Dash: Exploiting Hydra/Blacklight for Repository-Agnostic Data Curation

Stephen Abrams, California Digital Library, Stephen.Abrams@ucop.edu
Shirin Faenza, California Digital Library, Shirin.Faenza@ucop.edu
Marisa Strong, California Digital Library, Marisa.Strong@ucop.edu
Bhavitavya Vedula, California Digital Library, Bhavi.Vedula@ucop.edu

Session Type (select one)
- ✓ Presentation

Abstract
University libraries and IT groups are increasingly being asked to support research data curation as a consequence of funder mandates, pre-publication requirements, institutional policies, and evolving norms of scholarly practice. Any repository-like service targeting academic researchers must provide both high service function and intuitive user experience in order to compete successfully against free commercial alternatives such as figshare or Dropbox. To meet these goals, the UC Curation Center (UC3) is developing a second generation of its Dash curation service. Dash is not a repository itself, but rather an overlay submission and discovery layer sitting on top of a repository and supporting refactored drag-n-drop upload, metadata entry, DOI assignment, faceted search/browse, and a plug-in mechanism for extension by reconfiguration rather than recoding. Dash is now based on a forked version of the Hydra/Blacklight platform that has been genericized to permit integration with any repository supporting SWORD deposit and OAI-PMH harvesting protocols. This presentation will introduce the new Dash architecture and describe UC3's process of enhancing Hydra for applicability beyond Fedora. UC3's deployment of enhanced service function through the composition of loosely-coupled, protocol-linked components exemplifies a useful approach for the streamlined creation of new innovative curation and repository services.

Conference Themes
Select the conference theme(s) your proposal best addresses:
- ✓ Managing Research (and Open) Data
- ✓ Integrating with External Systems
- ✓ Building the Perfect Repository

Keywords
Data curation, Hydra/Blacklight, loosely-coupled integration

Audience
This presentation is likely to appeal to repository managers, research datacenter managers, data science librarians, and data producers.

Background
The Dash project is outcome of UC3's desire to provide better services for managing research and open data. Dash is integrated with a number of external systems for authentication (Shibboleth and OAuth), and identifier management (EZID and ORCID). The loosely-coupled Dash architecture provides a useful example of extending repository function by composition with small, independent protocol-linked components rather than additions to large monolithic systems. UC3 believes that this approach, fully consistent with its long-standing promotion of micro-services architectures, is the most effective and efficient manner for constructing and enhancing sophisticated curation services.

Dash is a submission and discovery overlay layer that is applicable to any standards-compliant repository supporting the SWORD submission and OAI-PMH metadata harvesting protocols. While UC3's Merritt repository [2] provides native ingest and search interfaces, these were designed with trained institutional users – librarians, archivists, and other curators and content managers – in mind; while sufficient for that purpose, experience has shown that the existing interfaces present a technical barrier to use by individual researchers, who are generally unfamiliar with digital library and repository systems. To address these concerns, UC3 wanted to develop alternative interfaces that would provide simpler, more intuitive experience for the researcher community, akin to well-known services such as figshare and Dropbox, while still offering the same high level of underlying curation service.

Presentation content
This presentation will cover the following topics: research data curation and the need for effective yet intuitive services for researchers; an introduction to UC3's Dash service; the design and implementation of Dash, including the modifications to the Hydra/Blacklight platform necessary for applicability to repositories beyond Fedora; and the benefits of system construction by composition of loosely-coupled components, where new function can be introduced by reconfiguration rather than recoding. From a functional perspective, Dash has three main components: submission, harvesting, and discovery (see Figure 1). Protocol support is provided by pluggable modules conforming to the APIs of internal abstraction layers for authentication, metadata entry and serialization, persistent identifiers, repository packaging and submission, and metadata harvesting. Initial protocol support includes Shibboleth and OAuth authentication, DataCite metadata, EZID identifier management, SWORD submission, and OAI-PMH
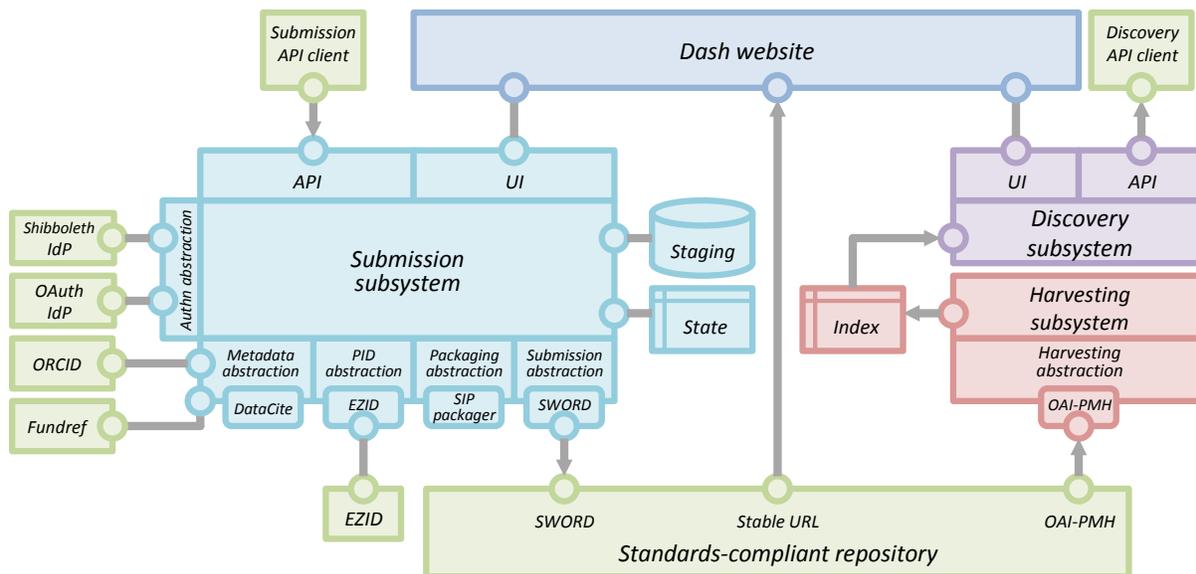
harvesting.

Figure 1 -- Dash functional architecture

UC3 considered a number of alternative approaches to Dash development, including incremental enhancement of the prior version's locally-developed codebase [1]. Strongly encouraged, however, by the successful integration of Hydra with a locally-developed DAMS by UCSD colleagues, UC3 decided to start fresh with a new Dash implementation also based on the community-supported Hydra/Blacklight platform. Hydra/Blacklight offers a stable and robust platform providing a number of important baseline functions out-of-the-box, permitting UC3's development resources to be refocused on added-value enhancements.

Dash's use of Hydra/Blacklight is shown in Figure 2. The standard ActiveFedora and Rubydora components are replaced by their Dash analogs, which rely upon the SWORD and OAI-PMH protocols rather than the native Fedora API, and which make Dash applicable to any standards-compliant repository. This activity was greatly facilitated by Hydra's highly modular architecture, which permitted UC3's efforts to be concentrated at a single point low in the technology stack.
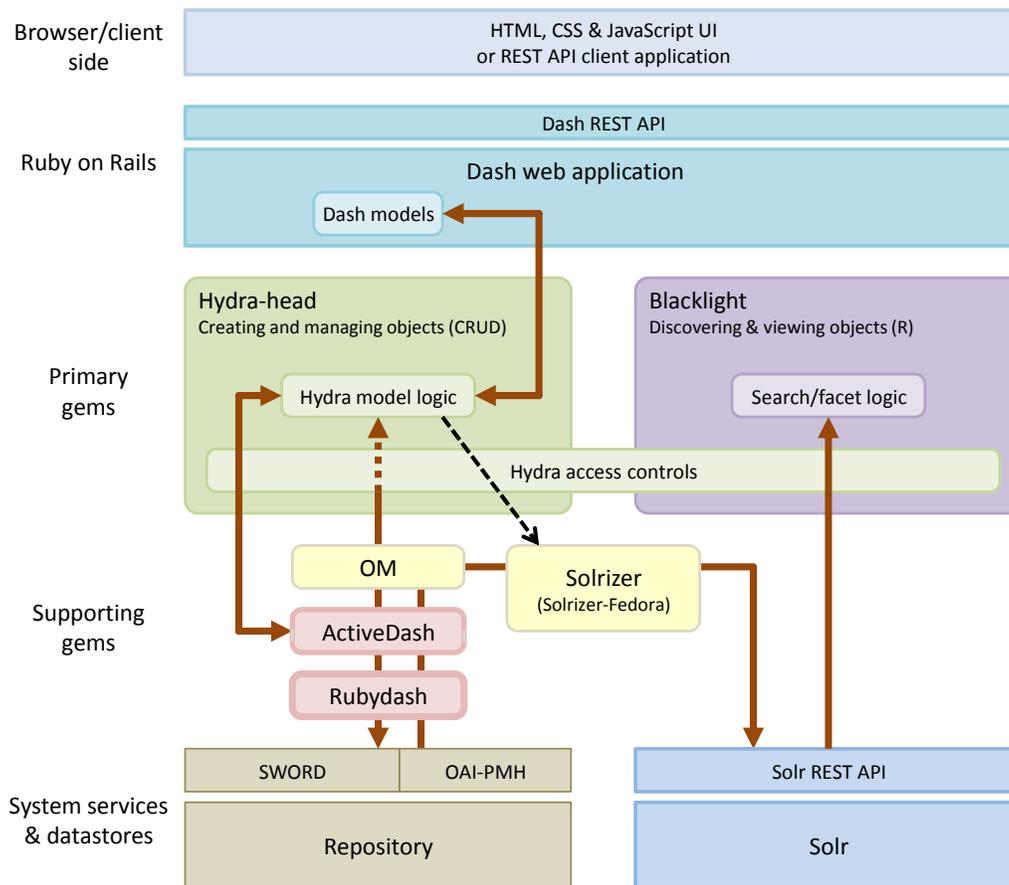
Figure 2 -- Dash Hydra/Blacklight architecture

Conclusion

The new version of Dash described in this presentation is the further evolution of an existing service in response to changing expectations by the research community. In order to streamline the development of simplified submission and discovery interfaces to its Merritt curation repository, UC3 used the Hydra/Blacklight platform as the basis for a loosely-coupled, protocol-linked overlay layer that, through localized modifications to Hydra, is now applicable for integration with any standards-compliant repository supporting the SWORD and OAI-PMH protocols. UC3's use of Hydra/ Blacklight provides a useful example of how the Hydra platform can be extended beyond its original design parameters to meet new functional goals of interest to the wider repository community

References

[1] UC Curation Center, *Dash: Data Sharing Made Easy* <dash.cdlib.org>.
[2] UC Curation Center, *Merritt* <merritt.cdlib.org>.