Open Science Grid Data Management Plan

Version 2

Open Science Grid Project

February 21, 2011

Execution of the proposal "The Open Science Grid – The Next Five Years: Distributed High Throughput Computing for the Nation's Scientists, Researchers, Educators, and Students" will not result in collection or use of private data. The program of work will result in production of the following types of data:

1. Software
2. Documentation
3. Accounting Information
4. Papers

The data generated by the project is in the public domain.

We will develop disaster recovery plans as part of the program of work. As long as the project exists we will keep the data during its useful lifetime. If the project ends we will make copies of the data and distribute it to the stakeholders that need it.

All data is currently backed up and archived by the institution providing the service. Operational services hosted at the University of Indiana have failover systems at a second physical site.

We will identify the means and cost of retention of information beyond the length of time of the project as part of the program of work and address the need by the end of FY14.

For each of the above the management, storage, and access to this data is as follows:

1. Software – Most software resulting from the OSG Consortium that is accepted as part of the OSG software stack is owned and managed by the project providing that software. It is incorporated into the OSG software release process and provided to any OSG member who wants to use the software. Software utilities and tools developed by the OSG are stored in a source code repository owned by the University of Wisconsin-Madison,

2. Documentation – Documentation on the technology, installation, and use of the software and OSG services is official documentation of OSG. It will be maintained in a project collaborative work area managed by the OSG Operations under a Service Level Agreement.   Although most of the information in the collaborative work area is available to anyone, editing the information requires x509 certification. Documentation is maintained by Indiana University and Fermilab.

3. Accounting Information – Accounting information is accessible through an accounting portal for ongoing monitoring of the use of the distributed infrastructure. Accounting information is maintained by the institutions deploying OSG accounting collection services (currently Fermilab, Indiana University and University of Nebraska).

4. Papers – Papers describing projects, trials, prototypes, performance results, or other reports will be placed in a document database that is publicly accessible or requires presentation of a valid certificate for any restricted information. The document database is currently hosted at Fermilab, though we plan integrate this service with the documentation service at Indiana University in the future.