lead to answers to questions such as how often a disastrous flood is likely to occur in an area; what is the impact of future land use changes, climate variability and change; what is the source of the pathogen that caused the water-borne illness; or what is the geographical extent of the severe drought. Data is also the key to forming early warning systems which allow local residents and communities to prepare for extreme events. Furthermore, with technological innovation, data is constantly becoming cheaper and of higher quality.

Water data is therefore important not only to managers who must guarantee the water security needs of their populations, but also to communities and businesses that must ensure consistent water supply or protect key infrastructure. It is also essential for water users at large to make better water use decisions and to adopt improved water practices. Including data in water security planning allows stakeholders to do the following:

- Identify hotspots and opportunities to intervene and take corrective actions to prevent or mitigate water security problems;
- Monitor the outcomes of interventions by collecting and analyzing data, leading to successful project management and anticipation of future needs; and
- Perform ongoing monitoring and analysis of historical and current data to help design more successful projects in the future, tailored to current conditions and needs.

Despite all its advantages, data can be complex and difficult to collect, manage, analyze, and report on. Poor-quality or missing data can lead to improper conclusions and poor decision-making. Unclear communication of data analysis results can cause confusion among stakeholders.

Data helps us understand water issues by identifying problems and their causes, measure their effects, and monitor the implementation of projects and programs. In order to achieve these benefits, data collection must be aligned to the desired outcomes, then analyzed and presented in a format that leads to clear understanding and good decision-making. One tool used to help interpret data for decision-makers is the use of targets and indicators that are aligned to the objectives of a project or policy.

3. WHAT WATER DATA TO COLLECT

A recurring mistake is collecting too much data, not collecting easy data such as precipitation, spending too much time and resources on collection, and not having the time and resources to properly analyze and use the results to improve decision-making. "It would be nice to know" is not justification enough, especially in developing countries where data collection can be a low priority and securing durable funding and resources for it is a daunting task.

Collecting data has cost implications. It can be tempting to design an extremely detailed data plan with many indicators, even if some are not specifically applicable to the problems at hand. This is beneficial in one sense because data needs may arise after design; however, it can create unneeded costs for tight monitoring budgets. The following are some recommendations for designing monitoring plans that are cost-effective:

- Define and qualify the geographic area of interest, and contextualize it with metadata on geomorphology and land uses.
- Define what problems are being targeted, and what types of solutions will be considered or are being implemented.
- Create a logical framework to ensure that each piece of data proposed to be collected is connected to specific outcomes and goals. If there is not a clear link between the collected data and how/to what end it will be used, that specific data should probably not be collected.
- Research existing data as well as monitoring activities in the area of interest. It is very possible that information already exists or that some key indicators are already being collected by government surveys, NGOs, research institutions, or private companies. Even if existing information or monitoring does not exactly match what is needed, it may be more cost-effective to enhance an existing effort than to create another entirely separate data activity.
- Consider using mobile technologies which can allow rural populations to participate in monitoring and reduce the need to send technicians to the field, which can drive up costs. Remotely sensed data (from satellite images) can be used, too; however, they require specialized software and skills to accurately analyze.

Given the costs associated with data collection, the complexities of data coming from many different sources and in varying formats, and the reality that data sets will always be at least partially incomplete, there may be a need to look to data analytics and big data techniques to address some of these gaps and challenges. Data analytics and big data techniques offer the potential to improve the quality of data sets by patching gaps in data, supporting interoperability of data in varying formats and in so doing saving time and costs related to data collection. These analytics also drive the ultimate decision-support processing that needs to be linked to data sets. Big data analytics also presents the potential of integrating a wider range of data sources (including those in handwritten format and citizen data) that were previously not accessible either because of perceived poor data quality or the time implications of manually having to process such data.

Below is a brief description of the main types of water data, with specific data sources available online listed in Section 7.

3.1. WATER QUANTITY DATA

Water quantity data are best measured using a permanent monitoring gauge that is placed in the body of water, such as a stream gauge. This allows for "continuous" monitoring, in which a measurement is taken once per time step, such as every minute, every hour, or every day. Since water volumes, flows, and depths change quite quickly, the permanent gauge allows for more numerous and consistent monitoring.

- **River or stream discharge**: If a water-related issue deals with water quantity one important piece of information is discharge or flow. Understanding historical flows is critically important to see how these flows may be changing due to external threats or understanding how an extreme event (such as a flood or drought) compares to extreme events in the historical records. Measurements of flows at the basin outlet can also help estimate groundwater recharge in the basin considering the hydrologic cycle. Discharge is generally reported as a volume of water over time (cubic meters or feet per second), and usually collected via stream gauges that are constructed in a river's channel.
- **Lake or other water body levels**: The equivalent to river discharge for water quantity issues faced where the main water supply is a lake or other unmoving water body, historical data on lake levels help to understand how much water is generally in the lake over the course of time. This allows for analysis of how the historical "regime" is changing over time due to external factors. Lake levels are generally reported as a depth below or above average levels, and sometimes as volumes. Lake sediments can also contain historical reconstructive evidence of levels in the past.
- **Precipitation depth**: Understanding how much precipitation (in the form of rain or snow) an area receives is also critical to understanding water quantity issues. Given that precipitation is the main source of water input in most watersheds, it is critically important to understand how much water an area naturally receives in order to make sustainable water decisions about how that water is used. Precipitation is generally recorded through rain gauges as a depth (millimeters) over a given time.
- **Groundwater levels**: Besides surface water in lakes and streams, groundwater is the other critical source of freshwater, and therefore being able to measure it is critical to understanding water quantity issues. Measurement of groundwater levels over time permits the analysis of whether or not levels are dropping (signifying unsustainable extraction) or increasing (signifying recharge) over the long term. These measurements are generally recorded by measuring the depth of wells at equal time increments over time. Measurement of groundwater levels also help estimate aquifer storage and change over time.

- **Water abstraction volume**: Water abstractions or water use refer to water being removed from its natural source, such as a river, lake, or aquifer, for human use. These volumes per unit of time (cubic meters per hour) are important to understand who is using an area's water and in what quantity. Abstraction volumes are generally measured in the same way as natural flow, using stream gauges or pipe meters. However, given that abstractions are often measured near engineered structures such as pipes, dams, weirs, and wells, it is often easier to install gauges, since they work best in engineered environments with predictable physical properties.
- **Other meteorological data**: In addition to precipitation, other meteorological variables are important as input to rainfall runoff models, including temperatures, relative humidity, solar radiation, wind, and pan evaporation.

The above are data normally collected for different water security assessments; however, each particular water security assessment requires different type of data in time and space to properly analyze for decision-making. Table 2 shows a summary of specific data needs for different type of assessments. Others may include data needs for infrastructure design such as reservoirs, flood protection barriers, and run-of-the-river hydro plants, among others.

Table 2. Data Useful for Water Security Assessments

Water Allocation Planning (WAP)	Flooding- Floodplain Delineation	Water Resources and Water Quality Assessments and Trends
Water Resources Availability • Topographic data or digital elevation models (DEM) for delineation of basins and sub-basins • River networks • Daily Precipitation and river flows from hydrometeorological and hydrometric stations • Hydrogeology and soil properties • Land use **Water abstraction surveys** • Abstractions by source type (spring, river, rainwater harvesting, dug well, water pan, borehole) **Water Demand Analysis** • Population by district/basin • Livestock population by type • Small-scale irrigation demand • Large-scale irrigation demand • Tourism water demand • Wildlife water demand by specie **Environmental Flows** • River flows • Water quality assessments • Geomorphology • Fish habitats and macroinvertebrates • Riparian vegetation **Climate Models**	• Topographic data or digital elevation models (DEM) for delineation of basins and sub-basins • Topographic data of the river and floodplain (bathymetries) • Digital terrain models • Time series of river flows and stages at different river locations • Land use and floodplain land cover and vegetation • Historical storms • Historical maximum annual precipitation depths with durations from 5 mins to a few days • Intensity-duration-frequency curves • Input hydrographs • Archive of sediment deposits in the stream channel cross-section for historical flood peaks estimates	• Topographic data or digital elevation models (DEM) for delineation of basins and sub-basins • River networks • Daily Precipitation • Daily river flows from hydrometric stations • Measured evaporation • Temperature • Relative humidity • Wind • Solar radiation • Cloud coverage • Hydrogeology and soil properties • Type of vegetation (land use) • River sediment concentration measurements • Time and quantity of application of agrochemicals **Water abstractions** • Water concessions and permits by type of abstraction (spring, river, dug well, borehole) **Climate Models** • Projections of temperature and precipitation changes for the planning horizon • Carbon Dioxide Concentrations

Water Allocation Planning (WAP)	Flooding- Floodplain Delineation	Water Resources and Water Quality Assessments and Trends
• Projections of temperature and precipitation changes for the planning horizon		
Main Outputs		
• Mean monthly and annual river flows, and flow duration estimates by river segments • Volume of water abstracted by source, and sub-basin • Total water demand by type and sub-basin • Environmental flow and river depth requirements for dry and wet periods by sub-basin • Guidelines for water allocation by water user type under wet and dry conditions for different planning horizons (5yr,10yr,20yr)	• Water depths and flows in floodplains for different frequencies such as 50yr,100yr, 500yr	• River flows at different river locations • Water and sediment yields at the sub-basin outlet • Total amounts of nutrients and bacteria transported by runoff • Concentrations of nutrients, pesticides or other pollutants in reservoirs or ponds • River flows at different river locations under different climate scenarios

Example of Models		
Water Allocation Planning (WAP)	**Flooding- Floodplain Delineation**	**Water Resources and Water Quality Assessments**
WEAP-Water Evaluation Planning Tool https://www.weap21.org/	SWMM-Storm Water Management Model https://www.epa.gov/water-research/storm-water-management-model-swmm HEC-RAS, Hydrologic Engineering Center's River Analysis System https://www.hec.usace.army.mil/software/hec-ras/ MIKE, https://www.mikepoweredbydhi.com/	SWAT-Soil Water Assessment Tool – https://swat.tamu.edu/ PRSM- Precipitation Runoff Modeling System- https://www.usgs.gov/software/precipitation-runoff-modeling-system-prms

3.2. WATER QUALITY DATA

Given the wide variety of pollutants found in water, there are a diverse set of indicators of water quality that can be collected. Many of these indicators test only for a specific pollutant, however some can be used to give a measure of general water quality. Here, key indicators are listed that are used to assess general water quality and, also those used to measure if water is sufficiently clean to be used as a potable water source. For drinking water quality, a comprehensive list of parameters can be found in WHO (2017)[1]. In contrast to most water quantity data that is generally recorded automatically over time, water quality data are generally measured by manually taking a sample of water and testing it.

- **Surface and groundwater quality key parameters**:
  - pH: a measure of water acidity; a change from normal levels may indicate a new pollutant has been introduced.

[1] World Health Organization (WHO) Guidelines for Drinking-water Quality, Fourth Edition (2017)

- Dissolved oxygen: a good indicator of the health of the organisms living in the water and therefore an indicator of overall water quality.
    - Fecal indicators: may indicate that wastewater has contaminated the water, which can bring with it many disease-causing pollutants.
    - Nitrate, nitrite, ammonia, and phosphate: indicators of possible pollution by fertilizer runoff from agriculture or urban areas.
    - Heavy metals and hydrocarbons: indicate contamination from specific industrial activities or urban runoff.
    - Water temperature: a change in normal water body temperatures may indicate a new pollutant has been introduced.
- **Drinking water quality key parameters**:
    - Indicator organisms to identify fecal pollution such as total coliforms and Escherichia coli (E. coli).
    - Dissolved solids: a measure of drinking water quality in general; water with more solids is likely to be more contaminated.
    - Arsenic: a pollutant harmful to human health at high concentrations; can be found naturally or from industrial/agricultural activities.
    - Fluoride: can be harmful to human health at high concentrations, despite its use as a decontaminant in water supplies.
    - Chlorine residual: can cause chronic problems to humans; found naturally in low concentrations but also from industrial runoff.
    - Iron: may be introduced into water supplies by corroding metal, industrial processes, or mining.
    - Chemicals including heavy metals and pesticides: A variety of heavy metals and pesticides can be harmful to human health even at low concentrations, often introduced by agriculture, mining, or industrial processes.
- **Measures of acceptable contaminant levels**: Understanding the concentrations at which contaminants become harmful is critical to evaluating water quality. Different regulators around the world have set different acceptable levels for their own jurisdictions, and these levels vary by water use and by contaminant. Locating and understanding both the scientifically established and regulated accepted levels is key to navigating water quality issues from both a health and legal standpoint. The World Health Organization's Guidelines for Drinking-water Quality (2017) are a good place to start when deciding what contaminant levels are safe for potable water sources.

3.3. AUXILIARY WATER DATA

Besides specific water quantity and quality data, there exist other more general data sets that can be useful to both water quantity and quality issues.

- **Base maps on the natural landscape**: spatial data on waterbodies, elevation, land cover, and soil types, among others, can be extremely useful in evaluating water issues. Elevation data sets can be used to better understand flow patterns and delineate watersheds. Land cover and land cover change maps can indicate where areas in the watershed are changing and potentially impacting water quantity and quality. Location and soil horizon properties (thickness, texture, organic carbon content) are very important for assessing water infiltration, subsurface runoff, groundwater recharge, and soil erosion.
- **Base maps on infrastructure**: spatial information on the location of buildings, roads, population, administrative boundaries, pipelines, extraction points, and point and nonpoint source polluters are very useful for water quality issues and to a lesser extent water quantity

issues, as they allow for identification of potential sources of pollutants and better characterization of water demand.

- **Water user surveys:** Surveys are important for understanding the perception of water and how the population interacts with their water. Therefore, critical data can be gleaned from surveys that can be used in decision-making. Perceived notions of an area's water quality, the perception of the utilities providing water and the willingness and ability (household incomes) to pay for water are examples of information that can be obtained via surveys.

4. DATA VALUE CHAIN

Once it is clear that data must be collected to address a water security issue, there is a process that occurs before the data is actually used in a meaningful way for this goal following the steps of the data value chain (Figure 1). The pertinent data indicators must be decided upon, the data collection process must be designed and implemented, the collected data must be managed and stored, and data analysis must be performed. Finally, the data analysis results must be reported, communicated, and scrutinized by stakeholders. These steps of the data value chain are described below.

Figure 1. Data Value Chain

4.1. DEFINE DATA NEEDS

Any data collection effort begins with defining what data to collect but this process often moves forward without proper consideration for how that data will inform decision-making. An indicator is anything that describes the state or level of something and is used to link measurements to desired outcomes. Often indicators are numbers that are used to measure the change in something over time; however, indicators can also describe whether a process is in place or something has been done. Usually managers will set a target for an indicator, which should be related to a desired outcome or an activity that should be taking place. One approach for establishing good indicators is to apply the SMART criteria, meaning the indicator is **S**pecific, **M**easurable, **A**greed upon, **R**ealistic, and **T**ime-related. Donors such as USAID, UN agencies responsible for the Sustainable Development Goals, the World Bank, and others have standard or key indicators for specific themes, water risks or water actors. Many indicators have recommended questionnaires, calculation methods, and reporting formats already defined, making it easier to use them. Also, careful consideration must be given to whether a data stream once initiated has the institutional support and funding to be sustained.

4.2. DATA COLLECTION

There is a wealth of global, national, and subnational water data freely available on the web, via research institutes, foundations, donors, international organizations, and others. For many basic water analyses, finding and accessing these data sources will provide a significant base layer of data for supporting decision-making. Section 7 contains a list of some of the most commonly used water data sets available on the internet.

However, for many decisions about water security, the required data is highly local and must be collected more frequently, such as local water levels or flow measurements. In this case, broader data sets must be supported by primary data collection. Collecting data manually can be time consuming and inaccurate, but there are many tools available for data collectors.

Smart devices like mobiles and tablets can enhance data collection by allowing for rapid, broadly dispersed data collection by teams of collectors using a common platform. This information is supplemented by the sensors available on most of these smart devices: images, geolocation, barometric pressure, and others can all be directly accessed by an app. Furthermore, a wide variety of third-party tools can connect directly to a smart device and transmit data to a central location. If smartphones are not available, reaching a broad number of data providers, such as observers or water users, can be accomplished using standard mobile "dumb" phones by using text messaging and voice calls to collect information about observable phenomena like rainfall, water levels, water use patterns, etc.

Recently, there has been a revolution in what has been called the maker movement. Inexpensive computer chips and open source software means that almost anyone can build custom electronics to accomplish any task. This movement has also opened up the doors to the construction of low-cost, customizable sensors that can remotely collect data for months at a time. Sensors for water level, flow, rainfall, temperature, and chemicals can all be deployed at relatively low costs. These can be custom-built or in some cases commercially provided. In any case, public and nongovernmental institutions must play a role in ensuring that data is collected in a consistent way so that it is comparable across locations and collection activities.

4.3. DATA TRANSMISSION

Though often overlooked, how data moves from one system to another is very important, especially when connectivity is an issue. A thorough understanding of the local environment will help decide the most efficient way to send collected data to a central database.

The first major factor is how the data is collected. It may be on a tablet, mobile phone, sensor device, or just written down in a notebook. From there, it can be transmitted by internet, via text message, manually entered into a computer, spoken over the phone, carried by USB drive, or any number of other feasible ways. Technology can help speed up and scale data collection and transmission, especially for data collected frequently in remote locations. The most important factor is that the method is as reliable as possible. It is generally helpful if a system is delay-tolerant; that is, if there is no transmission method available (i.e. no Wi-Fi), data can be saved and sent when ready. Redundancy is also preferred: if mobile phone service is out, for instance, it should be possible to pull data from a device and send it manually via another method.

4.4. DATA VERIFICATION

Ensuring high data quality is a critical step of any data value chain. A mountain of data can be collected, but if it is not of good quality it is largely useless. Decisions that are made with poor-quality data can be dangerous and lead to erroneous conclusions. Data verification refers to the process of ensuring the quality of data, which includes components such as completeness, consistency, and understandability of the data. For instance, for analysis of consistency of precipitation data from one station with questionable or missing data, a methodology called double-mass analysis can be used by comparing data for a single station to the composed pattern of the data from several other stations within the basin. This comparison allows to check for consistency and adjust values of the station in question. Usually both quality assurance and quality control (QA/QC) are included in the data verification process. Quality assurance involves ensuring quality of the data while it is being collected, while quality control refers to inspecting the data after collection.

The QA process begins with creation of standard operating procedures (SOPs) for each specific data collection process. Examples could include SOPs for installing and maintaining rain gauges, field missions to collect water quality samples, or carrying out a water use survey. Collectors should be trained on these SOPs before data collection begins. SOPs are a great way to ensure consistency when multiple people will be collecting data. Standard forms (which are often part of the SOPs) should be used so all collectors are collecting the same data in the same format which will facilitate of data transmission. USAID's Water Quality Assurance Plan (WQAP) guidance also provides templates with standard operating procedures for field data collection (https://usaidgems.org/wqap.htm). Some additional QA best practices include the following:

- Equipment should be properly calibrated from time to time (according to the equipment manual) to ensure consistent accuracy.
- If measurements are being recorded by a different person than the appointed data recorder, readings should be repeated by the data recorder to ensure they were heard correctly.
- The names of the data collection team should all be recorded along with serial numbers of equipment being used, in case questions arise later about the data.
- "Hot checks" should be performed, which involves a team leader or second collector observing the data collector to ensure they're following SOPs.

The QC process begins after data collection is complete. Data sheets should be reviewed to check for completeness and accuracy. Generally, it is good practice to check at least 10 percent of collected data for completeness and accuracy. It is important that the reviewer understand the measurements being taken so they have a general idea of the range of values to be expecting. This way, they can catch illogical values that may result from an erroneous unit conversion, typo, or other human error. In some cases, remeasurement can be done to see how original measurements compare—wide variation in measurements could indicate error.

4.5. DATA MANAGEMENT

As data is collected, it is important to think about how and where it is stored. To this end, a data management plan (DMP) can help data managers think through the variables that come into play. A DMP covers everything from storage formats and databases to metadata, ownership, and privacy and security considerations. A research-oriented DMP might be very complex, but a simple one would seek to answer:

- Who is collecting data and how is it transmitted/submitted?
- Where is data stored long-term? Are there costs or risks? Is it backed up?
- What kind of metadata do you need to keep to make this data useful to others? Metadata is additional information about the data set, including how it was collected, what standards or methods were used, and any limitations or gaps.
- How will the data be updated and linked to other data?
- Who has access to the data? How is/will it be shared? Are there privacy considerations?
- What financial or human resources will be needed to maintain the data system?

In many cases, data will be stored locally, on someone's laptop, smartphone, or a small server. This is okay, so long as it can be shared without too much difficulty and clear guidelines are established for how long data should be kept, as well as what should be done to archive both the electronic versions and raw data, such as field notes. Other options might include a low-cost cloud server or service, which ranges from free to a few dollars a month to maintain. This server, owned and maintained by a third party, could be a standalone database or come packaged with some value-add components, like data collection or analytical tools. Some examples are provided in Table 3.

Table 3. Tools for data management

Type	Tool	Description
Spreadsheet	Excel, Google Sheets, LibreOffice Calc	Located on a hard drive or on the internet, this is the simplest of data storage tools. Data entry is usually manual, as is sharing the data
Relational Database	MySQL, PostgreSQL, MongoDB	These are powerful data storage and analysis tools, but they are difficult to set up and use without expert help and other software
Specialty Software	mWater, Open Data Kit	Data collection software like Open Data Kit has its own database built in but requires technical expertise to operate; complete water data solutions like mWater also come with integrated databases

4.6. DATA ANALYSIS

Data analysis is the most specialized component of the value chain, because it requires a certain level of numeracy and comfort with basic statistics, and familiarity and understanding of underlying assumptions in water models. There are a large number of available analysis platforms that are appropriate for all levels of skill, from simple turnkey graphs to complex modeling software. The best software is going to be the one that is usable by its intended audience and capable of producing the appropriate level of analysis.

When choosing analytical tools, the most important factors are the user's and the sophistication of the desired output. For simple analysis to track trends in a time series, Excel or even pen and paper would be sufficient. For more complex decision-making needs, a deeper analysis of precise needs is necessary.

Table 4. Decision Support Tools

Type	Tool	Description
Spreadsheet	Excel, Google Sheets, LibreOffice Calc	Can do basic analysis with summary tables, graphs, and charts. Widely available and easy to use for some
Statistical analysis	R, SPSS	Both are advanced platforms for statistical analysis. Probably more sophisticated than necessary for most users, but accessible enough for an educated user
Modeling or decision support software	Water Models like the Water Evaluation And Planning (WEAP), Soil Water Assessment Tool (SWAT), and other Hydrologic and Hydraulic models (HEC-RAS, MIKE)	Generally used only by professionals and academics. Unlikely to be used by most people, but could be of use to a local hydrologist or engineer

Type	Tool	Description
Web-based platforms	IBM Watson Analytics, Klipfolio, Pentaho	Catchall term for the huge number of online analytical tools available to a savvy web user. Anything from interactive charts to graphs is possible. Can be shared easily with collaborators, but always requires an internet connection.
GIS	ArcGIS, QGIS	Desktop GIS software is generally left to specialists, but is an indispensable part of analyzing geographic features like river basins or water points
Specialty Software	ODK, Tableau, PowerBI, mWater	As above, ODK and other data collection platforms often come with basic analytics, which may be sufficient for general use, like basic time series data and simple maps. Similarly, analysis tools like Tableau have advanced analysis features

4.7. DATA REPORTING AND COMMUNICATION

Reporting on analysis is the act of making it available to other audiences including the public in an easily consumed format. Generally, this is a combination of making the data itself available, as well as a summary report conveying the appropriate message. This could be as simple as a single graph for a single decision-maker, or as complex as many interactive charts or maps intended for external groups. The exact details will depend on the audience as well as the message(s) conveyed.

Though a chart of numbers conveys some information, it takes better visualization to be able to draw conclusions quickly and accurately. Turning raw data into a chart, map, or graphic tells a story and builds consensus around that story among the intended audience. There are numerous platforms to turn analytical data into data visualizations for easy reading (Table 5).
Publishing data can be as simple as emailing or printing a report. For a broader audience or more complex data, data can also be shared via mobile apps or web dashboards. The complexity and cost may scale with the ambition of the reporting, depending on the tools chosen.

Table 5. Data Communication Tools

Type	Software Tool	Comment
PDF	Adobe Acrobat, FoxIt	Easily create a page that can be emailed or posted online
GIS Maps	ArcGIS Online, CartoDB, MapBox	Dynamic maps accessible online can be very powerful for viewing water data
Specialty Software	Tableau, Klipfolio, PowerBI, mWater	Dashboards with multiple data visualizations can present complex data in an easy-to-read format

Communicating:
Communicating the results of data collection and analysis to stakeholders is the critical last step in the data value chain. This step is the culmination of the entire process, where results are disseminated to those who will benefit the most from the conclusions. This communication involves organizing and

presenting the graphs, charts, and tables constructed in the Data Reporting section into a reporting format. The exact information that should be communicated varies by the type of water security issue being addressed (Table). The key to stakeholder communication is presenting the data in a clear, succinct and understandable manner. Too often communications are overly technical, since the data analysis is complex and normally performed by an expert. Experts may benefit from a communication expert or professional to distill the information in a way that is more understandable by the general public. Photos and diagrams are helpful to contextualize the results and explain them in a more intuitive manner. Public forums can also help disseminate the results to less-literate populations, especially in areas where internet access is limited, or education is basic.

Several data reporting formats exists. The following are some examples

- **Bulletin/one- or two-page "report card":** overview of water body health, especially water quality, usually with different indicators graded A-F.
    - *Examples*: Anacostia River Report Card, West Rhode River Report Card
    - *Benefits*: very brief, quick and easy to digest, very good for stakeholders with limited technical understanding.
    - *Drawbacks*: does not contain much information, data has to be modified to fit into easy-to-understand language and subjective "grades" without much explanation.
- **State of the river reports**: often long, 20-to-100-page documents that detail the state in which a water body is in, often explaining various indicators in depth and showing the data related to each indicator.
    - *Examples*: Mississippi River State of the River Report, Columbia River Basin: State of the River Report
    - *Benefits*: lots of information with in-depth description of each indicator so data is clearly and thoroughly presented, great for technical stakeholders.
    - *Drawbacks*: very long so hard to quickly digest, not good for stakeholders who need quick and easy information.
- **Status of aquifers**: includes aquifer conditions such level changes and water quality.
    - *Examples*: Status of the Aquifer report such as Tucson Water
- **Online interactive site**: contains videos and infographics showing water body health in an informative visual way.
    - *Examples*: Orinoco River Basin Report Card, Potomac Report Card
    - *Benefits*: potential for very good interactive graphics for increased understanding of issues, ability for stakeholders to quickly view in-depth information that is important to them while ignoring other sections they are not interested in.
    - *Drawbacks:* potentially costly to produce (need web developers), may not function well in low-bandwidth areas of the world.

Table 6. Key information to communicate to stakeholders according to water security issue.

<table>
<tr><th>Topic</th><th colspan="2">Water quantity</th><th>Water quality</th></tr>
<tr><th>Basic issue</th><th>Flooding</th><th>Droughts</th><th>Contamination</th></tr>
<tr><td rowspan="2">Key information to communicate to stakeholders</td><td colspan="2">• Recent rainfall (monthly average), correlated to historical average rainfall for the same month
• Predicted upcoming precipitation in the area/in the watershed
• Water reservoir (dams, etc.) levels and year trend
• Average daily river flows
• Average water use by relevant sectors (drinking water, irrigation, industry)
• Contact details of person/place where more information can be accessed</td><td rowspan="2">• Water pollution "heat map" showing major sources of pollution in a watershed
• Current levels of key indicators - turbidity, nutrient concentration (N, P), dissolved oxygen, coliforms, etc.
• Overall water quality index with ranges (good, poor, bad) based on individual indicator levels (low, medium, high) Important thresholds of these levels for human or ecological health
• Relevant water treatment plant alerts
• Relevant hypoxia/eutrophication maps showing size and intensity
• Basic reminders on how to improve water quality for human consumption (boiling)</td></tr>
<tr><td>• Maps of areas under risk of flooding
• Map or areas with recent land use change that could increase runoff risks (e.g. clearing of forests, large construction sites or new impervious surfaces)
• Relevant infrastructure alerts (state and conditions of floodgates, etc.)
• Current well levels</td><td>• Maps of areas under risk of severe drought
• Recent temperature (monthly average), and historical average temperature for the same month
• Prevision of upcoming water availability for households and relevant sectors (schedule of cuts) based on current/updated demand</td></tr>
</table>

5. CHOOSING DIGITAL TOOLS

Choosing the right digital tools for the intended purpose is a crucial first step for dealing with data, whether they are hardware or software, or both. Whether it is for collecting, managing, or analyzing data, there is a dizzying array of standalone tools and integrated platforms available at different prices and with a broad range of capabilities. The first critical factor is, of course, whether the tool performs the required job.

- Needs: What is the exact task that the user must accomplish? What are the inputs and outputs?
- Capabilities: What is the education level of the user? How much experience with computers/software do they have?
- Environment: What kind of equipment do they have access to (electricity, connectivity, money)?

The Digital Principles

The Principles for Digital Development are a series of principles that were collaboratively written by international development practitioners and endorsed by a key set of donors and implementers. There are nine principles that broadly lay out best practices in using information technology for development. For these purposes, several of the principles are particularly relevant:

Design with the User: A user's specific needs and capabilities dictate the functions of the software that they should use.

Understand the Existing Ecosystem: Take advantage of tools that are already in common use and that users are familiar with.

Design for Scale: The tools and processes should be replicable and adaptable to other contexts.

Build for Sustainability: Beyond usability, software should also be affordable and maintainable. Often, open source software is free, but harder to maintain, while commercial tools are well-maintained but expensive.

Use Open Standards, Open Data, Open Source, and Open Innovation: Using and contributing to open source software and open data will benefit the entire water sector.

6. OPEN DATA AND STANDARDS

Too often water-related data is controlled by government entities, private corporations, or individuals that do not share the data with stakeholders, who could greatly benefit from the data. Other times, data is available only for a fee or by insider knowledge. This limits the impact data can have in helping to solve water security issues and remains for the benefit of a select few, who have collected or obtained the data.

The idea of "open" data refers to data being publicly available for free, for any use, without restrictions. Open data policies allow anyone to easily access data, which encourages innovative analysis and reporting along with scrutiny, which will ultimately improve the quality of the data sets. Many governments are moving toward open data policies (see Annex I for the USAID Open Data Policy) which allow public access to data that is collected and maintained using taxpayer funds.

Generally, open data repositories should have the following characteristics:

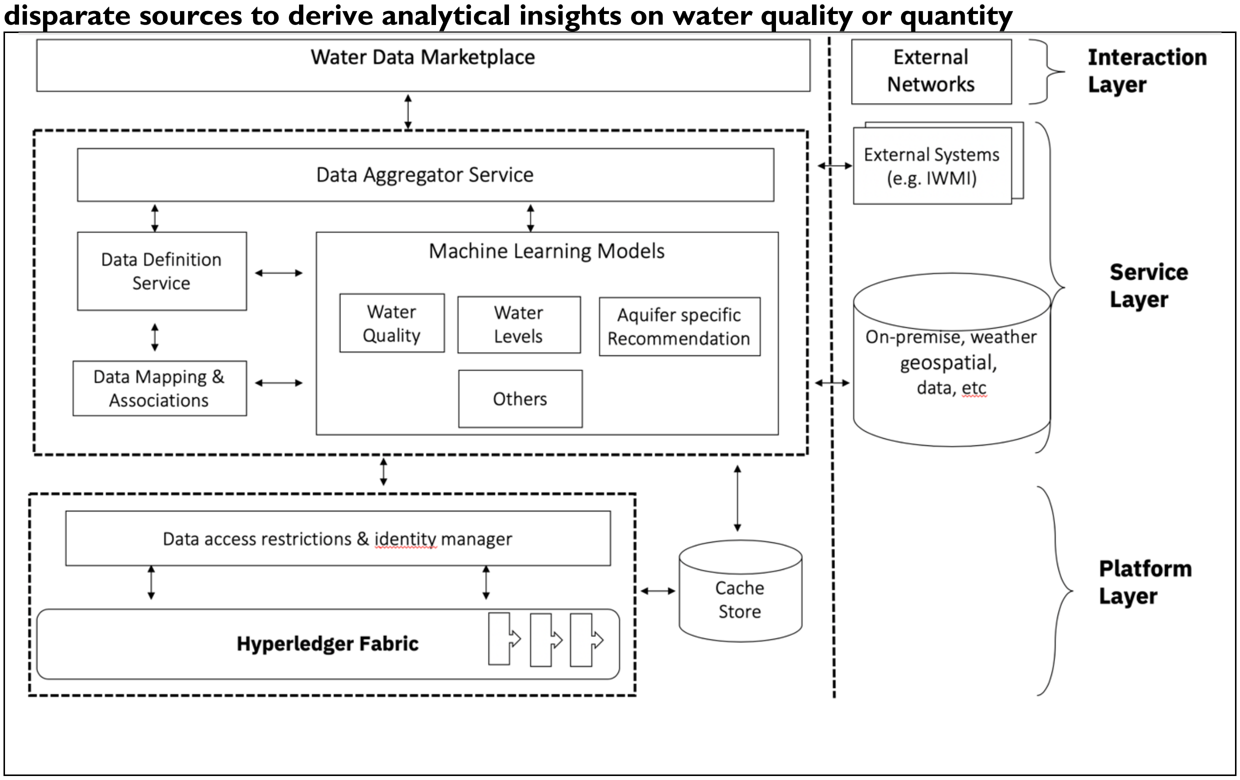
- The steps in the data value chain follow specific standards which are subject to a stakeholder review process and are publicly available.
- Data that is produced includes metadata, which explains when and how the data was created and by whom, along with a detailed description of the data attributes.
- The data is updated periodically according to a transparent schedule.

Open data initiatives make it possible to derive new insights from aggregated data coming from various disparate sources without the need of creating data lakes which can be costly in terms of time and financial resources. An important consideration is that this is facilitated in an online environment.

Figure 2 below provides an example of a water data analytics service analytic model where data aggregator services access the data needed for various modelling from the different distributed on-premise databases through applications programming interfaces (APIs). Though APIs, data owners are allowed to explicitly define data-related access rules and permissions which include a de-identification mechanism and maximum cache time that will be used to sync the data or to possibly avoid a future syncing process. A data definition service is used to clearly outline the structure of different data stored by all the stakeholders storing the data. The defined data structure will be useful in coming up with a data marketplace for water-related machine learning models for meaningful water insights (e.g. water quality and managing water levels). This means that the API on an application that runs analytics on aggregated data needs to be able to access the APIs of the various stakeholders from which it can identify the relevant data structures. The aggregation mechanisms include the use of data syncing processors that will be executed at different times to access and cache the data in the analytics service. Data mapping association is used to provide data provenance information which includes the synced data, identification of models that are using the data, users who are allowed to access the data with their explicit permissions, descriptive data definition, and privacy-related requirements.

Recent technological innovations have made it possible to enforce these associations by leveraging on blockchain technology, ensuring that all the water stakeholders have a trusted data marketplace where trust and data privacy is always enforced.

Figure 2. An illustration of an application environment for aggregating data from disparate sources to derive analytical insights on water quality or quantity



7. DATA RESOURCES

Global Water Data sets

Open data initiatives by government, academia, and the private sector have produced a large number of data sets of existing water data. Deciding which data sets can be useful to a particular project is challenging, because the data varies in quality and completeness. The tables on the following pages of this section resulted from a careful review of currently available water data sets. Each one is given an overall usefulness rating accompanied by details about the data hosted. The terms used in the table are explained below.

- **Type of data** – what measurements went into this data set and how was it collected? For example, time series are collected at the same places at regular time intervals.
- **Geographic coverage** – the areas of the earth covered by the data. Keep in mind that even global data sets still have some missing areas; for example, clouds can prevent satellites from seeing certain regions.
- **Geographic resolution** – the smallest area that has its own unique value in a data set. Sometimes this is given in numbers, such as 1 km, which means that the smallest area is a one-by-one-kilometer square, or grid. In other data sets the resolution is given as regions, such as country or district. Geographic resolution is important because data covering a large area is an average that can hide local differences.
- **Time period** – the range of dates for which data exists.
- **Time resolution** – the smallest step in time which has a unique data value. For example, a time resolution of one month means that the measurements are averaged or totaled (as in rainfall) over each month, resulting in a value for every month of the year.
- **Projection into future** – some data sets include predictions of what the values will be in the future, according to a model. The quality of these predictions depends on the uncertainty in the model and the data used to develop it.
- **Accuracy/source** – judgement of the likely quality of the data set, informed by the source of the data and the reputation of those who produced it.

1. General water resources

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ source	Notes
EU Global Surface Water Explorer	https://global-surface-water.appspot.com/	Water availability (watersheds and dams), location and temporal distribution of water surfaces over past 35yr, and statistics on their extent of change.	Global	Landsat (same as USGS Global Land Cover)	Data form 1984-2015	Annual	N	High/European Commission	high resolution and long-term changes of surface water.
FAO AQUASTAT	http://www.fao.org/nr/water/aquastat/data/query/index.html?lang=en	Large UN data set on geography (land use, population, etc.), water resources (precipitation, renewable water resources), water use (withdrawal by source and sector, pressure on water resources), irrigation and drainage, conservation and water harvesting, and Environmental health.	Global	Country results, with option of regional values	Variable - provides yearly data for each parameter from the period with information available in each country, from 1958 to 2017	Annual	N	High/UN data (FAO, WHO, etc.), and WRI data for flood occurrence interannual and seasonal water variability.	Very complete and very user friendly; just data, no maps. Interactive
IGRAC's Global Groundwater Network	https://ggmn.un-igrac.org/	Time series analysis of groundwater levels	Global	Groundwater site	Since 1950 to present	Decadal	N	High/Uses UNESCO's WHYMAP data, aquifer maps, soil maps, and land use maps	Time series analysis of groundwater well levels; plots historical data. Interactive.

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ source	Notes
IGRAC's Managed Aquifer Recharge	https://ggis.un-igrac.org/ggis-viewer/viewer/globalmar/public/default	Aquifer management by type of water, water use, management method, and sector. Drought hazard and frequency distribution.	Global	Groundwater site	Up to 2015, for now	Data for one year, but not the same year throughout the wells in the database	Climate change vulnerability scenarios 2005, 2050, 2100.	High/UNESCO's data and other sources (publications etc.), climate zone maps, population maps, and drought maps	Interactive and large data set, but limited information for now - the site does not include all existing data points, but is updated regularly. Very useful if the necessary well is on the map.
IHP Water Information Network System (WINS)	https://en.unesco.org/ihp-wins	Launched in Jan. 31 of 2017 with 212 data layers	Global	Varies per data layer	Varies per data layer	Varies per data layer	N	Good/ Intergovernmental Hydrological Programme (IHP)	Good data portal to exchange information on water resources
UNESCO WHYMAP network	https://www.whymap.org/whymap/EN/Home/whymap_node.html	River and groundwater basins capacity, aquifer vulnerability, groundwater extraction rates	Global	1 : 25,000,000 1 : 40,000,000 1 : 50,000,000	Only for years 2004, 2006, 2008, 2012, and 2015	Annual	N	High/UNESCO and CGMW (Commission for the Geological Map of the World)	Analysis of aquifer vulnerability to floods and droughts. Reported with detailed explanation of methods, etc. available in the web.

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ source	Notes
WRI's Aqueduct	http://www.wri.org/our-work/project/aqueduct/	Water stress based on withdrawals to annual supply and variability, with an economic growth component to project into the future. Data on flood occurrence, and drought severity.	Global	Country/ Province	Variable	Variable, monthly and annual	Y	Good/WRI	Interactive
WWF Global Lakes and Wetlands Database (GLWD)	http://www.worldwildlife.org/pages/global-lakes-and-wetlands-database	Large lakes and reservoirs, smaller water bodies, and wetlands divided by type. 3 levels of data available, depending on resolution.	Global	1: 1,000,000 to 1: 3,000,000			N	Probably accurate/WWF, data from diverse sources (e.g. USGS, ESRI, published manuscripts)	Good spatial reference data set but not temporally dynamic. Could be useful for flood risk assessment and coastal storm mitigation impact (because of the detailed level of wetland coverage information).

2. **Weather, precipitation and runoff**

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ source	Notes
Global Weather Forecast System NOAA	https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs	Weather Forecast- Temperature, wind, precipitation, soil moisture, atmospheric ozone concentration	Global	28 kilometers grid	Forecast out to 16 days	Hourly	Y	High/NOAA	Horizontal resolution to 70 kilometers after one week
Global stream flow forecasting presentation-an initiative under GEOGLOWS	https://www.arcgis.com/apps/Cascade/index.html?appid=bf1ecca8ff374691871ba82be570dcdb	Streamflow data predictions	Worldwide	River segments	Variable	15 days forecasting	Y	Depends if stream network has been digitized at the level of scale required /GEOGLOWS	Important for early warning systems, but still under development
NASA's GRACE	https://earthobservatory.nasa.gov/images/3666/earths-gravity-field https://tethys2.byu.edu/apps/newgrace/region/?region-select=2	Maps distribution and flow of water (calculates runoff from it), changes in water masses above- and belowground, and ocean currents based on gravity.	Global	Max. resolution is 4 x 4 degree	Data from 2003-2011	Monthly	N	High/NASA	Its best features are the time series with high resolution (time and geography) and the possibility to calculate flow and runoff. Two visualization tools of GRACE's data available at *ccar.colorado.edu/grace/about.html#img-7*

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ source	Notes
NASA's Global Change Master Directory	http://gcmd.nasa.gov/search/Titles.do?search#titles	Uploads NASA databases on groundwater (features, processes, and chemistry), surface water (chemistry, discharge/flow, runoff, features, depth, processes), etc.	Global	Variable	Since 1979	Variable. Most of them, monthly	N	High/NASA	Very complete, but level of detail (geographical and temporal) is very variable depending on the database accessed. Good starting point for water-related search; other databases are more user friendly and with actual data more accessible.
NOAA's Climate Data Online (CDO)	https://www.ncdc.noaa.gov/cdo-web/	Historical weather and climate data (includes temperature, precipitation, solar radiation, snowfall, etc.).	Global, but locations vary with searched parameter	N/A	1990-2016	Variable	N	High/NOAA	Historical data. Easy access. Most complete for USA; for other countries, data locations and abundance are variable.
World Bank Climate Change Portal	https://climateknowledgeportal.worldbank.org/	Temperature and precipitation	Global	Country, Region, Major Watersheds	1900-2016	Annual	Y	High/ The World Bank	Useful visualization of long-term average temperatures and rainfall by month

3. **Water quality**

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ Source	Notes
FAO AQUASTAT	http://www.fao.org/aquastat/en/overview/archive/river-sediment-yields	River sediment yields and wastewater (produced, collected, treatment, capacity of treatment facilities, posterior uses including irrigation).	Global	Country	2000 for sediment yields; 1988-2017 for wastewater.	Annual	N	High/UN data	Sediment yields have not been updated since 2000. However, this data set is the only source with this kind of information. Wastewater is part of the main database.
National Water Quality Monitoring Council	https://www.waterqualitydata.us/portal/	Physical, chemical, and biological data from hydrological stations.	A few countries (USA, Canada, Mexico, Nicaragua, Iraq…)	State level	Variable	Variable	N	US government data	For USA, seems good quality data but may not be reliable. Site selection seems too random.
WHO data repository	http://apps.who.int/gho/data/node.main.46?lang=en	% rural, urban, and total population using improved drinking water sources and improved sanitation facilities	Global	Country	1990, 2000, and 2015	Annual	N	High/UN data	Needs supplementary water data to calculate totals from these percentages (tip: use FAO's for consistent source of data).

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ Source	Notes
WRI's Eutrophication & Hypoxia Map Data Set	http://www.wri.org/resources/data-sets/eutrophication-hypoxia-map-data-set	Map of eutrophic, hypoxic, and recovering water bodies.	Global	Regional	variable, between 1850-2010	Annual	N	Published data, expected to be accurate/varies with the data point	Interactive. Unique source for global data.

4. **Auxiliary Data**

Database name	Database link	Type of data	Geographic coverage	Geographic resolution	Time period	Time resolution	Projection into future Y/N	Accuracy/ source	Notes
Africa GeoPortal	http://www.africageoportal.com/	Various-base maps, demographics, earth observations, urban systems, oceans, and historical maps	Africa	Variable per data set	Variable per data set	Variable per data set	N	Variable depending on source database	Useful website to share information
Nasa Global Land Data Assimilation System (GLDAS)	https://ldas.gsfc.nasa.gov/gldas	Various land data and assimilation sources	Global	2.5-degrees to 1 km	Variable	Variable	N	High/NASA	
NASA Earth Data	https://earthdata.nasa.gov	Multiple data sets from NASA's Earth Observing System Data and	Global	Multiple	Variable	Variable	N	High/NASA	

		Information System							
USGS Global Land Cover	https://archive.usgs.gov/archive/sites/landcover.usgs.gov/globallandcover.html	Land cover	Global	Landsat - 30m for tree, water, and bare ground; 2m for land cover	Variable - tree cover from 2010, water from 2000-2012, bare ground from 2010, land cover from 2010		N	High/USGS and U. of Maryland	Very useful, and very high geographical resolution, but similar information is already provided by other, more complete databases. Good backup.
USGS- Earth Resources Observation and Science (EROS) Center	https://www.usgs.gov/centers/eros	Land cover. Includes the Land Change Monitoring, Assessment, and Projection (LCMAP).	Global/ LCMAP US	Landsat - 30m		1985-Present	N	High/USGS	

5. **Web tools**

Database name	Database link	Type of data
UNESCO Global Water Forum	http://www.globalwaterforum.org/resources/data/	Accesses to some of the data sets from UN and WRI, and tools to calculate water use footprint.
IWMI Water Harvest Calculator	http://waterdata.iwmi.org/Applications/Rain_Water_Harvester/	Monthly calculation of water harvest and rooftop runoff based on collection efficiency, rooftop absorption rate, rooftop area, number of people, daily water use per person, and tank capacity.
FAO's Diagnostic Tools for Investment (DTI) in water for agriculture and energy	http://www.fao.org/nr/water/agwa/investment-tools/dti/tool/context/data	Water resources for agriculture and energy production in the dimensions of (1) Agriculture, (2) Irrigation, (3) Food security, poverty and food self-sufficiency, (4) Water resources and hydropower and (5) Environment and Climate change, using diverse specific indicators for each category.

ANNEX 1. USAID OPEN DATA POLICY

The United States Government is moving toward greater transparency, and a commitment to open data is a large component of this effort. In 2013, President Obama issued an Executive Order entitled *Making Open and Machine Readable the New Default for Government Information*, and shortly thereafter released a memo establishing a framework for an open data policy. The US CIO has been guiding the implementation of this policy across multiple departments and agencies. In October 2015, USAID released its own policy directive detailing open data requirements for implementing partners. Under ADS 579, all data collected under USAID contracts must be submitted to the Development Data Library (DDL) for the purposes of transparency and accountability.

This policy is part of a broader global movement to open up government data in the interest of increasing transparency, informing decision-making, and promoting private-sector innovation. While the Americas and Europe are generally further along than the rest of the world, there are emerging open data frameworks developing in Sub Saharan Africa and the rest of the world as well. The World Bank has been instrumental in pushing a greater open data agenda, as has the Open Government Partnership, which has 75 member countries. The latter's website has a great deal of data about the availability of data in each country and the progress made toward open data policies.

How This Affects the Water Sector

The most immediate effect of this policy is that, as implementation progresses, there will be more water data publicly available for use by new and ongoing development projects. This will create opportunities to learn from past projects, as well as provide a wealth of information to guide new policies, programs, and activities in the water sector, for USAID as well as across other donors. Combining implementer data with other data-rich sources will give us a better picture of successes and failures and allow us to design and execute better water security activities at the mission and project level. It's unclear if open data will actually make government more accountable, but it can certainly help the right stakeholders make better decisions about water policy and usage.

As of right now, there is very little water data available under the Development Data Library, presumably because the policy is relatively recent, and it is taking some time for implementers to fall into compliance, but we should expect this to grow over the next five years as program data is submitted.

The other side of this is that it creates a new obligation to ongoing and future water projects. Because water analysis, planning, and monitoring have the potential to generate a great deal of data, this sector may be more affected than other, less data-intensive areas. Water-oriented projects now have a duty and responsibility to properly organize and tag their data so that it can be effectively used.

Requirements

Data must be submitted to the DDL in a machine-readable format:

- *Raw data sets used to calculate performance data*: Performance data are the data that directly contribute to the project's Monitoring & Evaluation Plan — specifically the data collected to inform your performance indicators. Aggregate and derived values (such as final performance indicator numbers reported to USAID) do not need to be submitted to the DDL. However, raw data sets used to calculate indicator values (such as household survey data) must be submitted to the DDL.

- *Any data used to create an Intellectual Work*: This could include any significant data sets that are generated as a result of project activities. Examples may include: baseline, mid-term, and end-line studies, knowledge attitudes and practices (KAP) studies, household-level surveys, institutional capacity assessments, and similar items. Any activity collecting data on surface water, water points, or water users is likely to qualify. All data referenced in project publications, such as quarterly or final reports, fall into this category.

Regardless of the type of data, if applicable, the project must also include supporting documentation describing the data set, such as code books, data dictionaries, data gathering tools, notes on data quality, and explanations of redactions to the DDL. This documentation is critical to ensure other users can properly make sense of the data.

Data should be submitted "within thirty (30) calendar days after the data set is first used to produce an Intellectual Work or is of sufficient quality to produce an Intellectual Work," as per ADS 579. If 30 days is not feasible, the project must work with their COR to find a reasonable schedule.

Data Submission

Data should be submitted online at https://www.usaid.gov/data/DDLSubmissions. This form includes a number of straightforward programmatic questions, as well as some metadata fields that will help with categorization and organization of data.

You may propose a restricted access level for your data submission. There are a number of reasons for doing this, such as national security or the presence of personally identifiable information. There are three access levels:

- **Public**: The data set will be publicly listed on www.usaid.gov/data
- **Restricted**: The data set can be made available upon request under certain circumstances, such as academic research
- **Non-public**: The data set can only be shared with other government agencies

If your data set is already publically available via internet, you can also supply a web address to USAID to download the data directly or access it via Application Programming Interface (API). For instance, rather than submitting a new data set every quarter, mWater has submitted access to their data API, allowing a visitor to the DDL to find the most up-to-date data. If you don't have an online data source, once you have completed this form, USAID will contact you with instructions for submitting the actual data.

DDL data must be in *non-proprietary, machine-readable* formats.

- **Non-proprietary**: Must be an open standard, i.e. cannot be a private corporate format like Lotus, Access, or Excel
- **Machine-readable**: Data must be structured, like a spreadsheet, database, or other precise format; Word, PDF or HTML or qualitative formats are not acceptable

Commonly used, acceptable formats are:

- Comma Separated Value (CSV)
- Extensible Markup Language (XML)
- JavaScript Object Notation (JSON)

CSV is likely to be the most useful for basic tabular data submissions. It is a simple way to save a table that can be read by any computer, regardless of what programs it has installed.

ANNEX 2. DATA MANAGEMENT PLAN TEMPLATE

This template is designed as guidance and should be adapted to whatever context one is operating in. It is not strict or comprehensive, but should get one to think about all the essential elements of collecting and managing data. A document like this should be created not only to help think through details, but also as a handover document in case of staff or volunteer turnover. It is a living document that should be updated as a project evolves.

1. Project and Contact Information
Summary description of the Initiative.

Who is the main point of contact for the project and its data?

2. Plan and Acquire
How will the data be acquired?

If acquiring existing data sets, include more information, such as sources, frequency of updates, access methods, restrictions.

If collecting new data, include more information, such as method of collection and transmission, as well as any software or devices required.

How much data will be collected, in terms of the number of data points or overall data size?

Will the data be static or frequently updated? How frequently?

Are the appropriate hardware, software, and human resources available at a reasonable and sustainable cost?

3. Describe/Metadata and Manage Quality
How many new data sets will be created?

What are the data types and formats in which it will be stored?

Briefly describe the data processing steps. How will the data be analyzed?

What metadata will be associated with each data set?

What procedures will be used for ensuring data quality?

4. Backup/Secure and Preserve
Where will the data be stored in the short-term?

What will be the approach for routine backup of the data?

Describe any potential access restrictions.

What will be the final format of the data product?

Where will the data and metadata be preserved in the long-term?

If costs are associated with long-term storage, how will they be provided for?

5. Publish and Share
How will the data itself be shared and made available to other institutions or the public?

Will there be access or use restrictions on the data?

How can someone overcome any access restrictions?

Where will your metadata be stored to provide an access point for discovery by users?