Containerterminal Bremerhaven in Weser Estuary; 2019

BAW
Federal Waterways Engineering and Research Institute

Peter Schade
Mohammad-Shafi Arif
Frank Kösters

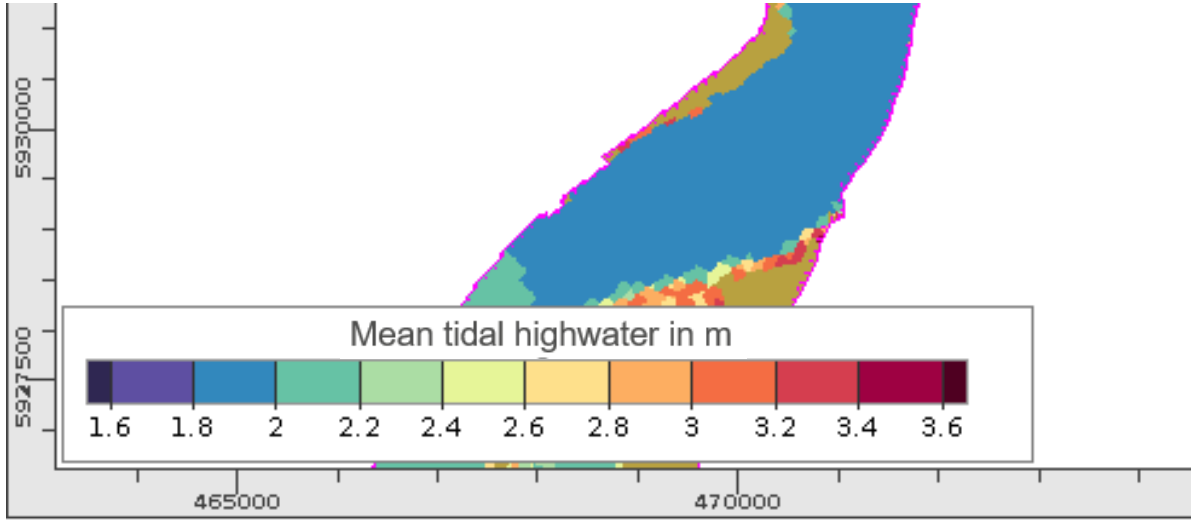
Automation of workflows for numerical simulation data and metadata

Data Science Symposium 2023 at Geomar

Kiel, 9. June 2023

Use case: precise finding of datasets

- Supposed you have a figure in a paper and would like to work with the dataset:
  - redraw it with new colour table or
  - create a figure of another variable or
  - reproduce the data



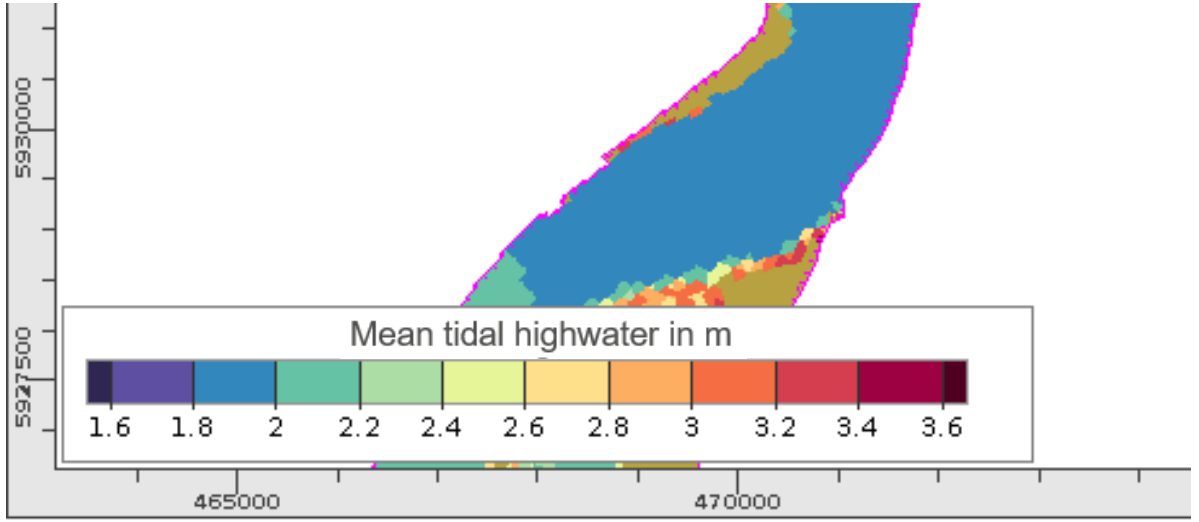
- Use a portal to find the dataset.
  - But if you search for “Weser“ and “highwater“, you may get hundreds of hits.
  - Even after precise finding reproducability is not guaranteed ➡ more detailed metadata (MD)

Use case: precise finding of datasets

- Supposed you have a figure in a paper and would like to work with the dataset:
  - redraw it with new colour table or
  - create a figure of another variable or
  - reproduce the data

- Solution by using Universally Unique Identifiers:
  - Mandatory to print the NetCDF UUID in figures.
  - Use UUID as Object-ID in metadata of a portal.
  - PIDs not suited since not known when the figure is drawn.
  - Implemented in an automated MD workflow.

- As user of a portal:
  - Search for UUID and find exactly one file.
  - Download files with detailed MD.

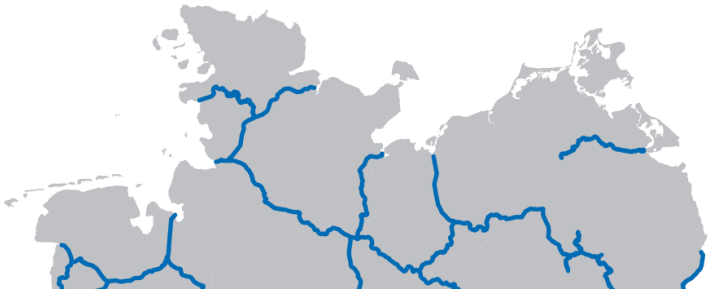


Content

1. Introduction into numerical simulations at BAW
2. How to record the history of a simulation workflow
3. Workflow for archiving data and metadata
4. Conclusions

Introduction into numerical simulations at the coastal department of BAW

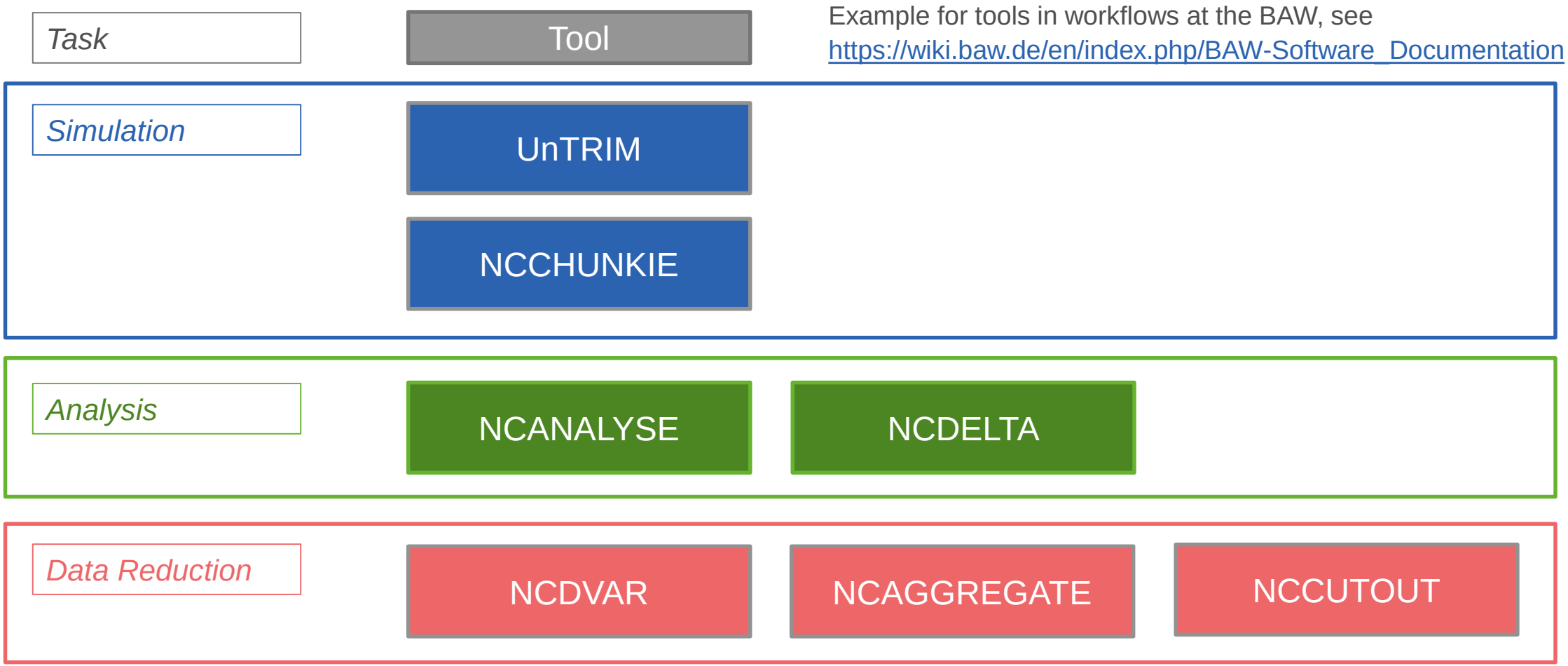
- Expertise and R&D, e.g., for adaption of fairways to seaports in North Sea and Baltic Sea.
- Assessment of physical processes such as hydrodynamics and sediment transport.
- High level of QA is needed due to legal aspects of expertise.



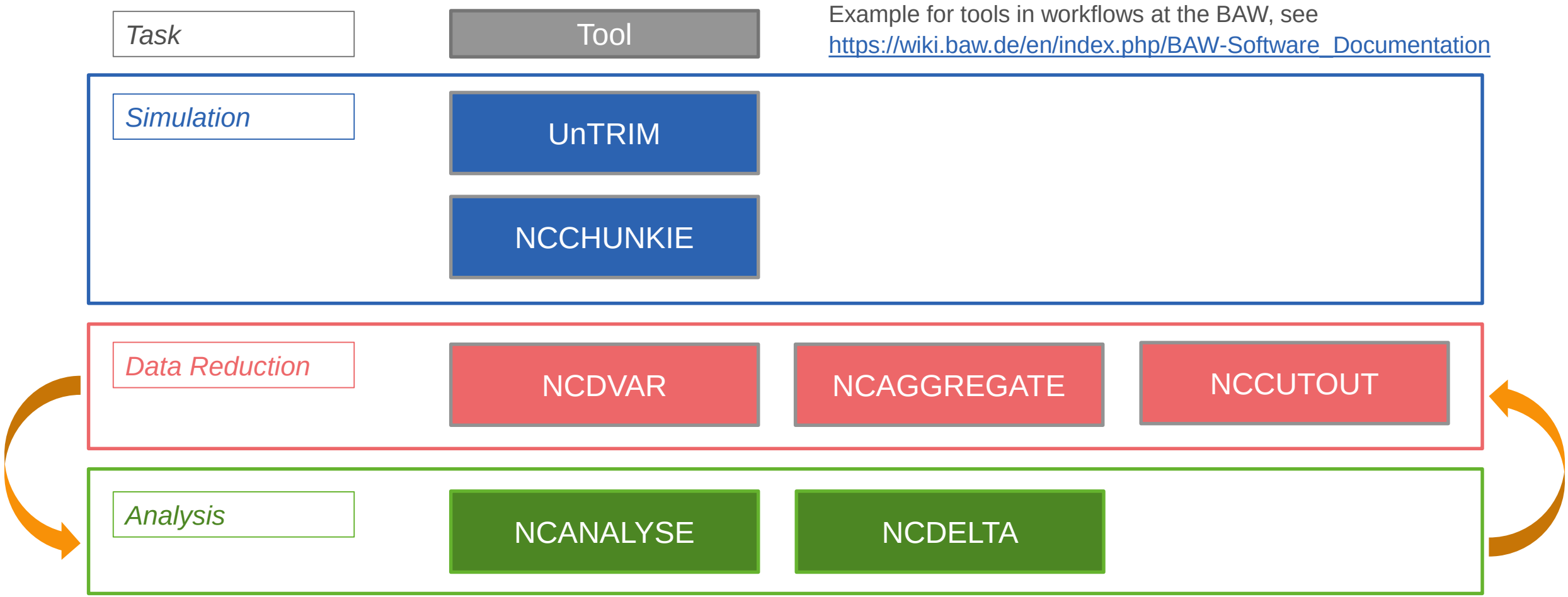
- Expertise includes forecasts, which are enabled by numerical simulation of the processes mentioned above and require:
  - High Performance Computing (HPC)
  - Numerical models by the current state of scientific knowledge
  - Staff with adequate skills
  - High reliability, e.g., for reproducing results.

- Automated workflows for data and metadata are a means of quality assurance.

How to record the history of a simulation workflow

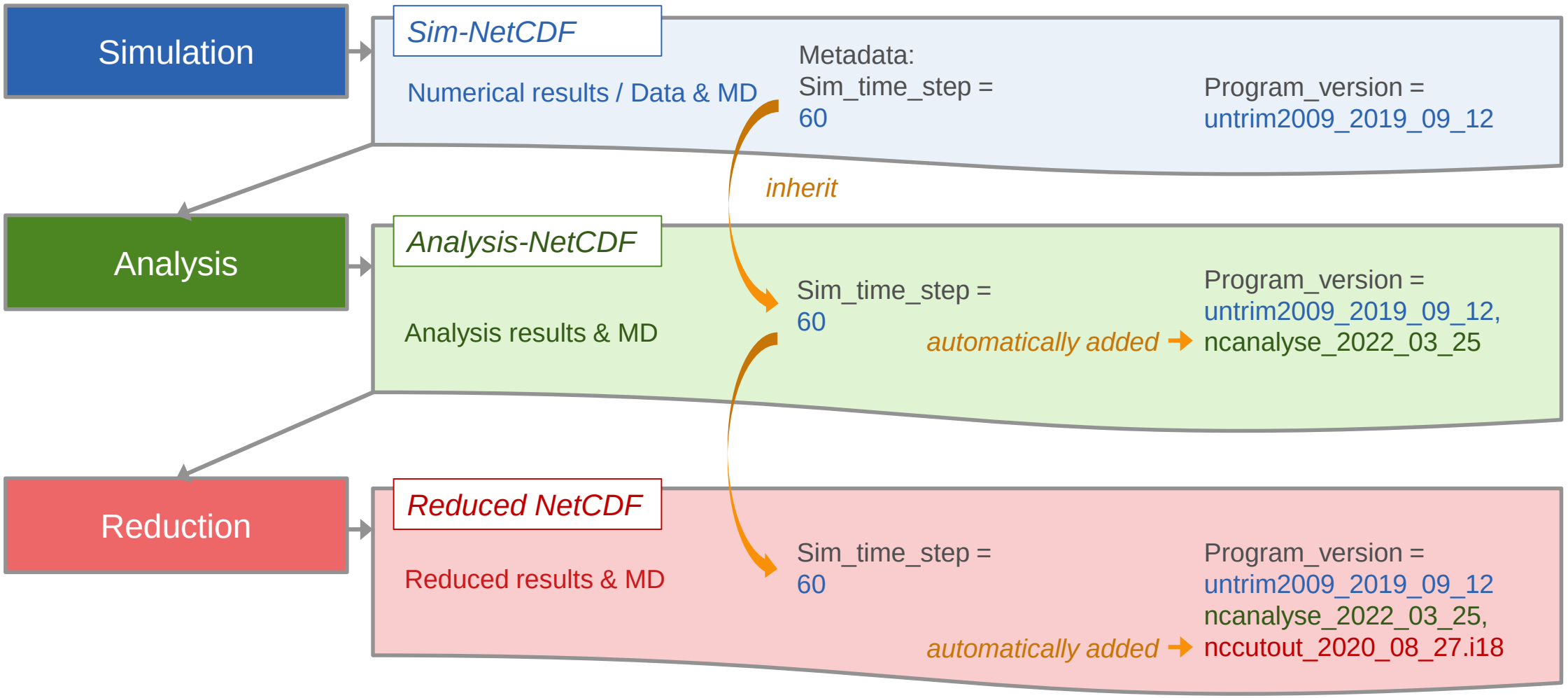


How to record the history of a simulation workflow

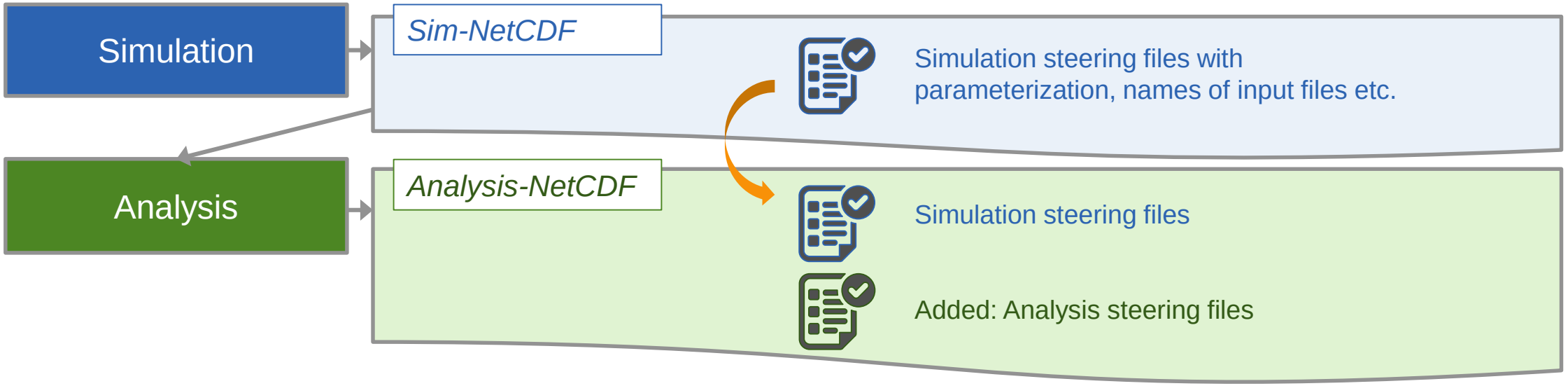


- How can you record these flexible workflows ?

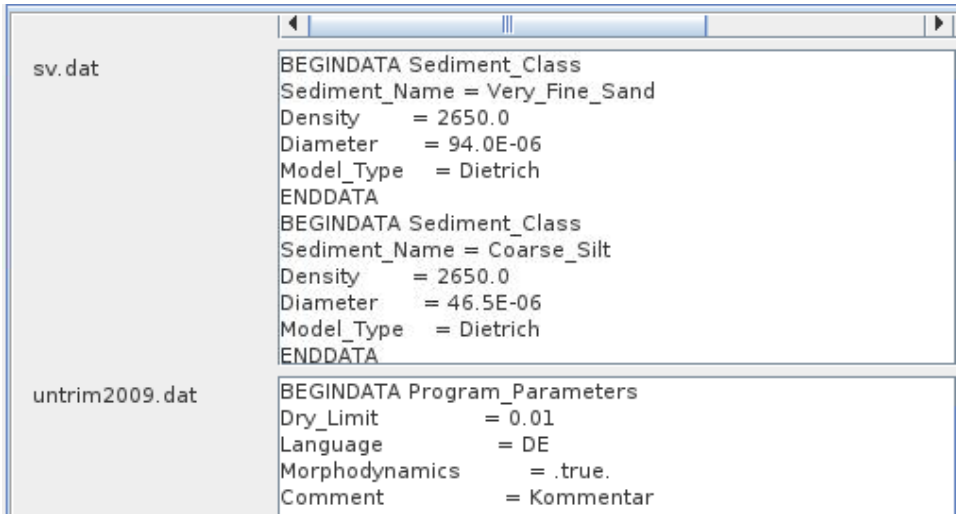
Each tool stores results and simulation-specific metadata in one CF-NetCDF file



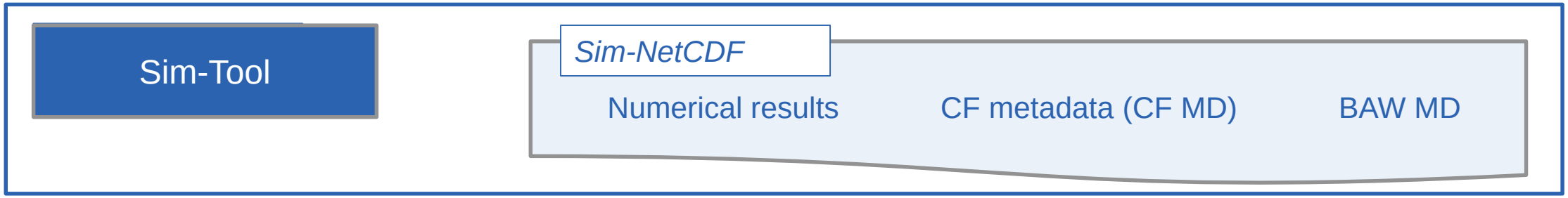
Storing steering file content in CF-NetCDF files



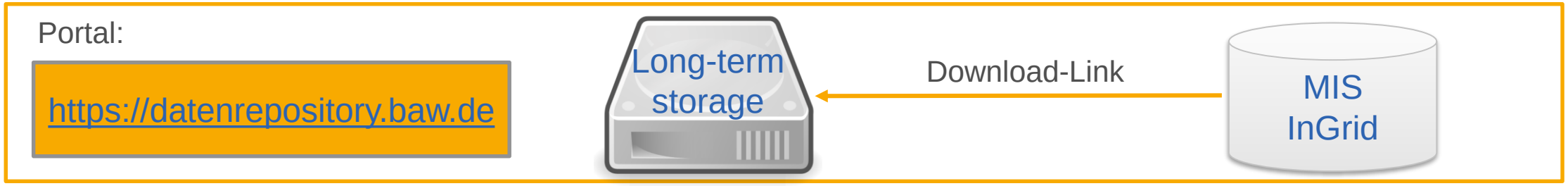
- Postprocessors, which “understand” BAW result files, give you access to the steering file contents of the complete processing chain.
- Reproduce processing even after years.



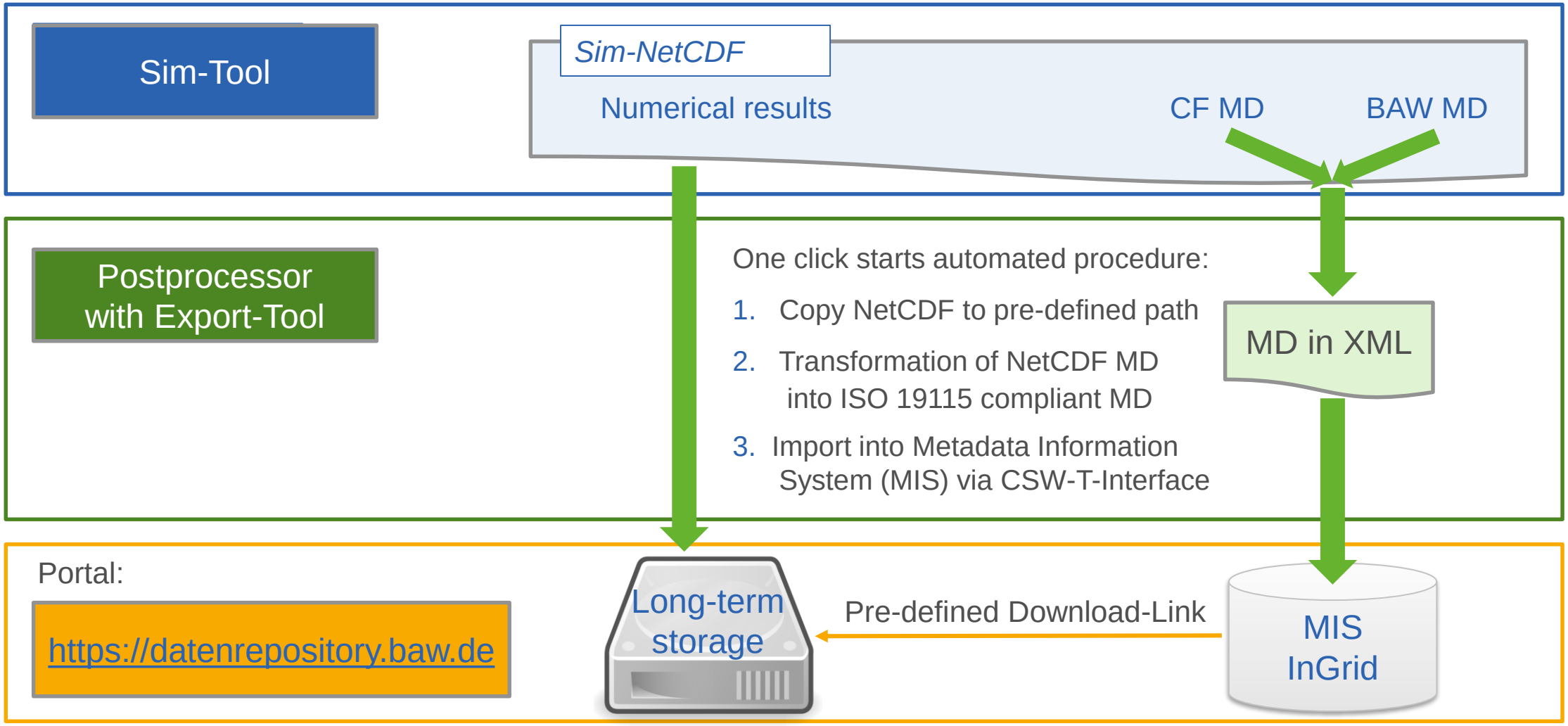
New workflow for archiving data and metadata



How can you automatically import the data and metadata into a portal?



New workflow for archiving data and metadata with an Export-Tool



Conclusions

- How to record the history of a simulation workflow:
  - Each tool in a processing chain inherits metadata from the former tool and adds its own.
  - Generic, so that the recording works for flexible workflows.

- Archiving data and metadata in a portal, e.g., https://datenrepository.baw.de:
  - Automatic transformation from NetCDF MD into ISO 19115 compliant MD.
  - Compared to manual editing less error-prone and more efficient.

- Use standardized metadata, but be creative with the possibilities:
  - UUIDs in figures.
  - Store steering file content in result files.
  - Your ideas: Other complex metadata to be stored in result files?

References

- Mohammad Shafi Arif et al (2023): From simulation to dissemination: automation of data and metadata management; In 2023 IOP Conference Ser.: Earth Environmental Science **1136** 012006 https://iopscience.iop.org/article/10.1088/1755-1315/1136/1/012006
- Bundesanstalt für Wasserbau: Portal https://datenrepository.baw.de
- Bundesanstalt für Wasserbau: Descriptions of BAW programs. Online available https://wiki.baw.de/en/index.php/Program_Descriptions, on 26.06.2023 recently checked.
- Bundesanstalt für Wasserbau (2022): Description of processor for automated archiving. Online available https://wiki.baw.de/en/index.php/DAVIT, on 26.06.2023 recently checked.
- Bundesanstalt für Wasserbau (2022): Description of simulation program. Online available https://wiki.baw.de/en/index.php/UNTRIM2, on 26.06.2023 recently checked.

Outer Weser; 2019

Thank you for your attention!

Bundesanstalt für Wasserbau
22559 Hamburg, Germany

www.baw.de