Data Management Plan

This plan will make certain that the data produced during the period of this project is appropriately managed to ensure its usability, access, and preservation. The data produced by the proposed project will consist of CI theory, new software, good practices for CI training, course materials, and program guidelines.

Publications and Lectures. The goal of this project is training CICs and CIUs to contribute to research, and we expect original research to arise in the training process. The participants, including (co-)PIs, senior personnel, external advisors, and students, will disseminate the results of their theoretical discoveries, their computational investigations, and their new insights into CI education as early as appropriate in the form of peer-reviewed journal articles, conference abstracts, and lectures at various conferences and institutions. Authorship will accurately reflect the contributions of those involved. Students will be particularly encouraged to publish their work. When allowed by publishers, pre-prints of publications will be posted on arXiv.

Software. Software packages of libraries resulting from this project will be stored on public repositories, such as GitHub, and made available for adoption and improvement by others. This is the practice already with our Guaranteed Integration Library (GAIL) [7] and many major software libraries developed by others. Software may also be published through ACM-TOMS and similar journals. Our software developments will be publicized through colloquium and conference talks and e-newsletters such as the NA-Digest.

Course Materials. Lecture notes and example code developed for our key courses, including the summer computational science course, the crash course for CoD students, the new professional practices course, the undergraduate parallel and distributed computing course, and the large-scale computation course, will be made available on public repositories such as GitHub or Google Drive.

Web Publication. The CISC website will serve as an index to the data generated by this project. This will include pointers to publications arising from this project, software arising out of this project, and course materials, as mentioned above.

We find that sample or template code, e.g., demonstrating how to run a job on a cluster, how to run a job utilizing multiple cores, or how to run a job based on a specialized package, are useful teaching devices. We will store these samples on public repositories and provide pointers to them on the CISC website.

Information about to our initiatives, including goals, policies, and benchmarks will also be available on the CISC website.

The best practices that we discover disseminated on public computational science forums, such as the newly established Better Scientific Software [5].

We will fully comply with all applicable guidelines and policies on model and data sharing as mandated or recommended by NSF. This Data Management Plan addresses NSF's policy on the dissemination and sharing of research results within a reasonable time. In accordance with this policy, this plan does not include preliminary analyses (including raw data), drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues.