HYPE metadata standard

⚠️*PAGE CONTENTS ARE WORK IN PROGRESS*⚠️

HYPE does formally not require any metadata to be saved along data contents in the various model files. In practice, it makes a modeller's life much easier to store metadata in a structured way along with the actual model files.

Below, we describe a best practice metadata standard in tabular format to be maintained along with the data files in a sub-directory *./meta* of a HYPE model set-up. Metadata tables are tab-separated text files with contents listed row-wise. Missing entries are denoted by NA, non-existing entries by empty strings and -9999 for numeric fields.

Mandatory metadata *denoted in bold face (SMHI-internal requirement).*

metaQobs.txt

One metadata entry per time series in Qobs.txt, detailing data origins, gauge details, and data statistics. Time series can be composed from multiple sources, in this case, only the most important data source is referenced here along with a multi-source flag. Additional sources have to be documented by the modeller in the Qobs.txt pre-processing workflow.

Column name	Unit/Format	Description
subid	*Integer*	ID number for HYPE sub-basins. Links metadata entry to Qobs.txt data file.
country	*Character*	Comma-separated country codes as defined in ISO_3166-1_alpha-2.
rivname_[cc]	*Character*	River name with language code, as extracted from external metadata sources, e.g. naturalearthdata.com
licence	*Integer*	Data licence under which the source data was shared/published as integer ID which relates to a licence table with full agreement details.
src_name	*Character*	Short name of the principal source data base incl. version number, if any. Data series in Qobs.txt can consist of merged data sources. In that case, the most important data source should be noted here, and additional data sources should be tracked in the relevant post-processing folders.
src_statid	*Character*	Station ID used in the principal source data base (type character to allow for alphanumeric source IDs).
src_statname	*Character*	Station name as given in source data base.
src_rivname	*Character*	River name, as given in source data base
multisource	*Logical*	TRUE/FALSE flag to indicate if multiple data sources were used for time series.
src_uarea	*km^2*	Upstream area as provided in principal source data base.
src_xcoord	*Float*	Source data base coordinates, geographic coordinates in decimal notation.
src_ycoord	*Float*	

Column name	Unit/Format	Description
src_epsg	Integer	EPSG code of the source coordinates, as listed on e.g. spatialreference.org
adj_xcoord	Float	Adjusted coordinates, i.e. most probable location on the river. Source coordinates can e.g. indicate gauge hut position, lack detail in decimals, or be plainly wrong. Adjusted coordinates are the accepted coordinates after quality control. 🔧Fix Me! ARE THESE ALWAYS IN THE SAME COORDINATE SYSTEM AS THE SOURCE COORDINATES?
adj_ycoord	Float	
mod_xcoord	Float	Model coordinates, most oftern the relevant SUBID outlet location.
mod_ycoord	Float	
hype_station_id	Integer	Model-internal unique station ID, if used in the within the HYPE set-up (can be useful in large model set-ups with multiple-source observations).
hype_uarea	km^2	Upstream area of sub-basin.
start_date	YYYYMMDD[HH:MM]	
"end_date"	YYYYMMDD[HH:MM]	
"resolution"	Character	Keyord, *day*, *month*, *hour*.
missing	Integer	Number of days/month/hours with missing data between start and end date.
missing_YYYY_YYYY	Integer	Number of days/month/hours with missing data for a focus period (e.g. for calibration).
regulation	Logical	Evidence of upstream regulation in the station's hydrograph.

metaXobs_[variable].txt

Column name	Unit/Format	Description

metaGeoData.txt

Column name	Unit/Format	Description