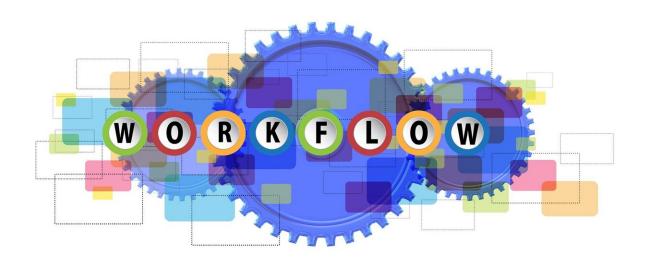# New workflow tool tackles modelling hurdles of data, scenario and collaboration management

December 10 2021

Credit: Pixabay/CC0 Public Domain

Controlling large amounts of data, running a multitude of scenarios and sharing information between a number of modelers are common challenges in modeling endeavors regardless of the field—be it energy systems, process design or epidemic modeling. Even when using state-of-the-art modeling tools, problems in pre- and postprocessing, sharing and maintaining different versions of data can reduce the efficiency and impair the quality of projects. Open-source workflow management tool

Spine Toolbox developed in a four-year EU-project focused on complex data handling, ease of scenario building as well as remote execution and division of labor within a modeling team. This allows using it in many fields even if the same data should be fed into multiple models in the same workflow.

Ease of building scenarios and complex workflows for better decision making

When models are used to evaluate future options and to understand complex systems in any domain, accounting for uncertainties is often a key factor for reliable and repeatable modeling. While conventional generic workflow tools can be powerful for executing tool chains, additional data management capabilities can be important: Firstly, support for creating and comparing scenarios—and secondly, support for managing not just data but also arbitrary data structures.

For scenario work, the capability to manage alternative values for data parameters, using them to build scenarios and comparing them systematically can improve the modeling process and the management of sensitivities in the input data. For complex workflows with several models that use partially the same data, data has to be converted into model specific formats and structures. This can be better facilitated if data includes structural information like relationships between entities and classes that categorize entities. Spine Toolbox stores data in SQL with a graph-like structure that allows storing and editing not just data but also those relationships within the data.

Workflow collaboration requires remote execution, ease-of-use and flexibility

A third starting point in Toolbox development was the ease of

collaboration. Spine Toolbox has graphical interfaces for managing and editing data, editing the workflow and for importing and exporting tabulated data. Any particular workflow can be a local project, but it can also be shared through a shared git-repository. Local workflows can also include shared elements, like databases or tools from git-repositories. This allows for flexible division of labor within modeling teams. Workflows or parts of workflow can be executed locally or using a remote server that has better computational capabilities. Parallelization can also speed-up the modeling process and Toolbox supports this not just across tools but also across scenarios and sensitivity runs.

Advanced tools offer quickly prototyping new models and linking with commercial projects

While workflows and data can be edited by regular users, Spine Toolbox has additional features under the hood for development oriented users. It's written in Python to allow for easy integration of Python based tools that are widespread in the research community. Furthermore, SpineInterface package allows quickly building and testing new optimization models using Toolbox and Julia/JuMP. All data and data structures from Spine databases can be directly used when writing equations for optimization models. Spine Toolbox is fully open-source, and is also available for commercial utilization and linking with commercial models. While the tool is recently published and continuously developing, it lacks many specific data processing capabilities present in the more mature tools. However, Spine Toolbox workflows can incorporate other data processing tools available in the open source community.

Efficient workflow management wastes time in the short run and saves it in the long run

The developers of Spine Toolbox include experienced energy system modelers, who have built own or projects specific workflow management systems and used existing tools. They have first hand experience of what can go wrong in complex projects with many collaborators, wasting time on non-essential parts of the project. As the resources to design and implement a workflow tool for betted data and scenario management, they set out to develop a tool that researchers, engineers and project managers could spend a few weeks deploying, and save it many times over in future years not having to spend time with imprecise data management.

  More information: Read more about model functionalities, installation and technical properties: github.com/Spine-project

Provided by VTT Technical Research Centre of Finland