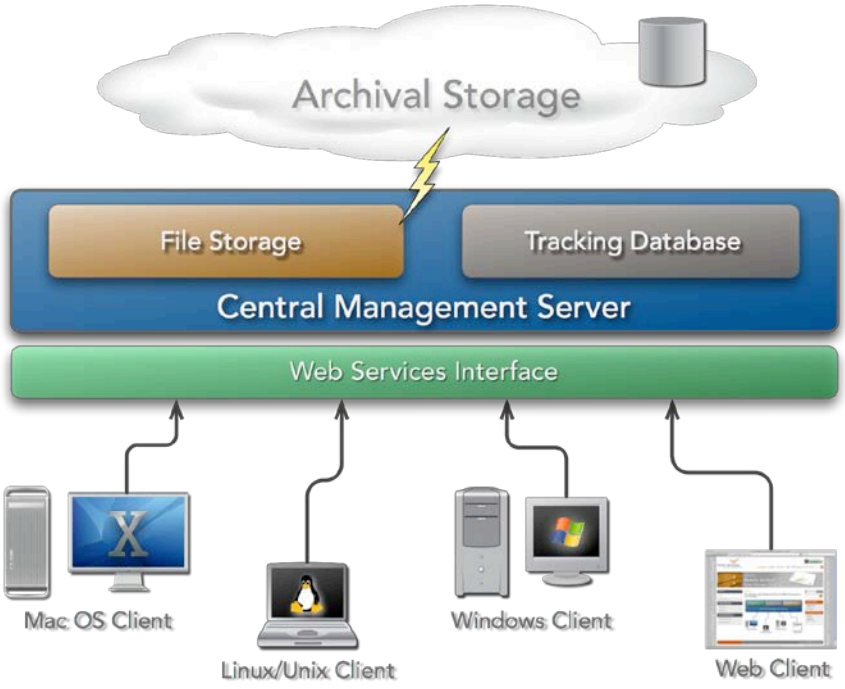
EMSL Research and Capability Development Proposals

Facility-Wide Management and Storage for Scientific Data

Project Start Date: Summer 2008

EMSL Lead Investigator
Ken Auberry
Instrumentation Development Laboratory, EMSL, PNNL

As greater numbers of collaborators, journals, and funding agencies require data retention associated with a given project, preservation of experimentally generated results has become an increasingly important challenge in science. In many small- to medium-scale laboratory environments, this task has traditionally been carried out using offline optical media (recordable CDs and DVDs) or externally connected commercial hard drive units. Along with the raw storage issues that must be addressed, additional challenges await in the correlation of stored data to contextual information about the experiments and methodologies employed to produce them. These approaches are tractable up to a certain limit—as long as the apparent convenience benefits seem to outweigh the more developmentally intensive alternatives. At some point, however, the process of manually tracking the location and pedigree of these data among all the various possible storage entities becomes more of a burden than the researcher is willing to bear.



Generally, this results in some commitment being made to develop a custom-built tracking and storage solution, with the requisite costs in time, money, and resources. In this case, the true issues at hand involve ease of use and user inertia (in the form of a reluctance to change) rather than the presence of locally available storage capacity or even the capture and cataloguing of metadata for curation purposes. A large part of bridging this gap requires constructing a straightforward and usable system where researchers can submit data for archiving and just as easily locate and retrieve it at some point in the future. Moreover, they must be able to do so in a manner that only minimally affects their current workflow.

This project is intended to provide a data storage and metadata tracking system coupled with curation and search and retrieve capabilities. In this way, individual research groups need not invest substantial amounts of funding in reengineering custom storage solutions. Instead, they can use a facility-provided service with some amount of customization available.

Products and Output

New Capability for EMSL Users

The desktop client architecture and much of the server-side architecture are currently being used as prototypes in the myEMSL lab-wide data management infrastructure. This system will not only provide EMSL users with a unified mechanism to preserve their data, but it eventually will provide a robust search capability intended to drive scientific progress through synergistic discovery approaches.