Dash Curation Service Infrastructure Enhancement: An Informed Extension & Redesign

Nancy J. Hoebelheinrich
California Digital Library
University of California Office of the President
415 20th Street, 4th Floor
Oakland, CA 94612-2901
+1-510-987-6482
nancy.hoebelheinrich@ucop.edu

Stephen Abrams
California Digital Library
University of California Office of the President
415 20th Street, 4th Floor
Oakland, CA 94612-2901
+1-510-987-6482
stephan.abrams@ucop.edu

ABSTRACT
University libraries and data repositories 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. While free commercial alternatives such as figshare and Dropbox provide high service functionality and intuitive user experience that serve research data creators well, they do not offer long term preservation reliability, nor do they necessarily share the increasingly important value of open data. From the perspective of the research data creator, however, all of these factors are important and desirable, so a preservation repository service targeting the needs of researchers should provide them. The UC Curation Center (UC3) at the California Digital Library created its *Dash* research data portal to address these needs. Following the initial deployment of the *Dash* service UC3 received feedback from users that additional functionality and a redesigned user interface would be desirable. With funding from the Alfred P. Sloan Foundation UC3 has re-factored the infrastructure behind *Dash*, and improved the front-end user experience of the existing deposit service. The *Dash* submission, harvesting, and discovery components are being extended to apply to any standards-compliant repository supporting the SWORD submission and OAI-PMH metadata harvesting protocols.

General Terms
Institutional opportunities and challenges; Infrastructure opportunities and challenges; Frameworks for digital preservation; Preservation strategies and workflows;

Keywords
Data curation; Data repository micro-services, SWORD submission protocol, OAI-PMH metadata harvesting protocol

1. INTRODUCTION
University libraries and data repositories are increasingly being asked to support research data curation in response to funder mandates, publication requirements, institutional policies, and evolving norms of scholarly practice. While free commercial alternatives such as figshare and Dropbox provide high service functionality and intuitive user experiences, they do not offer long-term preservation reliability, nor do they necessarily share the increasingly important value of open data. From the perspective of the research data creator, however, all of these factors are important and desirable, so a preservation repository-service targeting the needs of researchers should provide them. The UC Curation Center (UC3) at the California Digital Library created its *Dash* research data portal to address these concerns. *Dash* is not a repository itself, but rather a submission and discovery overlay layer sitting on top of CDL's Merritt curation repository that provides drag-n-drop upload, metadata entry, DOI assignment, and faceted search/browse.

2. ENHANCEMENT PLAN
After several years of providing the *Dash* service, UC3 received feedback from users that additional functionality and a redesigned user interface would be desirable, so a proposal was made to and funded by the Alfred P. Sloan Foundation to re-factor the infrastructure behind *Dash*, and to improve the front-end user experience of the existing self-service deposit workflow. The *Dash* service upgrade will continue to use the underlying micro-services architecture of extending repository functionality by developing small, independent, protocol-linked components rather than by adding to large, monolithic systems. Thus, the *Dash* submission, harvesting, and discovery overlay layer is being extended to apply to any standards-compliant repository supporting the SWORD submission and OAI-PMH metadata harvesting protocols. For a more complete picture of the components, and their interactions, see Figure 1, Dash Functional Architecture.

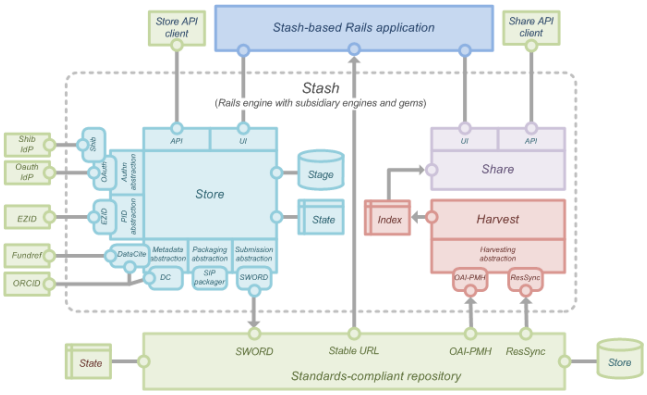

Figure 1. Dash Functional Architecture

3. PROTOCOL SUPPORT

Protocol support is provided by pluggable modules conforming to the APIs of internal abstraction layers for authentication, metadata entry and serialization, persistent identifiers including DOIs and ARKs, repository packaging and submission, and metadata harvesting. Besides supporting the SWORD and OAI-PMH protocols, the *Dash* service will support Shibboleth and OAuth authentication, DataCite and Dublin Core metadata schemes, and EZID metadata management. The front-end user experience is being informed by a more complete suite of user stories in order to provide a simpler, more intuitive interface designed with the individual researcher in mind. Researchers will be able to document, preserve, and publicly share their own data with minimal support required from repository staff, and be able to find, retrieve, and reuse data made available by others.

4. UI/UX-INFORMED DESIGN

One of the reasons often given to explain why researchers do not use repository tools for data submission is the poor design of their user interfaces. Often, the user interface does not take into account the user's experience (or inexperience) and expectations. Because so much of researchers' activities are conducted on the Internet, they are exposed to many high-quality, commercial-grade user interfaces in the course of a workday. Correspondingly, researchers have high expectations for clean, simple interfaces that can be learned quickly, with minimal need for contacting repository administrators. By means of extensive research into the user experience and usability testing, *Dash* is being designed with a simple, intuitive interface that will allow researchers to document, preserve, and publicly share their own data with minimal support required from repository staff, and also be able to find, retrieve, and reuse data made available by others.

5. SEEKING COLLABORATORS

By describing the *Dash* enhancement work in progress, CDL UC3's innovative and generalizable approach will show how the proven *Dash* research data portal, targeted to the needs of individual researchers, is being extended. Besides expanding awareness of the *Dash* service, CDL UC3 staff would like to identify potential collaborators from the digital preservation / open source communities who would be interested in participating and further developing the *Dash* software. More information about the current *Dash* service can be found at: http://dash.cdlib.org and about the *Dash* enhancement project at: https://confluence.ucop.edu/display/Stash/Stash+Home.