

OFES Data Sharing Expectations for Large-Scale Computational Projects

Introduction and Background

Large scale computations—including projects supported by the Office of Science’s Scientific Discovery through Advanced Computing (SciDAC) program—have increasingly assumed a vital role in addressing the scientific challenges facing the U.S. Fusion Energy Sciences Program as equal and complementary partners with the more traditional approaches of analytic theory and experiment. High performance simulation codes developed under the auspices of SciDAC and the OFES base program have provided us with new and significant insights into the physical mechanisms responsible for turbulent transport from the core and edge of magnetically confined plasmas, the interaction of electromagnetic waves with plasmas, and the mechanisms responsible for the macroscopic stability of present and next generation fusion devices. In the coming era of burning plasmas and ITER, large scale integrated simulations will be essential for the design, operation, and interpretation of results from future experiments.

With such responsibility comes the challenging task of verifying and validating the predictions of these codes. Considering the difficulty and significant computational expense of replicating results from massively parallel codes as well as the inherent limitations of analytic theory as a comprehensive verification approach for the highly nonlinear problems addressed by these simulations, cross-benchmarking among different codes is an indispensable and often-used verification tool which also leverages the OFES investment in the development of codes based on different technical approaches. The success of this approach relies heavily on the availability and sharing of simulation data and other supporting materials in a timely fashion and at no more than incremental cost by investigators engaged in large-scale simulation research.

Data Sharing Guidelines

For the reasons stated above, the OFES expects the timely sharing of simulation data among OFES-funded researchers engaging in large-scale computational research. At the same time, OFES recognizes the right of individual scientists and research groups to get fair credit for their work by establishing initial periods of *exclusive use*, defined as the period between the generation of the simulation data and publication and/or presentation of research results based on these data. During the period of exclusive use, PIs are not required to share their data with others.

In more detail:

- OFES strongly encourages the PIs of its large-scale computational projects to ensure that: (i) their most important findings be published in a timely way in peer-reviewed journals; and (ii) their results be presented at major conferences and workshops that are widely attended by the members of our community.
- Following publication in a peer reviewed journal, or after a year following a major conference presentation—whichever comes first—simulation data should be available for sharing upon request

- Users of shared data should consult with the donating author and try to reach a consensus on any technical issues pertaining to these data and the simulations that generated them before publishing or presenting results based on these data
- Users of shared data are expected to give proper credit to the researchers that generated them in any publication or presentation that makes use of these data
- Providing data and assisting in their interpretation can be time and resource consuming, especially for smaller research groups. Those requesting data should be aware of this fact and be reasonable in their requests, including being prepared to possibly share in the associated costs and labor
- Research teams involved in large-scale simulations should designate a point-of-contact person responsible for data sharing issues
- In the event that data sharing disputes or issues emerge, OFES will work to help resolve associated problems in consultation with the investigators