BAW
Federal Waterways Engineering and Research Institute

Competence for Waterways

Metadata Hierarchy

for Enhanced Management of Hydro-Numerical Simulation Data

Background

Numerical simulations are conducted for a better understanding of coastal and estuarine physics as well as environmental impact assessments. These assessments require a large number of simulations, e.g. for different scenarios covering natural variability. Individual numerical model runs may only differ in some minor variation of parameters, which makes it difficult for the user to keep track of differences among models or reproduce the model. On the other hand, a reliable documentation is the key to data quality assurance in a data management approach, which can be tedious and error prone.

Motivation

The aim of this work is to obtain a metadata management system in order to:

1) Utilize the benefits of digitization and advantages of web-enabled geospatial services such as searching, filtering and cataloguing,
2) Achieve a FAIR (Findable, Accessible, Interoperable, and Reusable) data management,
3) Guarantee the quality of metadata via consistent, standardized and detailed documentation, and
4) Attain a reliable metadata for data storage and reproduction.

Method

Addressing the complexity of data management for a simulation project, the R&D project Data Management and Quality Assurance in hydraulic engineering (DMQS) has introduced and initiated a hierarchical metadata management (HMM) approach at the Federal Waterways Engineering and Research Institute (BAW), which is depicted in Figure 1. Five levels of metadata are used:

1) The project, as the root element on top of the hierarchy,
2) The variants, defining the bathymetry of the model area,
3) The scenarios, defining the parametrization e.g. the hydraulic roughness,
4) The simulation run, representing further distinct parameters, and
5) The result files, which are distinguished by their Universally Unique Identifiers (UUID).

These levels of hierarchy for Metadata are further depicted in Figure 2.

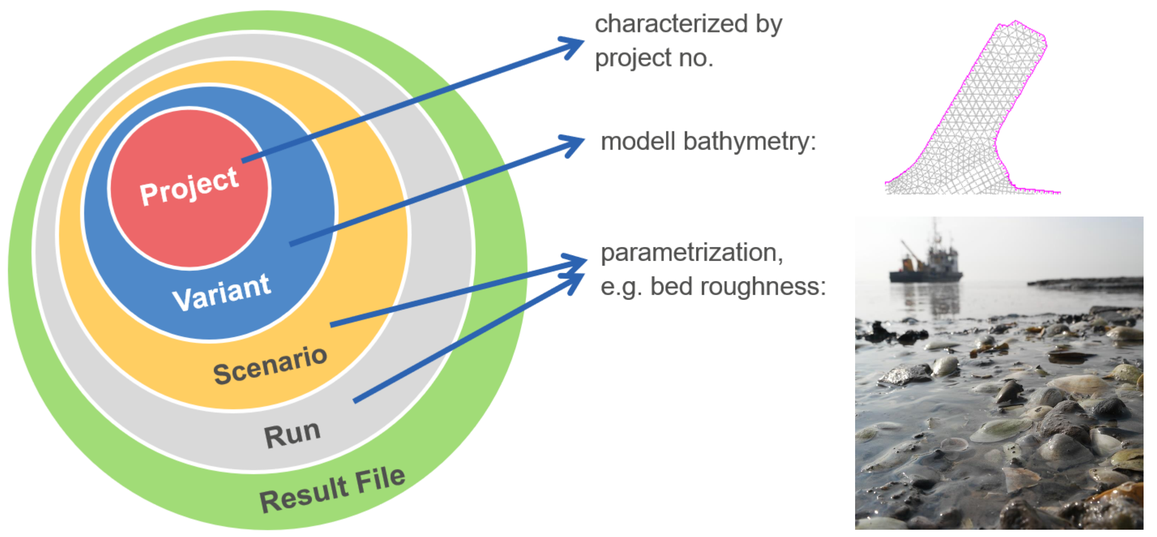

Figure 1: Hierarchy for simulation projects.

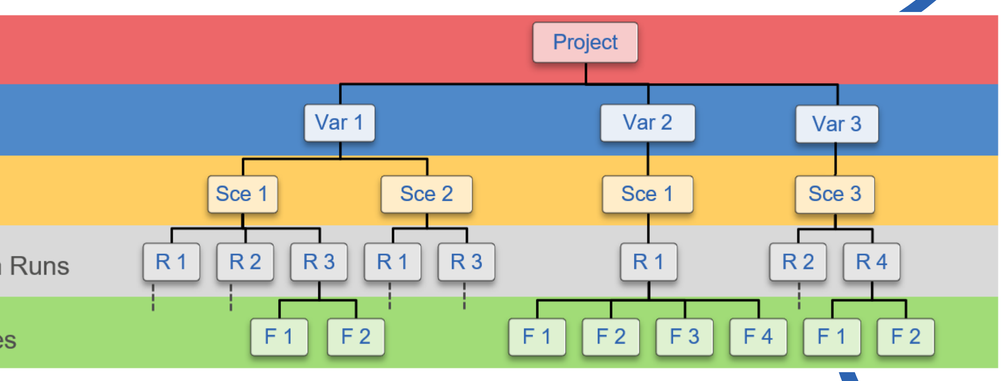

Figure 2: Metadata hierarchy in a tree structure for simulation projects.

Aiming at a reduction of the efforts associated with metadata recording and generation, a three-step approach was proposed:

- While starting a project: specify a few project related elements such as contact information.
- Before running a simulation: fill out only four meta elements.
- The rest of the parameters, e.g. the bounding box: computed from input numerical simulation file variables.

This way, 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 Form (NetCDF) files as extended Climate and Forecast (CF) metadata.

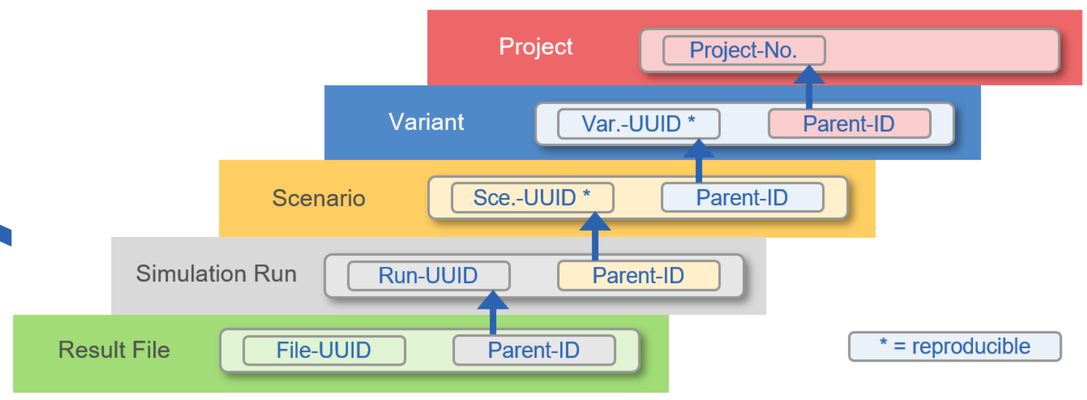

Figure 3: Application of UUIDs as parent identifiers.

Results

The NetCDF CF metadata which are gathered via the method mentioned above is then converted to XML (Extensible Markup Language) data based on ISO schemas, INSPIRE technical guidelines and the BAW metadata profile (GDI-BAW 1.3). For enhanced interoperability the BAW metadata profile then further controls the use of domain ontologies via the code lists integrated into this profile. The use of UUIDs in an HMM system represents a reliable method to identify the result files especially of those from simulations with small differences and allows establishing the tree structure. Metadata titles of the level are concatenated to a string that is used as input to the UUID generation. Figure 3 depicts the use and application of UUIDs as parent identifiers (`gmd:parentIdentifier` and `gmd:hierarchyLevel`) through which the metadata hierarchy is built in a metadata information system (MIS). This HMM is also able to construct a hierarchical catalogue view directly from the imported metadata sets.

Conclusion

It is found that an integrated data and metadata organization is achievable via the proposed hierarchical metadata management method following the ancillary automation. Further advantages of the HMM approach are:

- Reproducibility as well as management of the data and metadata
- Reducing erroneous metadata and deciphering the bottleneck of a rival manually recording
- Mitigating the impediments to the data management and to metadata organization in a metadata information system.

The automatic generation of XMLs will be described via upcoming works.

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.

Bibliography:
FAIR Principles, viewed on 05.05.2020, <https://www.go-fair.org/fair-principles/>
CF Conventions and Metadata, viewed on 15.08.2019, < https://cfconventions.org/>

April 2021 · Authors: Mohammad Shafi Arif, Peter Schade, Frank Kösters · Department of Hydraulic Engineering in Coastal Areas · daten-k@baw.de