

Templates for reproducible science

Thursday, 3 October 2024 12:30 (30 minutes)

Software engineering best practices favour the creation of better quality projects, where similar projects should originate from similar layout, also called software templates. This approach greatly enhances project comprehension and reduces developers' effort in the implementation of high-quality code. As an example, reproducibility and reusability are the key aspects of this software engineering process, the use of packaging tools and containers is a common practice to achieve robustness and portability for long-term software maintenance. However, these tools are not always easy to use and require a certain level of expertise to implement from scratch. Software templates are known to be an excellent way to reduce the complexity load on the use of such tools on the developer's side.

There exist various tools to create such templates and routinely generate projects from them. One such Open Source tool is cookiecutter [1], a cross-platform command-line utility where a new project is replicated according to a set of files and directories that are pre-configured to provide the base structure. These templates, or cookiecutters, can be re-used and freely hosted on software version control platforms e.g. GitHub, where customization is achieved by filling in placeholders in the template files using project-specific values.

In this contribution, we show you how to develop custom modules within the AI4OS dashboard and make use of the best software development practices for this scientific framework. We present a new service that provides a collection of best templates on a marketplace/hub and use them to generate new projects on-fly through a web interface without requiring the installation of the cookiecutter tool on the client side. The platform features a GitHub repository to collect metadata about templates, a python-based backend, and a javascript Web GUI with the authentication via EGI Check-In. From data preprocessing to model evaluation, this session covers the most critical steps in the process of module development, empowering participants to achieve the desired reproducibility and reusability in their projects.

[1] <https://github.com/cookiecutter/cookiecutter>

Topic

EOSC Developments and Open Science: Reproducible Open Science

Primary authors: ESTEBAN SANCHIS, Borja (Karlsruhe Institute of Technology); LAURES, Christophe; Dr HARDT, Marcus (KIT-G); Dr KOZLOV, Valentin (Karlsruhe Institute of Technology)

Presenter: ESTEBAN SANCHIS, Borja (Karlsruhe Institute of Technology)

Session Classification: Demonstrations & Posters