A Metadata Hierarchy for Enhanced Management of Hydro-Numerical Simulation Data

Mohammad Shafi Arif, Federal Waterways Engineering and Research Institute, (Germany),
mohammad-shafi.arif@baw.de
Peter Schade, Federal Waterways Engineering and Research Institute (Germany),
peter.schade@baw.de
Frank Kösters, Federal Waterways Engineering and Research Institute (Germany),
frank.koesters@baw.de

In hydraulic engineering numerical simulations are commonly conducted for a better understanding of coastal and estuarine physics. Moreover, hydro-numerical simulations are carried out in the framework of environmental impact assessments, e.g. to evaluate the impact of the construction of coastal structures or the adaptation of approach channels on estuarine hydrodynamics and morphodynamics. These assessments require a large number of simulations, e.g. the model calibration for different scenarios covering natural variability. Individual model runs may differ in some minor variation of parameters only, which makes it difficult for the user to keep track of differences among models.

In terms of quality assurance a reliable documentation is therefore required but can be tedious and therefore error prone. Here the incentives of digitization and advantages of web-enabled geospatial services, as well as FAIR (Findable, Accessible, Interoperable, and Reusable) data management principles come into play with the utilization of metadata. Consistent and detailed documentation is of importance for both storage and reproduction of data as well as services such as searching, filtering and cataloguing. In spite of numerous benefits which metadata can offer, manual recording the parametrization attributes as metadata with a complex data structure is considered challenging, i.e. labor intensive and user unfriendly. Hence, the provision of an adequate data structure as well as a partial-automation for the recording is considered as a prerequisite in order to achieve reliable metadata.

Addressing the complexity for data management, the research and development project Data Management and Quality Assurance in hydraulic engineering (DMQS) has been initiated at the Federal Waterways Engineering and Research Institute (BAW). A hierarchical metadata management approach was chosen for the data organization in a simulation project, which is depicted in Figure 1 as a detailed tree diagram for a project with an integrated data and metadata organization.

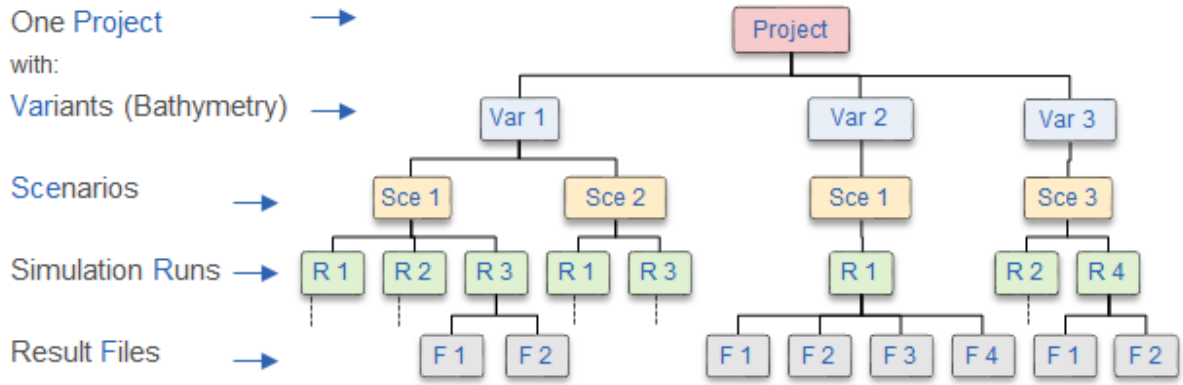

Figure 1: Metadata hierarchy in a tree structure for simulation projects.

Five levels of metadata are used with the project as the root element on top of the hierarchy. This method highly mitigates the impediments to the data management and to metadata organization in a metadata information system. The root is followed by the variants, one of the main features of hydro-numerical simulations. Variants are defining the bathymetry of the model area, consisting of a horizontal mesh and the water depth. Variants are followed by scenarios defining the parametrization e.g. the hydraulic roughness or the amount of head water. The subsequent hierarchy level is represented by the individual simulation run. Each run can generate one or more result files, which are distinguished by their Universally Unique Identifiers (UUIDs).

Aiming at a reduction of the efforts associated with metadata recording and generation, a three step approach has been chosen. At first, while starting a new project, the user is required to specify a few project related elements such as contact information just once. Pursuing this further, the user has to fill out only four meta elements related to the simulation run. At last, the rest is acquired from input data for the numerical simulations, which are mostly proprietary meta elements. For instance, the bounding box is computed by analyzing the extent of the input bathymetry, and the post processing tools protocol their lineage information so that the workflow up to the final analysis is retraceable.

All these metadata are then written and stored in Network Common Data Format (NetCDF) files as extended Climate and Forecast (CF) metadata. Later on, the NetCDF CF metadata is converted to XML (Extensible Markup Language) data based on ISO schemas, INSPIRE technical guidelines and the BAW metadata profile (GDI-BAW 1.3). The use of UUIDs represent a reliable method to identify the result files. This allows establishing the tree structure from automatically generated UUIDs on all levels.The data aggregate elements of ISO 19115 facilities the cataloguing tree as well. The UUIDs are reproducible so that later simulation runs can be linked to the same project or variant. Metadata titles of the level are concatenated to a string that is used as input to the UUID generation. Figure 2 depicts the use and application of UUIDs as parent identifiers.

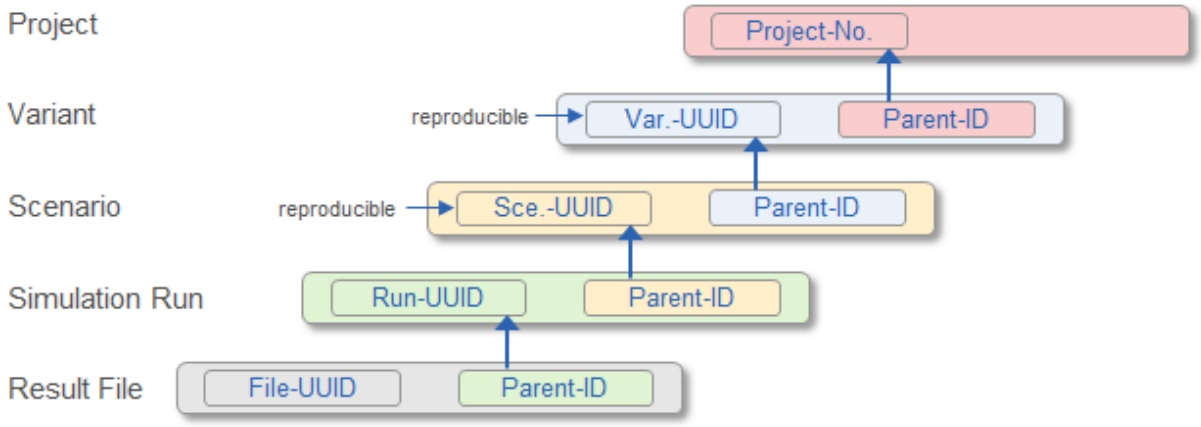

Figure 2: Application of UUIDs as parent identifiers.

To conclude, the proposed automation method has tackled the metadata generation bottleneck of a rival manual recording by using the metadata of NetCDF simulation files. Workflow automation has reduced erroneous metadata recording and enhanced reproduction as well as management of the metadata. Through the implementation of a metadata hierarchy users can better keep track of simulation results, especially of those from simulations with small differences. The convergence of CF conventions and ISO standards ensures interoperability and consistency for emerging data exchange.

Acknowledgements

The authors would like to thank Georg Carstens, Vikram Notay, Günther Lang and Shamimul Islam for their contributions towards the development of the DMQS project.