

Example data management plan

In the spirit of open scientific inquiry, all data produced in this project will be stored in standard, open formats and made freely available on the Internet under appropriate open-source licenses, and using standard, open distribution technology.

1. Types of data. The project team will produce data from experimental observations and numerical simulations of a wide variety of time-evolving, spatially extended fields in thermal boundary layers and piston engines. Software will include high-performance computing codes for boundary layer and engine flows.

2. Standards for data, metadata, content. All data will be stored in standardized, open formats, in order to facilitate sharing and longevity. For example, numerical data will be stored in open formats such as HDF5, with compact metadata such as grid sizes and field types embedded as metadata within the datafile. Datasets will also be accompanied by README files in ASCII format describing discretization procedures, the physical meaning of the data, and other production details. Software will be written in contemporary languages (C, C++, Fortran90), will utilize standard libraries (MPI, LAPACK, BLAS, FFTW), and will be packaged following open-source software distribution standards, with automated configuration and installation scripts, automated test suites, documentation describing usage and design principles, and detailed comments with the source code.

3. Policies for access and sharing. All data and K-12 outreach materials will be made freely available to the public on the Internet. Project participants will collaborate internally by documenting their ongoing research on a specially created internal project Wiki. This will ensure that all team members have access to the most up-to-date information regarding the status of the project, and will help coordinate the writing of manuscripts for publication.

4. Policies for re-use, re-distribution, and production of derivatives. Re-use, redistribution, and production of derivatives will be encouraged by publishing data and K-12 outreach content under appropriate open licenses, such as the Open Data Commons Attribution License for data or through the Engine Combustion Network maintained by the Engine Combustion Department of Sandia National Laboratories.

5. Plans for distribution and archiving. Data and K-12 outreach content will be distributed via free access on the Internet, using standard, open-source distribution systems such as Wiki systems or the Engine Combustion Network for datafiles. Systems and the data they host will be maintained at least five years past the date of associated publications. The longevity of the data is assured by the choice of open standards for distribution systems and data formats.