

# NCI HPC DME – High Performance Computing Data Management Environment

## ***Getting Started***

### **Brief Introduction:**

The HPC DME, High Performance Computing Data Management Environment, is an open ended data storage environment supporting storage and management of scientific data, produced from various HPC, genomic sequencing or lab information management systems. HPC DME provides flexible capabilities for storing(archiving), managing, transferring and sharing data/metadata across different systems securely and effectively.

HPC DME stores and associates user defined metadata to any registered data objects at different levels of data life cycle, enabling the environment not only to help identify the data but also to assist users in better organizing their research data/metadata and enhancing the search capabilities to empower user adoption, cross-group collaboration and custom control of data access to their own comfort zone.

### **Metadata - Organizing Unstructured Data:**

The HPC DME provides flexibility for each labs or groups to have their own way of organizing/structuring how their data/metadata is presented or preserved to enhance collaboration and connect unstructured data files with flexible metadata attributes attached. Standardization of such metadata is not enforced by the HPC DME but may be introduced through each group's own policy. Check out "Metadata Definition and Logical Entities of Hierarchies" at below Git training document [HPC DME General Training](#)

### **How to Get on Board:**

- For application Pre-Requisites, check out the [HPC DME User Guide](#).
- For Application Support or account creation with HPC DME, consult HPC DME administrator at [HPC\\_DME\\_Admin@nih.gov](mailto:HPC_DME_Admin@nih.gov) after obtaining your group or DOC approval of using the HPC DME.
- There will be an initial archive storage assessment and requirement from requested labs or groups to sustain application scalability though archive storage can currently be scaled up on as needed basis. Data storage need will be assessed, estimated and monitored as more groups start exploring the HPC DME for pilot use.
- Each group or DOC will define its own policy file or collection of validation rules and enforce/implement security on data/metadata. In addition, each group or DOC will have its own separate archive area in the storage vault – The group administrator has full authority on the access control and can implement authorizations of who will own, read, write or no

permission at all at different path or levels organized by their group. See [HPC DME General Training](#) for details.

- HPC DME Client Utility provides accessible command line interface to interact with [HPC DME APIs](#) to enable [batch file registration](#) and perform [integrated search](#).
- Access HPC DME with Soap UI or CURL command-line tool to [perform common use scenarios](#)

### **Core Features/Functionalities in Place:**

- Register a collection (PI Lab, Project, Run, Sample or dataset)
- Register a single data file/object into storage archive using Globus Connect
- Register a single data file/object into storage archive from a local share
- Perform update on a metadata attribute
- Subscribe to a known event
- Generate a report
- Update/assign permission to enable sharing
- Perform simple search functions
- Download a data file/object to a Globus end point
- Download a data file/object to a local directory

### **Considerations for Web (Coming soon):**

- Provides many GUIs to work with distributed IT resources, users with little knowledge of technology backend can quickly be productive with these interfaces.
- Nothing to deploy, install, or configure at the user desktop.
- Can be used with most any browser and any operating system.
- Custom enhanced search functions via web will maximize collaborators' use and promote sharing
- Group admin functions via web will alleviate common application administrator support and enhance client adoption
- Ideal for Internet connectivity as communication with the browser user is through the HTTP Server which requires only a single port to be open to the network.
- No middleware for running desktop applications (like ODBC, .NET, etc)