Purpose and Scope of the Extended Hydrofabric Data Model

The Extended Hydrofabric data model was developed to support reproducible Water Management Modeling (WMM) by integrating water management infrastructure (e.g., diversion points, reservoirs) with the national reference hydrofabric through persistent COMID identifiers.

Unlike data inventory systems, this structure is explicitly designed to provide **model-ready, topology-aware data** that can be programmatically integrated into WMMs and coupled with hydrologic models such as the National Hydrologic Model (NHM) and National Water Model (NWM).

The data model consists of:

- Extended geospatial layers (GeoPackage format)
- A normalized relational database (SQLite)
- Persistent linkage between spatial features and time-series records

This document describes the structural design, relationships, and intended modeling use of the Extended Hydrofabric database.

Design Principles

The Extended Hydrofabric database follows the following engineering principles:

1. **Persistent Topological Linkage**
   All water management features are linked to the reference hydrofabric using `COMID`, preserving hydrologic connectivity.
2. **Normalization and Relational Integrity**
   The schema follows normalized relational design to prevent redundancy and support scalable queries.
3. **Automation-Ready Structure**
   Tables are structured to allow programmatic extraction and direct use in model construction workflows.
4. **Basin-Agnostic Architecture**
   The schema can be applied to any basin with sufficient metadata.
5. **Separation of Spatial and Temporal Layers**
   Spatial attributes are stored in the GeoPackage, while dynamic records (time series, rights, rule curves) are stored in the database.

Minimum Metadata Requirements for POD Integration

To enable automated linkage of a Point of Diversion (POD) to the hydrofabric, the following metadata are required:

Required Fields

- Persistent native identifier (e.g., WDID)
- Latitude
- Longitude
- Diversion type (Physical or Aggregated)
- Water source name

Strongly Recommended

- GNIS ID of source stream

Absence of GNIS identifiers increases reliance on fallback matching procedures and may require manual quality assurance.

Hydrofabric Linkage Algorithm (Diversions)

Diversions are linked to the reference hydrofabric using a tiered matching strategy:

1. **Tier 1 – GNIS ID Match**
   If `SOURCE_GNIS_ID` is available, the algorithm selects hydrofabric segments with matching GNIS ID and chooses the closest geometry.
2. **Tier 2 – GNIS Name Match**
   If GNIS ID is unavailable, normalized stream names are matched to hydrofabric `gnis_name` fields.
3. **Tier 3 – Stream Order Fallback**
   If no name-based match is found, the nearest segment with stream order ≥ 3 is selected.

Unmatched or uncertain cases are flagged for manual review.

This approach maximizes network completeness while preserving hydrologic plausibility.

Enhanced Hydrofabric Layers Feature Table

DIVERSION_POINTS

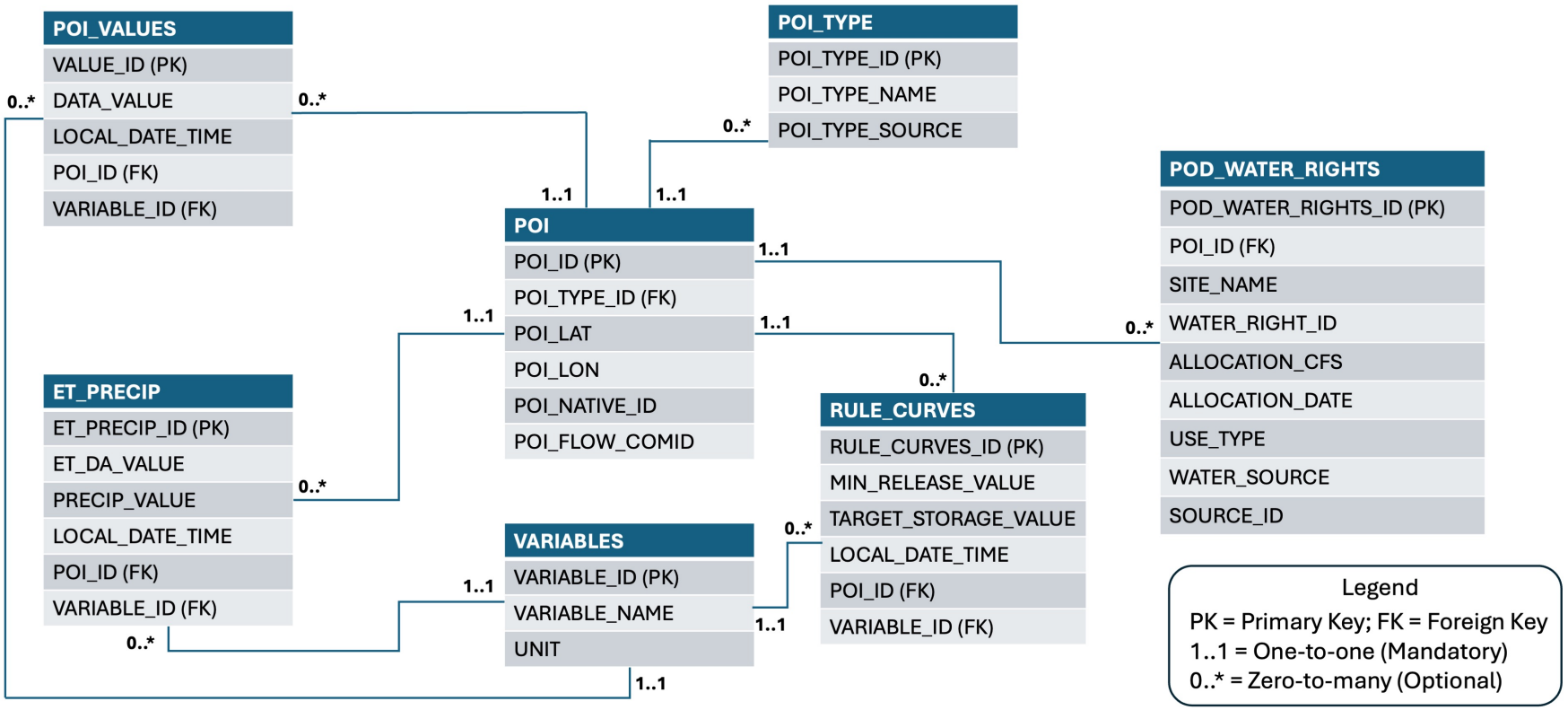
Feature	Description
Feature	Description
WDID	ID of diversion reported from the source
LATITUDE	Latitude of POD reported from the source
LONGITUDE	Longitude of POD reported from the source
TYPE	1. Physical 2. Aggregated Diversion
SITE_NAME	Site name of POD reported from the source
BENEFICIAL_CATEGORY_USE	Designated use for the water right license associated with the POD reported in WaDE
WATER_SOURCE	Name of the stream source of the POD reported from the source
SOURCE_GNIS_ID	GNIS ID of the stream source of the POD derived from the “Water Source”
SOURCE_COMID	“comid” of the closest segment in reference fabric with the same GNIS ID

RESERVOIR_POINTS

Feature	Description
DAM_ID	Unique DAM_ID for each dam in ResOpsUS database
DAM_NAME	Name of the dam
STATE	State where the dam is located
AGENCY_CODE	Unique code for agency reporting data
LONG	The longitude of the dam's location
LAT	The latitude of the dam's location
TIME_SERIES_START	The initial starting date of the time series data for this dam in ResOpsUS database
TIME_SERIES_END	The final date of the time series data for this dam in ResOpsUS database
INCONSISTENCIES_NOTED	Any inconsistency in data reported in ResOpsUS database
NID_ID	Native ID of the dam in National Inventory of Dams
RIVER	Name of impounded river

CAP_MCM	Maximum storage capacity of reservoir in million cubic meters
DAM_HGT_M	Height of dam in meters
AREA_SKM	Surface area of reservoir in square kilometers
MAIN_USE	Main purpose of reservoir: Irrigation; Hydroelectricity; Water supply; Flood control; Recreation; Navigation; Fisheries; Pollution control; Livestock; or Other
SOURCE_COMID	“comid” of the closest and most downstream segment in reference fabric
POI_NATIVE_ID	Native ID of dam reported from the source

Database Schema



Database Tables Documentation

This document contains an explanation of the tables created in the SQLite database along with their structures.

1. POI_Type Table

This table stores information about different types of Points of Interest (POIs).

Column	Data Type	Description
POI_TYPE_ID	INTEGER (Primary Key)	A unique identifier for each type of POI.
POI_TYPE_NAME	TEXT	The name of the POI type.
POI_TYPE_SOURCE	TEXT	The source or origin of the POI type data.

2. POI Table

This table stores individual Points of Interest (POIs), which represent geographical locations of interest.

Column	Data Type	Description
POI_ID (PK)	TEXT (Primary Key)	A unique identifier for each POI.
POI_TYPE_ID (FK)	INTEGER (Foreign Key)	References POI_Type table for type of POI.
POI_LAT	REAL	The latitude of the POI.
POI_LON	REAL	The longitude of the POI.
POI_NATIVE_ID	TEXT	An optional native identifier for the POI.
POI_FLOW_COMID	INTEGER	NHD segment ID associated with the POI.

3. Variables Table

This table stores information about different variables related to the POIs

Column	Data Type	Description

VARIABLE_ID (PK)	INTEGER (Primary Key)	A unique identifier for each variable.
VARIABLE_NAME	TEXT	The name of the variable.
UNIT	TEXT	The unit of measurement for the variable.

4. POI_Values Table

This table stores the actual values of different variables measured or recorded at specific POIs at different times.

Column	Data Type	Description
VALUE_ID (PK)	INTEGER (Primary Key)	A unique identifier for each value record.
DATA_VALUE	REAL	The actual recorded data value.
LOCAL_DATE_TIME	TEXT	The timestamp of the data recording.
POI_ID (FK)	INTEGER (Foreign Key)	References POI table for the POI where data was recorded.
VARIABLE_ID (FK)	INTEGER (Foreign Key)	References Variables table for the variable measured.

5. ET_Precip Table

This table stores evapotranspiration (ET) and precipitation data values for reservoir POIs.

Column	Data Type	Description
ET_PRECIP_ID (PK)	INTEGER (Primary Key)	A unique identifier for each record.
ET_DA_VALUE	REAL	The recorded evapotranspiration value.
PRECIP_VALUE	REAL	The recorded precipitation value.
LOCAL_DATE_TIME	TEXT	The timestamp of the data recording.
POI_ID (FK)	INTEGER (Foreign Key)	References POI table for the POI where data was recorded.
VARIABLE_ID (FK)	INTEGER (Foreign Key)	References Variables table for the variable measured.

6. RULE_CURVES Table

This table stores rules and parameters related to release and storage values for reservoir POIs.

Column	Data Type	Description
RULE_CURVES_ID (PK)	INTEGER (Primary Key)	A unique identifier for each rule.
MIN_RELEASE_VALUE	REAL	The minimum release value from the reservoir or water body.
TARGET_STORAGE_VALUE	REAL	The target storage value for the water body.
LOCAL_DATE_TIME	TEXT	The timestamp for the rule.
POI_ID (FK)	INTEGER (Foreign Key)	References POI table for the POI the rule applies to.
VARIABLE_ID (FK)	INTEGER (Foreign Key)	References Variables table for the variable measured.

7. POD_WaterRights Table

This table stores information about water rights for Points of Diversion (POD) associated with specific POIs.

Column	Data Type	Description
POD_WATER_RIGHTS_ID (PK)	INTEGER (Primary Key)	A unique identifier for each water rights record.
POI_ID (FK)	TEXT (Foreign Key)	References POI table for the POI associated with the water right.
SITE_NAME	TEXT	The name of the water right site.
WATER_RIGHT_ID	REAL	The water allocation in cubic feet per second (CFS).
ALLOCATION_CFS	TEXT	Appropriation date for the water right holder
ALLOCATION_DATE	TEXT	The type of water use.
USE_TYPE	TEXT	The source of the water.
WATER_SOURCE	REAL	An identifier for the water source.

Database Query Examples:

1. To retrieve time-series storage levels for a specific reservoir by POI_NativeID. This SQLite query example fetches the historical storage levels for a given reservoir.

```
SELECT POI_Values.LocalDateTime, POI_Values.DataValue
FROM POI_Values
JOIN POI ON POI_Values.POIID = POI.POIID
JOIN Variables ON POI_Values.VariableID = Variables.VariableID
WHERE POI.POI_NativeID = 'RESERVOIR_ID'
AND Variables.VariableName = 'storage'
ORDER BY POI_Values.LocalDateTime;
```

2. To retrieve time-series demands for a diversion by POI_NativeID. This SQLite query fetches the historical water demand for a specific diversion.

```
SELECT POI_Values.LocalDateTime, POI_Values.DataValue
FROM POI_Values
JOIN POI ON POI_Values.POIID = POI.POIID
JOIN Variables ON POI_Values.VariableID = Variables.VariableID
WHERE POI.POI_NativeID = 'DIVERSION_ID'
AND Variables.VariableName = 'demand'
ORDER BY POI_Values.LocalDateTime;
```

3. To retrieve Allocation_Date for a specific water right by POI_NativeID. This query fetches the date when a particular water right was allocated.

```
SELECT Allocation_Date
FROM POD_WaterRights
JOIN POI ON POD_WaterRights.POIID = POI.POIID
WHERE POI.POI_NativeID = 'WATER_RIGHT_ID';
```

Guide to Simplify Hydrofabric for Water Management Modeling

The hydrofabric simplification process involves filtering a reference flowline dataset to retain only the essential features necessary for water management modeling. The filtering operation iteratively removes flowline segments that do not contribute to downstream connectivity or are not linked to critical water management features such as PODs, reservoirs, gages, or NHM endpoints. This is achieved by checking whether each segment's comid appears in the tocomid (downstream connection) list or is associated with significant hydrological features. If a segment does not meet these criteria, it is removed in each iteration until no further deletions are necessary. Once the filtering is complete, the script creates a simplified version of the original geopackage, preserving all other datasets but replacing the reference_flowline layer with the reduced dataset. This ensures that the essential hydrological structure remains intact while eliminating unnecessary flowlines, reducing computational complexity, and improving the efficiency of the subsequent water management model.