Data-Driven Modeling of Brain Circuits and the SONATA Data Format

Anton Arkhipov, Allen Institute, Seattle, USA and **Eilif Muller**, Blue Brain Project, Switzerland

The ever-expanding amounts of data from experimental neuroscience and increasing computing power have made it possible to build and simulate large-scale, intricately complex models of brain circuits that contain thousands to millions of neurons and a wide range of detailed information about cells, cellular compartments, and synapses. In this talk, we will highlight the directions of research in large-scale modeling pursued at the Allen Institute and the Blue Brain Project and will describe a path forward for standardization and sharing of such models. A significant bottleneck for scientists developing large-scale network models is the lack of efficient data formats. An open specification is needed that is compact, computationally fast, yet also easy to read and edit. To meet these demands, the Allen Institute and the Blue Brain Project have jointly developed the SONATA (Scalable Open Network Architecture TemplAte) Data Format - an open-source framework for representing neuronal circuits. The framework utilizes both organizations' expertise with large-scale high-performance computing and network simulations, visualization, and analysis. It was designed for memory and computational efficiency. The format represents nodes (cells) and edges (synapses/junctions) of a network using table-based data structures, hdf5 and csv, employs a systematic schema for storing simulation output in hdf5-based format (including variables like spike times, membrane potential, and Ca2+ concentration), and relies on JSON-based specifications for configuring simulations. The resulting set of standards and tools permits efficient representation of large-scale models and enables sharing of results between different labs. The SONATA Data Format and framework (https://github.com/AllenInstitute/sonata) are open to the community to use and build upon with the goal of achieving greater transparency, collaboration, and reproducibility in the field.