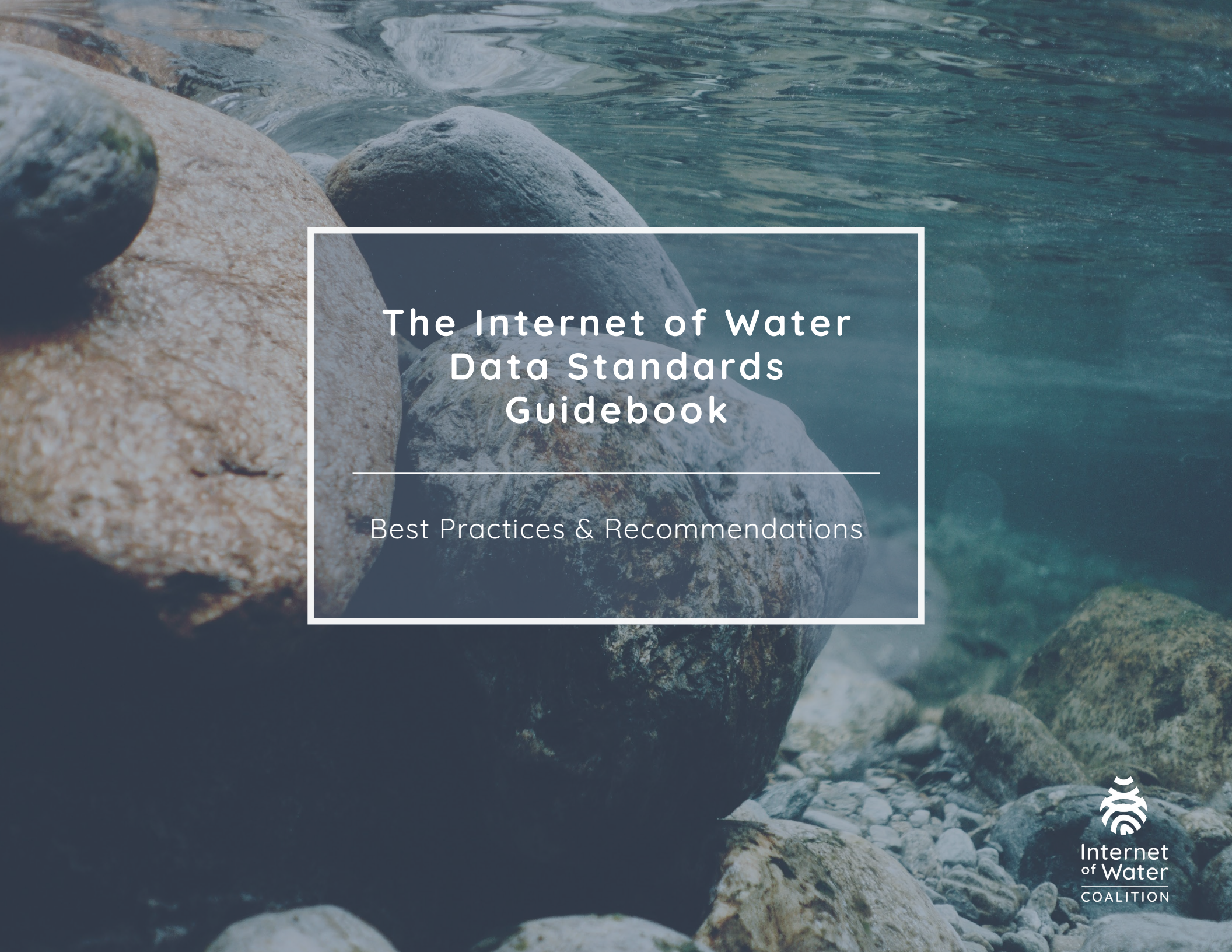

The Internet of Water Data Standards Guidebook

Best Practices & Recommendations

Internet of Water COALITION

CONTENTS

Introduction 1

Topic Hierarchy 2

Categories of Water Data and an Introduction to Linked Data 4

- Observational & Modeled Data 4
- Hydrologic Features 6
- Anthropogenic Features 7
- Relationships and Interactions 8
- Linked Data 8

Recommended Elements and Relevant Data Standards 10

- Hydrologic Features 10
- Anthropogenic Features 14
- Observational & Modeled Data 27

INTRODUCTION

Data standardization is crucial for effective water resource management, yet the landscape of water data standards remains complex and fragmented. The following text outlines the challenges of water data interoperability and presents the Internet of Water's approach to harmonizing diverse water data standards while preserving their contextual richness.

There are many different subtopics within the overarching concept of "water data," such as "hydrography", "water quality", and "water infrastructure." Each has a variety of data standards developed by federal and state agencies, international community standard-setting bodies, and private data management system vendors. In the interest of creating interoperability across the United States and globally, the Internet of Water recommends using international and federal standards where possible. However, the available data standards often do not explicitly reference interrelationships between data on different yet related topics.

Water datasets represent a network of interrelationships between observations about water as it courses through the water cycle. Each dataset carries its own significant contextual information. As such, the Internet of Water recommends augmenting the existing data standards with a means to link each data record to a framework common across as many datasets as practicable.

This guidebook introduces an overall categorization of water-related data and the concept of "Linked Data," then provides a high-level introduction to the Internet of Water Recommended Elements for data and metadata to be provided for each category to enable interoperability. The inclusion of these elements is sufficient to allow subject matter experts to integrate data about the same topic from different sources across the country. In addition, this guidebook provides descriptions of specific data standards in use by federal agencies where applicable, or other state agencies where there is no relevant official international or federal data standard. These standards are often highly complex and use-case specific and sometimes omit certain elements that the Internet of Water recommends. Thus the Internet of Water recommends using these standards if they are applicable, while also augmenting them with the Recommended Elements.

TOPIC HIERARCHY

The Internet of Water recommends that data portals implement a “topic hierarchy”. A topic hierarchy is a structured framework that organizes subjects or themes in a systematic and logical order, often from general to specific. This structure helps in understanding and navigating complex information by breaking it down into smaller, more manageable parts. Each level of the hierarchy provides a deeper dive into more detailed aspects of the overarching topic. Datasets represented in a data portal should be tagged with one or more relevant topics within the hierarchy, which can then be used to search and filter for datasets relevant to data portal users. The Internet of Water recommends the following topic hierarchy, which may be expanded to higher levels of detail. Note that the topics are non-exclusive, and a given dataset may be tagged with multiple topics in a cross-cutting manner. For example, a database of drinking water quality samples might be tagged as Water Quality-Physical/Chemical, Water Quality-Biological, and Infrastructure-Drinking Water-Distribution. The hierarchy is elaborated below:

HYDROGRAPHY

- Surface Water
  - » Streams & Rivers
  - » Lakes & Reservoirs
  - » Wetlands
- Groundwater
  - » Aquifers
  - » Wells and Boreholes
  - » Springs

HYDROLOGY

- Precipitation
- Evapotranspiration
- Runoff/Streamflow
- Storage
- Infiltration
- Models
- Forecasts
- Observations
  - » Land-based
  - » Remote sensing

WATER QUALITY

- Physical/Chemical
  - » Parameters (pH, turbidity, contaminants)
- Biological

    - Parameters (benthos, microbiological species)
- Ecological
    - Habitat assessment
    - Biodiversity

WATER RIGHTS AND USE

- Water Withdrawals and Consumptive Use
    - Agriculture
    - Industrial
    - Public Supply
    - Domestic
- Water Rights and Regulation
    - Ownership
    - Allocation
    - Points of use
    - Points of diversion
    - Transfer records
    - Withdrawal permits
    - Discharge permits

WATER INFRASTRUCTURE

- Reservoirs
    - Dam
    - Impoundment
- Drinking Water
    - Storage
    - Transmission
    - Distribution
    - Treatment
- Wastewater
    - Collection
    - Treatment
    - Discharge
- Stormwater
    - Collection
    - Treatment
    - Discharge

WATER MANAGEMENT, PLANNING AND POLICY

- Boundaries
    - Water management areas
    - Service areas
    - Government Jurisdictions
- Policy and Planning
    - Infrastructure finance
    - Water prices
    - Management plans

CLIMATE

- Observations
    - Land-based
    - Remote sensing
- Models

CATEGORIES OF WATER DATA AND AN INTRODUCTION TO LINKED DATA

While a topic hierarchy helps organize data for discovery of relevant datasets, interoperability between datasets can be fostered by ensuring that data are formatted in a standard way according to category. Water data categories for data standards purposes are somewhat independent of the Topic Hierarchy. For example, a single data standard can be sufficient to format observational data from field samples, sensor stations, and reports relevant to hydrology, water quality, water use, or climate. Figure 1 shows an overall nested categorization of water data by specific types of "features". A feature is a specific, singular unit of observation or analysis, about which a dataset might record static attributes (e.g. location in latitude and longitude) or dynamic observations (e.g. water flow at a specific time). At a high-level the Internet of Water recommends augmenting existing water data standards for given feature types by recording some common attributes so that it is easy to link and analyze datasets about different topics together. A feature could be a streamgage, a dam, a lake, a stream reach, a drinking water system, or a specific component of a drinking water system such as a water treatment plant.

The categorization and interrelationships of water data types are described below:

OBSERVATIONAL & MODELED DATA

Data collected from or simulated about natural and anthropogenic systems for analysis and decision support.

- **Surface Water Sites:** Locations where surface water data are collected, essential for monitoring ecosystem health and managing water resources.
- **Groundwater Sites:** Points of groundwater data collection, key for understanding the current state and trends of groundwater levels and quality.
- **Anthropogenic Monitoring Systems:** Systems monitoring attributes on infrastructure, or summarizing data about anthropogenic boundaries. For example, systems monitoring water quality of effluent from wastewater systems or summarizing water use within a jurisdictional boundary.
- **Observations:** Recorded data points describing specific measurements of water quality, quantity, or other

relevant parameters, foundational for scientific analysis, policy-making, and public information.

Observed or modeled datasets should include fields (example watershed id) that enable the users to link the data to the hydrologic or anthropogenic feature/s it monitors

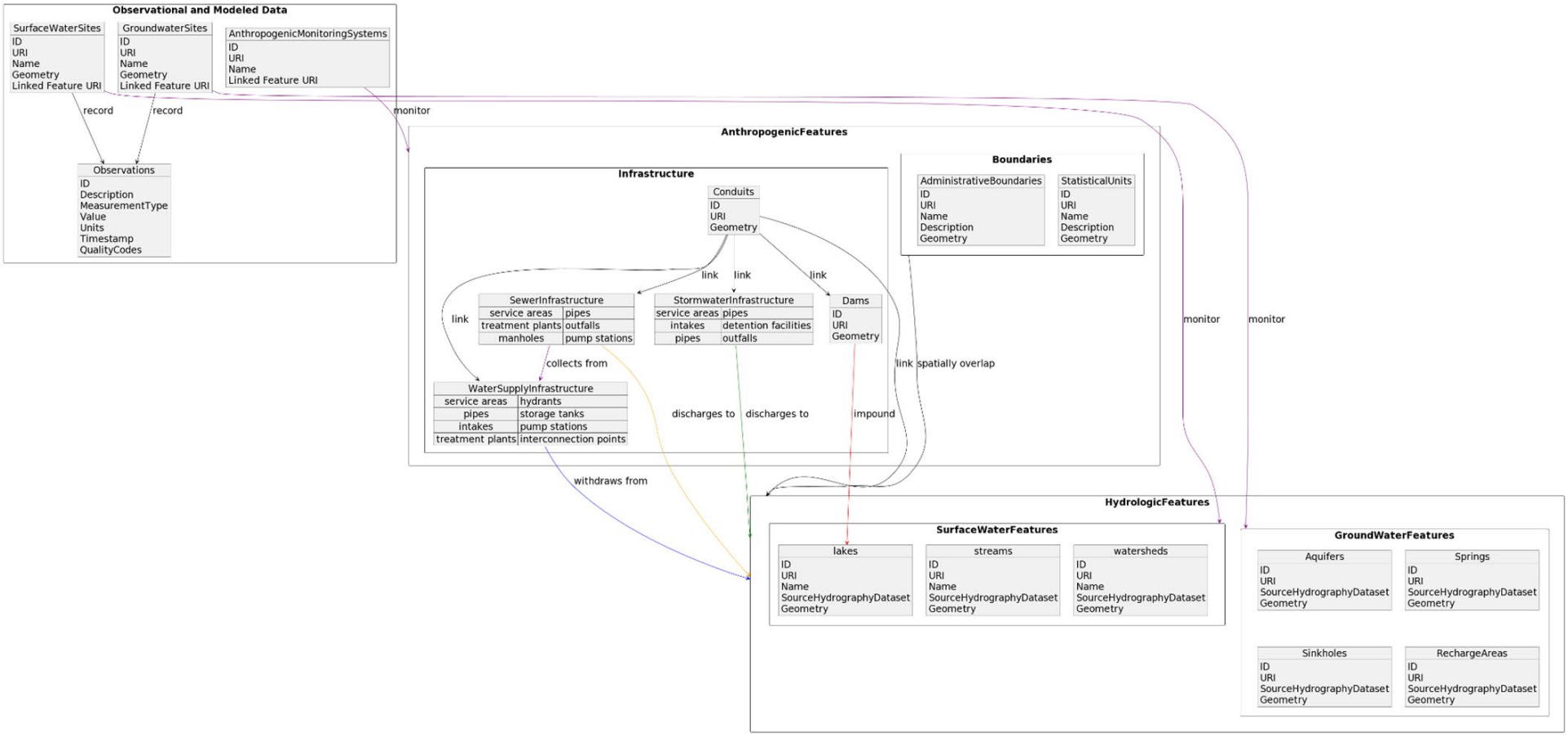

Figure 1. Overall Categorization of Water Data and Relationships between Categories. Click to download full-size image

HYDROLOGIC FEATURES

Features representing natural water systems and their components. Datasets commonly catalog these features, assigning them identifiers, geometrical representations, and sometimes attributes such as names or calculated values such as length, area, or hydrological summary statistics.

Surface Water Features: Natural bodies of water exposed to the atmosphere.

- **Watersheds:** Land areas that drain rainfall to a common outlet, crucial for understanding hydrological cycles and managing water resources.
- **Streams:** Bodies of flowing water moving in one direction, critical for ecosystem health, water supply, flood management, and recreation
- **Lakes:** Inland bodies of standing water, important for biodiversity, recreation, and as water sources.

Groundwater Features: Various formations and phenomena associated with the movement of water below the Earth's surface.

- **Aquifers:** Underground layers of water-bearing rock, sand, or gravel, vital for groundwater supply and impacting surface water interactions.
- **Springs:** Natural outlets where groundwater flows out of the ground onto the surface.
- **Sinkholes:** Depressions or holes in the ground caused by the collapse of a surface layer, often resulting from the dissolution of limestone or other soluble rocks by groundwater.
- **Recharge Areas:** Locations where water from precipitation and surface water is able to infiltrate the ground and replenish an aquifer.

ANTHROPOGENIC FEATURES

Human-made or influenced features impacting or delineating water resources.

Boundaries: Defined edges of areas used for administrative, statistical, or management purposes.

- **Statistical Units:** Areas defined for the collection and analysis of statistical data, helping to understand

demographics, land use, and economic activities. Examples include Census tracts and Metropolitan Statistical Areas

- **Administrative Boundaries:** Politically defined borders of government jurisdictions and service areas, important for governance, policy implementation, and regulation.

Infrastructure: Constructed systems and facilities serving the community. Infrastructure data can become quite complex depending on the granularity of analysis. Only high-level categories are provided here. Further elaborations are provided in the detailed data standards.

- **Dams & other control structures:** Structures built on rivers or streams to control water flow, used for hydroelectric power, water storage, navigation, and flood control.
- **Water Supply Infrastructure:** Systems and facilities for treating, storing, and distributing water for use in drinking water or irrigation systems. This infrastructure can have data on components like service areas, hydrants, pipes, wells, injection wells, storage tanks, and treatment plants.
- **Sewer Infrastructure:** Facilities for collecting, transporting, treating, and discharging wastewater, critical for sanitation and environmental protection.
- **Stormwater Infrastructure:** Systems designed to manage runoff from precipitation, crucial for preventing flooding and protecting water quality.
- **Conduits:** Channels or pipelines facilitating the flow of water between different parts of the hydrologic system and water infrastructure, representing a modification of natural hydrology for water storage and distribution.

RELATIONSHIPS AND INTERACTIONS

- **Spatial Overlap and Impacts:** Anthropogenic features intersect with natural hydrologic features, affecting water bodies through processes like impoundment, withdrawal, collection, and discharge.
- **Monitoring Data and Features of Interest:** Observed and modeled data generally have “features of interest” – that is, the feature(s) that the data is about. These features may be anthropogenic or hydrologic features.

The interplay between different datasets provides a multidimensional perspective on water resources, their management,

and their sustainability. For instance:

- The relationship between natural water bodies and the data collected from monitoring sites is foundational. Streams' flow rates and lakes' water quality measurements offer insights into ecosystem health and inform conservation strategies.
- The interaction between dams and rivers represents a direct human intervention in natural water cycles. Analyzing these interactions can help assess the impacts of infrastructure on water availability and ecosystem services.
- Statistical units and administrative boundaries intersect with water quality data, providing a socio-economic context to environmental data. This intersection aids in targeting mitigation efforts and enforcement actions according to society's priorities.

This guidebook emphasizes the importance of linking datasets in a meaningful way to address specific questions about water resource management. By elucidating the potential interrelationships among the datasets identified in the UML diagram (Figure 1), we aim to equip users with the knowledge to explore these connections further, fostering a holistic understanding of water systems and informing more nuanced and effective decision-making processes.

LINKED DATA

The Internet of Water recommends a linked data approach. Linked data enables interconnected, structured data across the web, facilitating easier access and integration. URIs (Uniform Resource Identifiers) are essential alongside native IDs because they provide unique, global identifiers for resources, ensuring consistent reference across different systems and datasets. URIs take the format of URLs, and can serve a dual purpose of directing to a resource on the internet that describes that item. While IDs might be unique within a specific dataset or system, URIs guarantee uniqueness across the entire web, making them critical for linking data across diverse sources and enabling seamless data discovery, sharing, and reuse in a global context.

Consider a scenario where a datahub includes a dataset of streamgages and their readings and a dataset of dams. If dam operations data needed to be linked to stream inflow and outflow data, one way to do this would be to download all of the streamgage data, all of the dam data, and a stream network dataset, perform complex GIS operations to identify which streamgages were associated with each dam, and then proceed from there. Alternatively, the Internet of

Water recommends populating the streamgage dataset and the dam dataset with URIs for the relevant stream reaches that each streamgage and dam are on. To make the connection explicit, both the streamgage and the dam would be linked to the same stream URI, such as https://geoconnex.us/ref/mainstems/1431511 for the North Santiem River. This URI represents the stream itself, allowing for a clear connection between the streamgage (https://waterdata.usgs.gov/monitoring-location/444508122171000) and the dam(https://geoconnex.us/ref/dams/1075935) as both interact with the North Santiem River (https://geoconnex.us/ref/mainstems/1431511). This linkage provides a cohesive framework for analyzing the stream's flow data, the dam's effects, and their combined impact on water management and ecological health.

Thus, as shown in Figure 1, the Internet of Water recommends that all Surface Water, Groundwater, and other "sites" where data is collected include a metadata field called "URI" and a "Linked Feature URI". The URI identifiers the site itself, and the "Linked Feature URI" identifies the relevant surface or groundwater or anthropogenic feature. Geoconnex provides URIs for many features of interest nationwide, including named rivers, dams, aquifers, census geographies, streamgages, etc. These can be found at https://reference.geoconnex.us Other URIs from Federal databases such as the Water Quality Portal (water quality samples from federal and state agencies and non-governmental organizations across the U.S) , HydroSource (Oak Ridge National Laboratory Hydropower), Water Data for the Nation (USGS water data services including stream gages and monitoring wells), and the Reclamation Information Sharing Environment (USBR data repository) can be found in the Geoconnex knowledge graph and web map. Learn more about Geoconnex here.

RECOMMENDED ELEMENTS & RELEVANT DATA STANDARDS

This section of the guidebook provides, for each data type from Figure 1, an overview of the Internet of Water's most Recommended Elements. For most types, the data can be represented in a single table, where each Recommended Element corresponds to a column. For complex data types that require multiple tables, a link to an example Excel template with the minimum fields filled out is provided. Where applicable, descriptions and links to more detailed documentation for relevant applicable data standards, which in many cases have more detailed requirements than Internet of Water Recommended Elements but may omit some of those elements, are provided.

HYDROLOGIC FEATURES

SURFACE WATER FEATURES

In general, it is recommended that datasets about surface waters reference a common hydrography. The Internet of Water recommends using the 3DHP (3D Hydrography Program) "mainstem id" concept as the common hydrographic reference for datasets related to surface waters. 3DHP, which will replace the NHD (National Hydrography Dataset) and WBD (Watershed Boundary Dataset) products, will be based on best-available LIDAR data that USGS will process into elevation-derived hydrography products over the coming years. As such, the exact fine-scale flowlines and waterbodies may change over time. The "mainstem id" is a more persistent identifier designed to allow external datasets to be referenced to the 3DHP in a way that is interoperable over different versions of the 3DHP as it improves and evolves over time. The mainstem is the relationship between a given headwaters and outlet location or confluence. Thus, a mainstemID could identify the Mississippi River, or a very small, unnamed ephemeral tributary from its headwaters until its confluence with a named stream. The latest 3DHP data is available from a web service here: https://hydro.nationalmap.gov/arcgis/rest/services/3DHP_all/MapServer The 3DHP data model for flowlines with associated definitions is as follows1:

Flowlines are polylines representing linear water features (streams, including flowpaths through water bodies like lakes). They include x,y, and z coordinates. They also include the following attributes:

Name (Alias)	Definition	Type	Allow Nulls	Length	Domain (allowed values list, if applicable)	Default Values	Comments
id3dhp	Unique identifier for 3DHP features.	Text	Yes	7	-	-	This is a base-36 7-digit alphanumeric unique identifier that is not persistent. Applied to hydrolocation features in this case.

1 https://www.usgs.gov/ngp-standards-and-specifications/3d-hydrography-program-product-specification#:~:text=The%203DHP%20is%20comprised%20of,Hydrography%20Standards%20and%20Specifications%20website.

featuredate	Date the feature was loaded into the 3DHP core database.	Date	Yes	-	-	-	-
mainstemid	A cross-dataset identifier for all flowlines that represent the headwater to outlet path of a river.	Text	Yes	200	-	-	A mainstem is equivalent to a feature with the same levelpathid in NHDPlus HR. The text field will be of the form https://geoconnex.us/ref/mainstems/<id> or https://geoconex.us/usgs/mainstems/<id>
gnisid	A permanent, unique number assigned by the Geographic Names Information System (GNIS)** to a geographic feature name for the sole purpose of uniquely identifying that name.	Long Integer	Yes	-	-	-	gnisid = “null” if no name is associated with the feature. The gnisid is conflated from points and assigned to features. The most recent name, and any historic names, can be retrieved from the GNIS data base using the gnisid.
featuretype	Feature type description.	Short Integer	Yes	-	flowline featuretype	-	1 - River: Flowing body of water that receives inflow from upstream and surrounding catchment. 2 - Canal: Flowing body of water that receives inflow from upstream but not the surrounding catchment. 3 - Drainageway: Drainage pathway in a low drainage area setting (headwater) upstream of the onset of discernable channelization. 4 - Surface Connector: Abstract surface or near surface path used to connect upstream channelized features with downstream channelized features 5 - Waterbody Connector: Abstract connector over a portion of the landscape covered by water. 6 - Elevation Breaching Connector: A known or inferred connection that is used to breach values in the elevation surface that are blocking the natural downstream flow of a hydrologic feature. 7 - Hydro Unenforced Connector: Specific or abstract connector representing flow that is not determined by the surface-water hydrologic network

featuretypelabel	The name of the feature type.	Text	Yes	50	-	-	The featuretype label is based on the coded value description in the featuretype field. Allows label display in non-ESRI datasets.
lengthkm	Length of linear flowline feature. Value based on regional spatial reference systems.	Double	Yes	-	-	-	Computed in projections based on specific locations.
waterbodyid3dhp	The id3dhp of the waterbody that a waterbody connector flows through.	Text	Yes	7		-	-
flowdirection	Identifies the flow direction of a feature relative to the direction it was digitized.	Short Integer	No	-	Flow direction type	1	
onsurface	Defines vertical relationship of flowline features.	Short Integer	No	-	Onsurface type	1	Indicates whether the flowline feature is on the land surface, below the surface, or elevated above the ground and another hydrography feature.

The “mainstemid” then, which is a URI, can be used to tie a given location to a specific river (a relationship between a headwater and outlet or confluence). For example, https://geoconnex.us/ref/mainstems/1431511 identifies the North Santiam River. A metadata record for a streamgage could include attributes:

“featureOfInterest_type”: “stream”
“featureOfInterest_name”: “North Santiam River”
“featureOfInterest_uri”: “https://geoconnex.us/ref/mainstems/1431511”

GROUNDWATER FEATURES

Unlike surface water hydrography, high-resolution information about groundwater hydrogeologic units is highly local and fragmented. In addition, many types of groundwater features are possible to monitor, and these features are not interrelated in the same way as surface waters, which are generally part of a network that is straightforward to traverse “upstream” and “downstream”. The Internet of Water recommends the use of the OGC Groundwater Markup Language Standard when curating datasets about hydrogeologic units, including aquifers and groundwater basins. In general,

hydrogeologic units would be identified, associated with a geometry and attributes, and then have their identifiers referenced in relevant data about monitoring or extraction locations such as wells or springs. Many properties may be recorded about hydrogeologic units. The Internet of Water recommends the following data elements be required:

Name (Alias)	Definition	Type	Allow Nulls	Length	Domain (allowed values list, if applicable)	Default Values	Comments
groundwaterID	Unique identifier groundwater feature	Text	No	7	-	-	This is a base-36 7-digit alphanumeric unique identifier that is not persistent. Applied to hydrolocation features in this case.
uri	A cross-dataset identifier for all groundwater features across all organizations in the state	Text	Yes	200	-	-	Should be in the form of a URI. eg. https://geoconnex.us/owdp/gwfeatures/<id>
name	A cross-dataset identifier for all groundwater features across.	Text	Yes	200	-	-	A
featuretype	Feature type	Text	No	-	GroundwaterML2 Hydrogeologic Unit Types	-	Relevant types for groundwater include: 7=Groundwater basin 8=Unconfined aquifer 9=Confined aquifer 10=Artesian aquifer 11=Subartesian aquifer 12=Aquitard 13=Spring
featuretypelabel	The name of the feature type.	Text	Yes	50			See above
mediatype	Type of material or, by proximity, type of voids	Text	Yes		GroundwaterML2 Media types		Granular, fracture, karstic, mixed
Geometry	Machine-readable coordinates defining 2 or 3D geometry, including coordinate reference system	Geometry	No	-			Point, polyline, polygon, multipolygon are posssible

ANTHROPOGENIC FEATURES

BOUNDARIES

The Federal Geographic Data Committee provides standards for representing boundaries. Many boundaries, such as Census units, and state, county, and local government jurisdictional boundaries, are already curated and published by the U.S. Census Bureau according to these standards. Other boundary types, such as state agency administrative districts, water management areas, drinking water service area boundaries, etc. will need to be curated using these standards.

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
unitID	Identifier assigned to the governmental unit, administrative unit, statistical unit, or other unit.	Text	No	-	Framework::Identifier	-	-
unitType	Type code or abbreviation for the feature unit type.	Text	Yes	-	-	-	Examples: Water ServiceArea Sewer ServiceArea Groundwater Management Area Urban Growth Boundary Urban Area
instanceName	“Official” feature name.	Text	No	-	-	-	Preferably from GNIS.
instanceAlternativeName	“Unofficial” or variant feature name.	Text	Yes	-	-	-	-
officialDescription	Descriptive information about the geographic area.	Text	Yes	-	-	-	-
instanceCode	Specific code identifying the geographic area.	Text	No	-	-	-	-
codingSystemReference	Reference identifying the instance code type.	Text	Conditional	-	-	-	Required if code exists.

effectiveDate	Date on which a geographic area change took effect.	Date	Yes	-	Any valid geometry type from ISO 19107	-	YYYY-MM-DD
geometry	Geometric representation of the feature.	Geometry	Yes	-	GM_Surface	Defined in ISO 19107	-

INFRASTRUCTURE

WATER DISTRIBUTION SYSTEMS

Water Distribution Data Standards are based on those in use by the Kentucky Infrastructure Authority. Example datasets can be explored here.

SURFACE SOURCES

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
pwsID	Public Water System ID	Text	No	-	-	-	-
systemName	Name of the Water System	Text	No	-	-	-	-
sourceID	Identifier for the Water Source	Text	No	-	-	-	-
federalID	Federal Identifier	Text	Yes	-	-	-	Optional for Reservoir
surfaceSourceName	Name of the Surface Source	Text	No	-	-	-	-
intakeLocationDescription	Description of Intake Location	Text	No	-	-	-	-
surfaceSourceType	Type of Surface Source	Text	No	-	“River”, “Lake”	-	-
availability	Availability Status	Text	No	-	“Permanent”, “Reserve”	-	-
surfaceSourceintakeElevation	Elevation of Intake Source	Integer	No	-	-	-	Measured in feet

surfaceSourceAverageWithdrawal	Average Withdrawal (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
surfaceSourcemaxWithdrawal	Maximum Withdrawal (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
HUC12	Hydrologic Unit Code	Text	No	12	-	-	12-digit code
id3dhp	ID for 3D Hydrography Program stream segment	Text	Yes	-	-	-	-
mainstem_uri	Mainstem Unique Identifier	Text	Yes	-	See https://reference.geoconnex.us/collections/mainstems/items	-	-
geometry	Geospatial Data Point	Geometry	No	-	Point	-	Coordinates for location, in WGS 84 latitude/longitude

GROUNDWATER SOURCES

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
pwsid	Public Water System ID	Text	No	-	-	-	-
systemName	Name of the Water System	Text	No	-	-	-	-
groundwaterSourceID	Identifier for Groundwater Source	Text	No	-	-	-	-
wellName	Name of the Well	Text	Yes	-	-	-	-
groundwaterSourceName	Name of the Groundwater Source	Text	No	-	-	-	-
groundwaterSourceType	Type of Groundwater Source	Text	No	-	“Well”, “Spring”	-	-
wellDepth	Depth of the Well	Integer	Yes	-	-	-	Measured in feet

availability	Availability Status	Text	No	-	“Permanent”, “Reserve”	-	-
groundwaterSourceElevation	Elevation of Groundwater Source	Integer	No	-	-	-	Measured in feet
groundwaterSourceAverageWithdrawal	Average Withdrawal (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
groundwaterSourceMaxWithdrawal	Maximum Withdrawal (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
aquifer_uri	Aquifer Unique Identifier	Text	No	-	-	-	-
geometry	Geospatial Data Point	Geometry	No	-	Point	-	Coordinates for location

DRINKING WATER PIPELINES

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
pwsid	Public Water System ID	Text	No	-	-	-	-
uri	Unique Resource Identifier	Text	No	-	-	-	-
dwPipeID	Drinking Water Pipe Identifier	Text	No	-	-	-	-
dwPipePurpose	Purpose of the Drinking Water Pipe	Text	No	-	“Distribution”, “Transmission”	-	-
dwPipeSize	Size of the Drinking Water Pipe	Integer	No	-	-	-	Measured in inches
dwPipeYearConstructed	Year the Pipe was Constructed	Integer	Yes	-	-	-	-
dwPipeMaterial	Material of the Drinking Water Pipe	Text	No	-	“PVC”, “Cast Iron”, “AC”, “Ductile Iron”, “Unknown”, “HDPE”, “Other”	-	-
geometry	Geospatial Data of the Pipeline	Geometry	No	-	Polyline	-	Line coordinates for pipeline

WATER TREATMENT PLANTS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
pwsid	Public Water System ID	Text	No	-	-	-	-
wtpID	Water Treatment Plant ID	Text	No	-	-	-	-
wtpName	Name of the Water Treatment Plant	Text	No	-	-	-	-
wtpConstructionDate	Construction Date of the Plant	Date	Yes	-	-	-	-
wtpDesignCapacity	Design Capacity of the Plant (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
wtpAverageDailyProduction	Average Daily Production (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
wtpMaxDailyProduction	Maximum Daily Production (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	Coordinates for plant location

STORAGE TANKS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
pwsid	Public Water System ID	Text	No	-	-	-	-
storageTankID	Storage Tank Identifier	Text	No	-	-	-	-
storageTankName	Name of the Storage Tank	Text	No	-	-	-	-
storageTankConstructionDate	Construction Date of the Storage Tank	Date	Yes	-	-	-	-
storageTankMaterial	Material of the Storage Tank	Text	No	-	“Steel”, “Concrete”	-	-
storageTankCapacity	Capacity of the Storage Tank (Gallons)	Integer	No	-	-	-	-
storageTankType	Type of Storage Tank	Text	No	-	“Elevated”, “Standpipe”, “Ground Storage”	-	-

storageTankOverflowElev	Overflow Elevation of the Tank	Integer	Yes	-	-	-	Measured in feet
geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	Coordinates for tank location

WASTEWATER SYSTEMS

Wastewater System Data Standards are based on those in use by the Kentucky Infrastructure Authority. Example datasets can be explored here.

SEWER LINES

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
npdesPermitID	NPDES Permit ID	Text	No	-	-	-	-
wastewaterSystemName	Name of the Wastewater System	Text	No	-	-	-	-
sewerLinePurpose	Purpose of the Sewer Line (e.g., collector)	Text	No	-	“Collector”, “Interceptor”, “Outfall”, “Relief”, “Overflow”	-	-
sewerLineYearConstructed	Year the Sewer Line Was Constructed	Integer	Yes	-	-	-	-
sewerLineType	Type of Sewer Line	Text	No	-	“Gravity”, “Force”	-	-
sewerMaterial	Material of the Sewer Line	Text	No	-	“PVC”, “AC”, “VCP”,”Concrete”, “Cast Iron”, “Steel”, “FRP”, “HDPE”, “Other”	-	-
geometry	Geospatial Data of the Sewer Line	Geometry	No	-	Polyline	-	Line coordinates for sewer line

WASTEWATER TREATMENT PLANTS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
npdesPermitID	NPDES Permit ID	Text	No	-	-	-	-
wastewaterSystemName	Name of the Wastewater System	Text	No	-	-	-	-
wwtpName	Name of the Wastewater Treatment Plant	Text	No	-	-	-	-
wwtpConstructionDate	Construction Date of the WWTP	Date	Yes	-	-	-	-
wwtpAverageDailyFlow	Average Daily Flow Processed (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
wwtpMaxDailyFlow	Maximum Daily Flow Capacity (MGD)	Float	No	-	-	-	MGD = Million Gallons per Day
wwtpTreatmentProcess	Treatment Process Type	Text	No	-	“Primary”, “Secondary”, “Tertiary”	-	-
geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	Coordinates for WWTP location

WWTP OUTFALLS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
wwtp_uri	Uri of the associated WWTP	Text	No	-	-	-	-
waterbody_uri	URI of the water body it discharges to. Can be any Surface of Groundwater featurte	Text	Yes	-	-	-	-
Geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	Coordinates for lift station location

SEWER LIFT STATIONS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
outfallID	ID of the Sewer Lift Station	Text	No	-	-	-	-
numberPumps	Number of Pumps	Integer	No	-	-	-	-
Stormwater	Whether it Handles Stormwater	Text	No	-	“Yes”, “No”	-	-
Use	Use of Lift Station	Text	No	-	“In Line”, “Influent”, “Wet well”, “Other”	-	-
Geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	Coordinates for lift station location

STORMWATER SYSTEMS

These standards are based on the Metro Stormwater Data Standard.

STORMWATER PIPELINES

Name	**Definition**	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
strormwaterPipeID	Identifier for the Pipeline	Text	No	-	-	-	-
stormwaterPipeType	Type of Stormwater Pipeline	Text	No	-	“Culvert”, “Collector”, “Interceptor”, “Retention”, ”Outfall”	-	
stormwaterPipeShape	Shape of the Pipeline	Text	No	-	“Elliptical”, “Circular” “Arch”, “Horseshoe”, “Box”, “Other”	-	

stormwaterPipeMaterial	Material of the Pipeline	Text	No	-	“PVC”, “Concrete”, “Steel”, “AC”, “Ductile Iron”, “Other”	-	-
stormwaterPipeCoverage	Coverage Type of the Pipeline	Text	Yes	-	“Covered”, “Exposed”	-	
stormwaterPipeYearConstructed	Year the Pipeline was Constructed	Integer	Yes	-	-	-	-
geometry	Geospatial Data of the Pipeline	Geometry	No	-	Polyline	-	Line coordinates for pipeline

CHANNELS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
channelID	Identifier for the Channel	Text	No	-	-	-	-
channelType	Type of Stormwater Channel	Text	No	-	-	-	Examples: Natural, Engineered
channelLining	Lining Material of the Channel	Text	Yes	-	-	-	Examples: Concrete, Rock, Vegetative
channelShape	Shape of the Channel	Text	Yes	-	-	-	Examples: Trapezoidal, Rectangular
channelYearConstructed	Year the Channel was Constructed	Integer	Yes	-	-	-	-
geometry	Geospatial Data of the Channel	Geometry	No	-	Polyline	-	Line coordinates for channel

BASINS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
basinType	Type of Basin	Text	No	-	“Lake”, “Pond”, “Retention”, “Detention”, “Constructed Wetland”, “Other	-	

basinLining	Lining Material of the Basin	Text	Yes	-	“Grass”, “Natural Vegetation”, “Riprap”, “Polyethylene”, “PVC”, “Other”	-	
basinOrigin	Origin of Basin Formation	Text	Yes	-	“Natural”, “Constructed”, “Modified Natural”, “Restored to original” “Unknown”	-	
basinImpaired	Whether the Basin is Environmentally Impaired	Text	Yes	-	“Impaired”, “Not impaired”, “Prior impairment”, “Maintenance”, “At-risk”, “Not applicable”, “Unknown	-	
basinYearConstructed	Year the Basin was Constructed	Integer	Yes	-	-	-	-
geometry	Geospatial Data of the Basin	Geometry	No	-	Polygon	-	Polygon coordinates for basin area

HYDRAULIC CONTROL STRUCTURES

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
hcsID	Identifier for the Hydraulic Control Structure	Text	No	-	-	-	-
hcsType	Type of Hydraulic Control Structure	Text	No	-	See codelist	-	Examples: Gate, Valve, Spillway
hcsYearConstructed	Year the Structure was Constructed	Integer	Yes	-	-	-	-
geometry	Geospatial Data Point of the Structure	Geometry	No	-	Point	-	Coordinates for structure location

STORMWATER LIFT STATIONS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
stormwaterLiftStationID	Identifier for the Stormwater Lift Station	Text	No	-	-	-	-
stormwaterLiftStationNumPumps	Number of pumps	Integer	No	-	-	-	-

stormwaterLiftStationYearConstructed	Year the Lift Station was Constructed	Integer	Yes	-	-	-	-
geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	Coordinates for lift station location

INLETS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
stormwaterInletID	Identifier for the Inlet	Text	No	-	-	-	-
stormwaterInletShape	Shape of the Inlet	Text	No	-	“Rectangle”, “Square”, “Circular”, “Trapezoid”, “Other”, “Unknown”, “Not applicable”	-	
stormwaterInletYearConstructed	Year the Inlet was Constructed	Integer	Yes	-	-	-	-
geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	Coordinates for inlet location

OUTLETS

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	
stormwaterOutletID	Identifier for the outlet	Text	No	-	-	-	
outletType	Type of Outlet	Text	No	See Codelist	”	-	
outletYearConstructed		Integer	Yes				
id3dhp	ID for 3D Hydrography Program stream segment	Text	Yes	-	-		
mainstem_uri	Mainstem Unique Identifier	Text	Yes	-	See https://reference.geoconnex.us/collections/mainstems/items		
geometry	Geospatial Data Point Location	Geometry	No	-	Point	-	

BEST MANAGEMENT PRACTICES

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
uri	Unique Resource Identifier	Text	No	-	-	-	-
bmpID	Identifier for the BMP	Text	No	-	-	-	-
bmpType	Type of BMP	Text	No	-	See Codelist	-	Examples: Bioretention
bmpFilterMaterial	Filter Material Used in BMP	Text	Yes	-	See Codelist	-	Examples: Sand
bmpGroundCover	Ground Cover of BMP	Text	Yes	-	See Codelist	-	Examples: Grass
geometry	Geometry	Geometry	No	-	Point or Polygon	-	Coordinates or area of BMP location

OBSERVATIONAL & MODELED DATA

RECOMMENDED ELEMENTS

There are many data standards for observed and modeled data about water. Many are designed for specific purposes, such as high-frequency environmental sensor data, discrete water quality samples, or for regulatory reporting, scientific data sharing, or data visualization. In general, for water observation and model data to be interoperable, the following information is generally required, although the exact format and vocabularies used vary. In general, the Internet of Water recommends organizing data about three types of entities for observational and modeled data: site, methodological metadata, and data.

Site entities include attributes of the location where data was collected and the feature of interest of the data. The feature of interest can be a member of ANY of the above data categories (hydrologic features, boundaries, or any of the infrastructure subcategories). Including this information allows observational or modeled data about any important water feature, natural or human-made, to be integrated. This is why it is important for there to be a URI for all environmental or human-made water features. These URIs can be used as feature of interest identifiers for any kind of observed or modeled data about those features, including water quality samples, continuous sensor readings, administrative or regulatory determinations, etc.

Metadata entities include information about what variables and in what units data was collected or modeled and the methods used to collect or generate the data. Data includes the actual observed or modeled values and any data quality information.

This data model is applicable to any dataset where some characteristic of water is being measured, estimated, or modeled and associated with a particular location or piece of infrastructure; ranging from streamgages to water quality sample locations to water treatment plants to points of diversion or use associated with water rights.

SITE METADATA

Name (Alias)	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
siteID	An alphanumeric string uniquely identifying the site.	Text	No	-	-	-	-
uri	A cross-dataset identifier for all features across all organizations in the state.	Text	Yes	200	-	-	Should be in the form of a URI. eg. https://geoconnex.us/owdp/sites/<id>
Geometry Longitude CRS	Machine-readable coordinates defining 2D or 3D geometry, including coordinate reference system Specific coordinate reference system used for latitude/ longitude	Geometry Decimal Text	No No No	- - -	- -	- -	Point, polygon, polyline - -
name	The official name of the site.	Text	No	-	-	-	-
organization	The entity responsible for the site's operation, monitoring activities, and/or sample collection.	Text	No	-	-	-	-
facilityType	The type of monitoring equipment or installation	Text	No	-	-	-	e.g., well, streamgage, sonde, buoy
featureOfInterestType	The type of feature that the data is about	Text	No	-	One of the above data types, whether within hydrologic features or among the Boundary or Infrastructure types	-	Examples": "Boundary", "surfaceWaterSource", "river", "canal", "aquifer", "Wastewater Treatment Plant Outfall" , "Best Management Practice"
featureOfInterest	The specific feature the data collection focuses on.	Text	No	-	-	-	Preferably a URI if available. If a surface water feature, can be the mainstemid from 3DHP.

METHODOLOGICAL METADATA

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
metadataID	A unique identifier for the metadata record.	Text	No	-	-	-	-
siteID	Links the metadata to its corresponding site.	Text	No	-	-	-	Should match the siteIDin the Site table.
observedPropertyName	The variable or parameter being observed.	Text	No	-	-	-	-
observedProperty_uri	A URI for the observed property, using controlled vocabularies (e.g., ODM2 VariableName Vocabulary).	Text	Yes	200	-	-	-
valueType	The nature of the data values (e.g., measurement, calculation), with URIs for reference (e.g., ODM2 ResultType Vocabulary).	Text	No	-	-	-	-
unitName	The unit of measurement.	Text	No	-	-	-	-
unit_uri	A URI for the unit, using controlled vocabularies (e.g., ODM2 Units Vocabulary).	Text	Yes	200	-	-	-
samplingMethodName	The technique or procedure used for sampling.	Text	No	-	-	-	-
sampleFraction	The part of the sample analyzed (e.g., filtered, unfiltered).	Text	Yes	-	-	-	-
samplingMethod_uri	A URI for the sampling method, if available (e.g., from NEMI Methods).	Text	Yes	200	-	-	-
analyticalMethodName	The procedure used for analyzing the sample.	Text	No	-	-	-	-
analyticalMethod_uri	A URI for the analytical method (e.g., from NEMI Methods).	Text	Yes	200	-	-	-
detectionLimits	Specifications of the method's sensitivity.	Text	Yes	-	-	-	-
accuracyBounds	Specifications of the method's accuracy.	Text	Yes	-	-	-	-

DATA

Name	Definition	Type	Allow Nulls	Length	Domain	Default Values	Comments
dataID	A unique identifier for the data record.	Text	No	-	-	-	-
metadataID	Links the data to its corresponding methodological metadata.	Text	No	-	-	-	Should match the `metadata_record_id` in Metadata table.
sampleTimestamp	The date and time when the sample was collected.	DateTime	No	-	-	-	-
resultTimestamp	The date and time when the result was determined.	DateTime	Yes	-	-	-	-
dataValue	The observed or measured value.	Float	No	-	-	-	-
comments	Any additional notes or comments about the data.	Text	Yes	-	-	-	-
dataStatus	Indicates the current status of the data (provisional/final).	Text	No	-	Controlled vocabulary	-	Examples: provisional
dataQuality	An assessment of the data's reliability	Text	Yes				Controlled vocabulary example: (good, fair, bad, unknown, none)

An spreadsheet template for these Internet of Water Recommended Elements for Observational and Modeled Data is available here: https://docs.google.com/spreadsheets/d/1tTUVtCQppMMbEEOM4MpgDxY63_FaqJv2KnbGc_gAP_s/edit?usp=sharing

OTHER APPLICABLE DATA STANDARDS

While the Internet of Water Recommended Elements would be sufficient for most water data analysts to use disparate datasets together, most data standards in active use by water data regulatory or scientific programs require additional metadata, and also do not require persistent URIs for features of interest methods, which hampers discoverability and interoperability. In addition, there has been a general divide between the water data standards used for discrete samples and for time-series data as might be produced by sensors. This is due to a divergence between the large amounts of metadata required for each sample data result by typical regulatory programs (e.g. project, reason, chain-of-custody information), and the inefficiency in storing such result-level metadata for high-frequency time series data. In recognition of this difference, the Internet of Water recommends that discrete sampled data use the US EPA-developed WQX data

standard, and that time series data use the Observations Data Model 2 (ODM2) data standard developed by CUAHSI, augmenting both by adding URIs where available for Site, Feature Of Interest, and Method related metadata. Both of these standards have significant overlap with the Internet of Water Recommended Elements. Both of these standards have well-developed documentation and Excel-based data entry templates linked below:

WQX

Documentation: https://www.epa.gov/sites/default/files/2020-08/documents/wqx_web_template_user_guide.pdf

Data Entry Templates: https://www.epa.gov/waterdata/water-quality-exchange-web-template-files

ODM2

Documentation: https://www.odm2.org

Data Entry Template: https://github.com/ODM2/YODA-File/raw/master/examples/time_series/v0.3.3/YODA_v0.3.3_TS_climate(wHeaders).xlsm