**High Performance Computing (HPC) Data Management Environment (DME) Getting Started: DOC Group Administrators**

**Purpose**

This document is intended to serve as a guide for getting started on using the High Performance Computing (HPC) Data Management Environment (DME) for the National Cancer Institute (NCI). Specifically, this guide is intended to help the reader understand the duties of a DOC Group Administrator, should the reader be designated such role in facilitating a DOC’s utilization of the HPC DME.

**Assumptions**

It is important that a DOC Group Administrator be familiar with how to use the HPC DME as an ordinary non-administrator end user.

**General Duties**

A DOC Group Administrator (Group Admin) is expected to be a knowledgeable user of the HPC DME, serving as a technical leader and guide in his/her respective DOC. Therefore, he/she plays a critical role in the adoption and ongoing utilization of the HPC DME by others who belong to the same DOC.

One of the responsibilities of a Group Admin is creation and/or maintenance of the metadata policy for their DOC. A metadata policy is a set of rules that must be followed for HPC DME users to register or modify/update metadata for Collections (logical containers for items) and Data Objects (items bearing potentially useful data). It is strongly recommended that every DOC utilizing the DME has a comprehensive metadata policy. With a well-devised metadata policy in effect, the DME users of a DOC shall find that the Data Objects they register in the HPC DME shall be easier to index, search, and organize, and the treatment of the Data Objects can be more consistent and uniform as they all adhere to specific rules about their metadata.

Currently, the mechanism for creating a metadata policy is part of a manual process that an HPC DME System Administrator (Sys Admin) executes to establish a DOC’s presence within the DME. It is envisioned that in the future, a Sys Admin or a Group Admin will have the capability to create or update a metadata policy using the HPC DME web GUI client. For production web client, go to <https://hpcdmeweb.nci.nih.gov>. For UAT web client, go to <https://fr-s-hpcdm-uat-p.ncifcrf.gov/>. Information on how to use HPC DME web client is available in [HPC DATA MANAGEMENT USER GUIDE](https://github.com/CBIIT/HPC_DME_APIs/raw/master/doc/guides/HPC_User_Guide.docx), Chapter 9.

Currently, a metadata policy is composed as a JavaScript Object Notation (JSON) file for the purposes of creation or update/modification. If a DOC Group Admin is comfortable working with JSON, he/she could, without the aid of a Sys Admin, write or edit the JSON that specifies the metadata rules that would be validated for any requests to register or update Data Objects in the DME. Any text editor application could be used for this purpose. However, it is outside the scope of this document to offer detailed instructions on how to write or edit JSON using any text editor application or other type of software product.

The following is a sample of a metadata policy expressed in the expected JSON form.

{"HpcMetadataValidationRules": {

"collectionMetadataValidationRules": [

{

"attribute": "Collection type",

"mandatory": true,

"validValues": [

"project",

"dataset",

"folder"

],

"ruleEnabled": true,

"DOC": "DOC-NAME"

},

{

"attribute": "Project name",

"mandatory": true,

"collectionType": "project",

"ruleEnabled": true,

"DOC": "DOC-NAME"

},

{

"attribute": "Project type",

"mandatory": true,

"defaultValue": "Unspecified",

"collectionType": "project",

"validValues": [

"Umbrella Project",

"Sequencing",

"Analysis"

],

"ruleEnabled": true,

"DOC": "DOC-NAME"

},

]

}

The following is a summary of the attributes per JSON object defining a metadata policy item.

1. **attribute**: Metadata attribute name
2. **mandatory**: Flag to indicate if it is required or not. Valid values are “true”, “false”
3. **defaultValue**: Default value of the attribute if no value is given
4. **collectionType**: Collection type name applicable for this attribute rule. Only one value is allowed.
5. **validValues**: List of valid values for this attribute
6. **ruleEnabled**: Flag to indicate if this rule is enabled or not. Valid values are “true” or “false”. If value is set to “false”, this rule will not be evaluated during validation process.
7. **DOC**: Division name applicable. Any registered user account is assigned one DOC in the system. Any specific DOC dependent rules may be attached and activated.

Since metadata of interest is widely variable across different organizations or work teams, an HPC DME Sys Admin would not be able to offer definitive guidance on what should be in any DOC’s metadata policy. The Sys Admin could assist with the mechanics of creating or updating a policy, but ultimately, a Group Admin is responsible for producing the content in his/her DOC’s metadata policy.

Before devoting much time to the mechanics of creating or updating a metadata policy, or even writing/editing JSON, a DOC Group Admin ought to spend some time thinking about what belongs in the policy for his/her DOC. Given the list of 7 attributes above used to define a metadata policy item, the Group Admin should decide what are the mandatory metadata attributes for their DOC’s Collections and Data Objects. For these mandatory attributes, the Group Admin can then determine the details of each such as default value and discrete set of valid values if applicable. After mandatory attributes have been determined and fleshed out, the Group Admin should then brainstorm additional optional metadata attributes that could be useful but may be omitted without any negative consequences on basic data quality. An HPC DME Group Administrator can devise a detailed metadata policy in plain English using the 7 listed attributes and then later concern himself/herself with converting the policy into a JSON form that the HPC DME system can consume. The ability to disable rules could be used to “draft” prospective rules without needing for them to be put into effect.

Another responsibility of an HPC DME Group Administrator is administering users and groups. Tasks related to this responsibility can be accomplished via the HPC DME web GUI client. For details about that, please refer to the [HPC DATA MANAGEMENT USER GUIDE](https://github.com/CBIIT/HPC_DME_APIs/raw/master/doc/guides/HPC_User_Guide.docx), and read the section about using the web GUI.