Milestone DigitalObjectRegistry: Plan to use Digital Object Architecture for experimenter tools and services.

1. Overview

Digital Object Architecture (DOA) defines a data model, the Digital Object (DO), which logically encapsulates related pieces of information, including metadata and unique resolvable identifiers, without requiring those pieces to be stored at the same physical location or using the same storage technologies. DOA provides three infrastructural components for managing DOs: 1) the Handle System, which provides a globally distributed persistent identifier system for widely distributed sets of digital objects, 2) the DO Repository, which provides a common interface for accessing and managing DOs, and 3) the DO Registry, which provides a registration and discovery service for finding and relating DOs. CNRI has designed several information management systems using the DOA to address problems and issues faced by various DOD agencies. For example, the ADL Registry (ADL-R) was designed for use by DOD agencies to register, discover, and reuse military training material[http://www.adlnet.gov/Technologies/adlr/ADLRDocuments/ADL%20Registry%20Documentation/adl-registry-and-cordra-volume-1.pdf]. The DOA is also the key technology for the DARPA Network Archive (DNA) effort, providing repository services for archiving and for secure sharing of information.

We believe the DOA is equally applicable to various GENI issues. This document discusses how the DOA may be used to support experimenter tools and services.

2. Experimeter Tools and Services

Over the past several months, we have studied the practices and approaches of the various clusters for performing experiments in GENI. We also studied the Experiment LifeCycle Document produced by the GPO, which highlighted some of the requirements for performing experiments in an interoperable way, including the need for an Experimental Specification Language, the requirements of a search engine to discover experiments and resources, the need for repository services for archiving experiment descriptions and results, and so on. While individual clusters have designed models for specifying experiments, we believe interoperability across those models is imperative in order to perform experiments that span multiple clusters. This does not preclude individual clusters from using custom models and languages, provided those custom specifications are interoperable with other specifications adopted in GENI. Figure 1 illustrates our understanding of the Experimental Specification Model, which would allow specifying experiments run in GENI, based on the Experiment LifeCycle Document, and the approaches taken by the cluster members. The illustrated model may be seen as a common mapping model between and among custom specifications, and may act as a starting point for gaining interoperability across the various clusters in GENI.

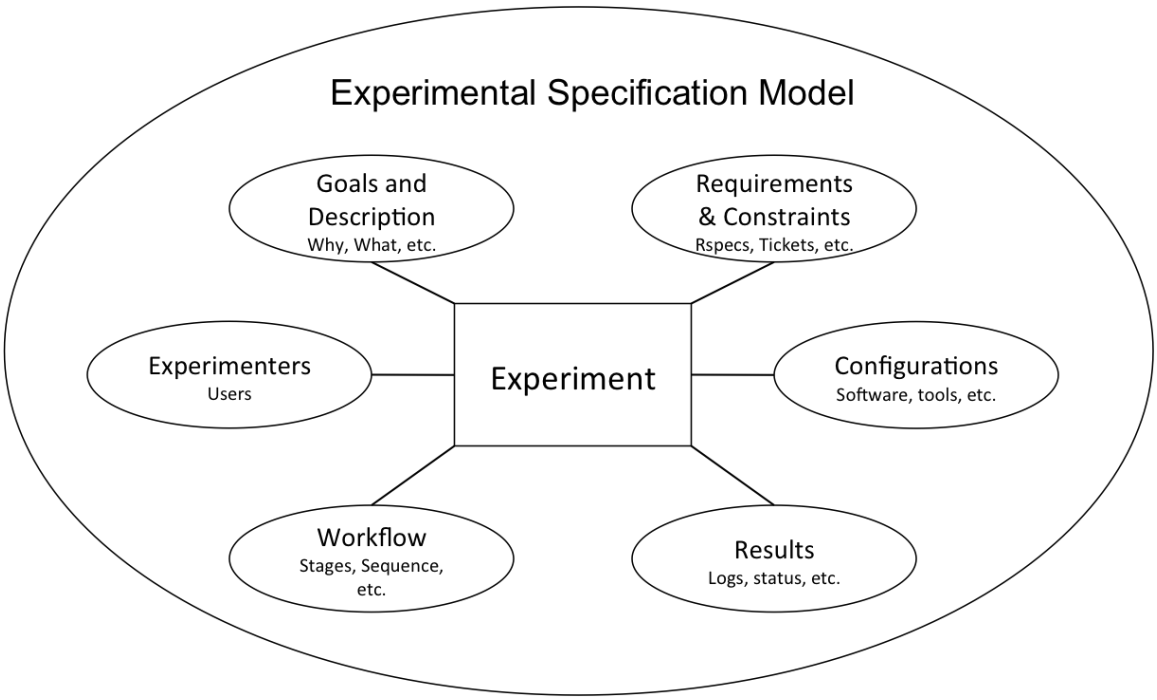

Figure 1: Proposed Experimental Specification Model

We also found that individual clusters and projects do not meet the search engine requirement at this point, at least not in an interoperable way. Ideally, if all the clusters conformed to the GPO requirements, then the DOA could be easily adapted, by implementing client tools, to perform each of those information management requirements. However, we believe, we can create an Experiment Specification Language based on the model illustrated in Figure 1. We can then integrate the DOA with the defined specification, the illustration of which is shown in Figure 2.

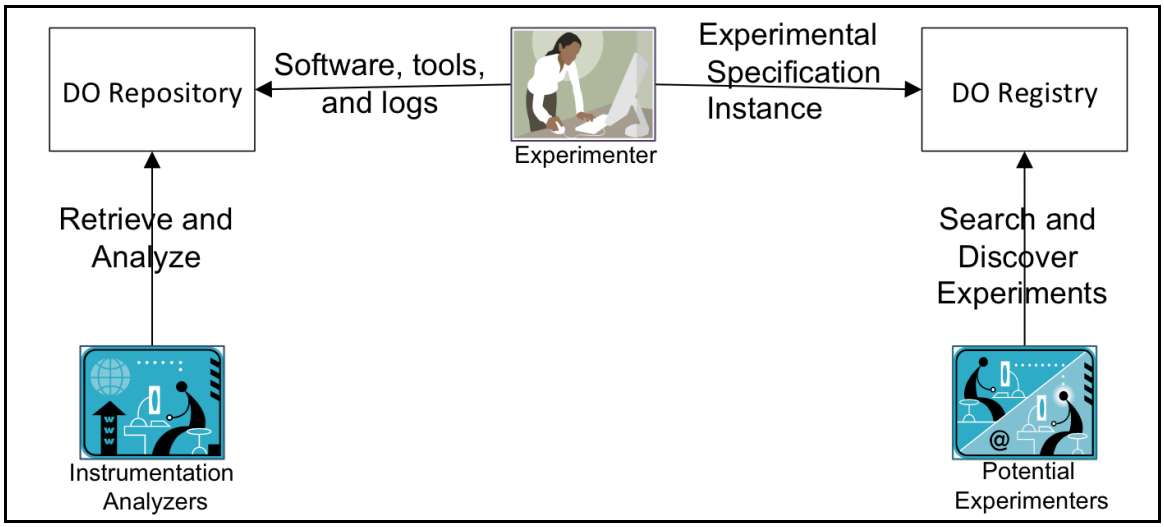

Figure 2: Proposed Integration of the DOA

The proposed customization of the DOA allows experimenters to register and store experiment-related information both during the period the experiments are performed and also at the end of those experiments. Experimenters may register, update, and delete information in a dynamic fashion at the time such experiments are performed, thereby keeping the state of the experiment up to date. We understand that it is not always possible to segregate information when experiments are underway for meaningfully registering the information in the DO Registry. While the DO Registry component requires associating metadata for registered objects, the DO Repository components allows storing snapshots of information, much like a database, but provides persistent reference to the information using the Handle System (that usually is missing in a database). The experiment controller, which manages experiments, may be integrated with the workflow of the proposed system to perform these steps. At the completion of the experiments, metadata may be associated with the information already stored in the DO Repository enabling the discovery of the archived information with the help of the search engine embedded in the DO Registry.

Another relevant aspect of the proposed architecture is the use of persistent identifiers to reference each of the pieces of information stored for experiments. That is, Handles are used to identify discrete pieces of information stored in the proposed system such as logs, software packages, results, descriptions, etc. The DO Registry and the DO Repository components logically group related pieces of information using those Handles, allowing various pieces of information related to a given experiment to be grouped together by reference. Aside from presenting a model for building associations easily, Handles also allow manipulations over existing associations. That is, the set of entities that are part of a group (experiment) may be easily changed. For cases such as combining two experiments together, or splitting an experiment into two, managing and changing related information would be a matter of managing the Handle references within the group.

In addition to providing the proposed system to GENI members, we would be pleased to assist various clusters and projects harness the DOA to meet the information management needs of those individual groups. That is, each of the groups would be able to use the deployed components of the DOA without changing their respective practices. Eventually, if and when interoperability among those groups starts to surface, the DOA could be adapted to the interoperable practices.

3. Spiral 2 Plan

We propose the following plan to allow GENI members to harness the DOA for meeting the experimentation requirements. Given time and funding, we would be pleased to implement the plan in Spiral 2.

We would design an Experimental Specification Model and define a language supporting that model based on current approaches and GPO recommendations, provided we get support from at least one cluster. We would then integrate the DOA to build experimental services as illustrated in Figure 2.

If getting collaboration from any of the clusters becomes challenging, we would still offer the following services for individual clusters to leverage.

- We would stand-up a Digital Object Registry (DOReg) on the public Internet, or on other networks as required, embedded with a Digital Object Repository (DORepo) and a Handle server. The DOReg could be used by GENI members to store experiment-related objects including experiment specifications, descriptions, results, logs, etc., by associating some form of metadata (presumably serialized in XML). That process, which we term "registration", would return a unique identifier, a Handle, that may then be used to retrieve the registered object. The DOReg would also allow discovering the registered objects through keyword searches that would be matched against the associated metadata.
- We would also stand-up a Digital Object Repository (DORepo), independent of the DOReg, that allows storing experiment-related objects that do not need any search-based discovery, but only require those objects to be returned in response to a direct request. The DORepo upon receiving an object for storage, would return a unique identifier, once again a Handle, which may then be used to retrieve the stored object. If required, the DORepo may be integrated with Amazon S3 or other cloud computing storage model, with the DORepo providing a unified interface to a diverse set of storage systems.
- We would also stand-up a Handle Server that would allow assigning unique identifiers, aka Handles, and associating resolvable information with those identifiers. Those identifiers may be used to uniquely identify any experiment related objects, or perhaps the experimenters themselves. Note that the Handles would allow authentication and trust evaluation as demonstrated during the GEC5. Currently, the Handle System supports PKI through SSL or through an open Handle protocol, and basic authentication using logins and passwords. If required, we can configure the deployed Handle server to provide such authentication services.

4. Conclusion

The Digital Object Architecture provides components that are required in any robust information management system. The flexibility of those components is manifested by their capability to store objects that conform to a specific standard as well as those that do not. While our plan at this point is to provide abstract storage, discovery, identity, and security services, it is our hope to continue our collaboration with the GPO and GENI members fine-tuning the level of specificity in the offered services as the requirements for those services firm up.