Hydra, DuraSpace and the Digital Public Library of America

The DPLA concept note

The DPLA concept note describes the DPLA as potentially beginning with existing digitized public domain works that already in public repositories, followed by adding additional materials as copyright issues can be resolved.

It also mentions the architectural likelihood of a distributed network of online resources.

The Hydra and DuraSpace Technology Family

Our initial submission suggested that we would explore a series of integrations of existing Fedora Repository, Hydra Project, DuraCloud, and Blacklight. These provide a good example of an approach that extends existing technologies and efforts in a fashion compatible with a DPLA. This combination can federate a number of sources, and is thus well placed to provide a desirable first step in the DPLA process.

While little progress has been made on a DPLA specific integration in the brief beta sprint period, there is substantial work proceeding in parallel that illustrates the applicability of the approaches in the Hydra Project community to the DPLA effort. Much of this work was already planned to occur during the beta sprint period, and so we are submitting it as an example of the potential.

We applaud the architectural concept of a distributed network of resources, and believe that the "stack" of technologies referenced here is one of the groups of building blocks that can form such a network. It is worth noting that a reasonably sized set of heterogenous approaches to a DPLA could well provide a more robust and vibrant solution than a single approach.

Relevant examples of Hydra and DuraSpace Technologies

The Hypatia project, an offshoot of AIMS using the above components with the exception of DuraCloud will complete its initial phase within a few weeks, and show a strong example of the technologies integrated together. A strong point of Hypatia is its use of Hypatia EAD conversion analysis which seems well suited to part of the DPLA problem space, and thus a natural candidate for further investigation as the project evolves.

DuraSpace has developed technology to move content between Fedora CloudSync, which opens the possibility of merging collections from multiple sources in a cloud environment with extensible resources for the purposes of curation activities prior to inclusion in a DPLA. In particular, utilizing cloud resources for format conversion activities is compelling, and has been demonstrated during the DuraCloud pilot program. This is compatible with and complementary to the standard Hydra / Fedora platform.

The Libra system at University of Virginia shows a fully operational Blacklight backed with Fedora, and gives a beginning vision of what part of the access to a DPLA might look like.

Areas of Interest

The Hydra framework tries to embrace and codify data models such that multiple families of both user facing and administrative applications can interoperate. Attempting this with DPLA's "enriched semantic metadata" as standards emerge should prove fertile ground. In general, Hydra and DuraSpace will continue to track the progress of content, interchange, and metadata standards for the DPLA as they evolve, as compatibility is clearly an advantage.