OSSI Proposal – Development and Maintenance of Dacapo

DaCapo is an open-source machine learning framework for the training and application of neural networks on large, multi-dimensional bioimage datasets. The framework facilitates the whole life-cycle of machine learning applications from data loading, architecture and loss selection, training, evaluation, to the application of a trained model on new data. DaCapo has a modular design that allows to combine different datasets, tasks, architectures, and trainers. Training and prediction runs can be submitted to a compute cluster to take advantage of parallel execution on multiple GPU nodes.

DaCapo has so far been successfully used in the Janelia CellMap project, mostly for the semantic segmentation of cell organelles in large FIB-SEM datasets. Here, we propose to further develop and maintain DaCapo to make it available to (1) the Janelia community, especially experimental labs with little coding or machine learning experience, and (2) the wider life science community.

The objectives for the OSSI initiative are:

1. **Simplify the execution of an experiment:** We aim to bring DaCapo to labs with little coding and machine learning experience. To that end, DaCapo should be entirely runnable by providing configuration files (selection of the datasets, task, architecture, etc.) and running a simple command line tool. Ultimately, we envision a web-based interface to create those configuration files, run experiments, and monitor their progress.
2. **Write documentation:** This entails both the documentation of DaCapo's API for developers and contributors, as well as the writing of tutorials and a best-practice guide for common image analysis tasks.
3. **Respond to Community Feature Requests:** DaCapo's modular design allows incorporating new model architectures, segmentation strategies, and post-processing algorithms. Based on requests from the community, we aim to support some of those requests as core modules in DaCapo, if those modules are general enough to be useful for the wider community.
4. **Cloud execution:** To make DaCapo more accessible to the wider scientific community we will work on capabilities for straight-forward deployment on cloud infrastructure.

We ask for funding to contract William Patton (Scientific Software Engineer, freelancer) for one year, with a weekly commitment of 10h and compensation of $100/h ($52,000 total). The amount includes potential help from SciCompSoft and other contractors.