

DuraCloud DfR

Project Brief: *DuraCloud for Research*

CNI Fall Meeting

December 13, 2011

Jonathan Markow
Chief Strategy Officer

DuraCloud DfR

A DuraSpace Project Funded by
the Alfred P. Sloan Foundation

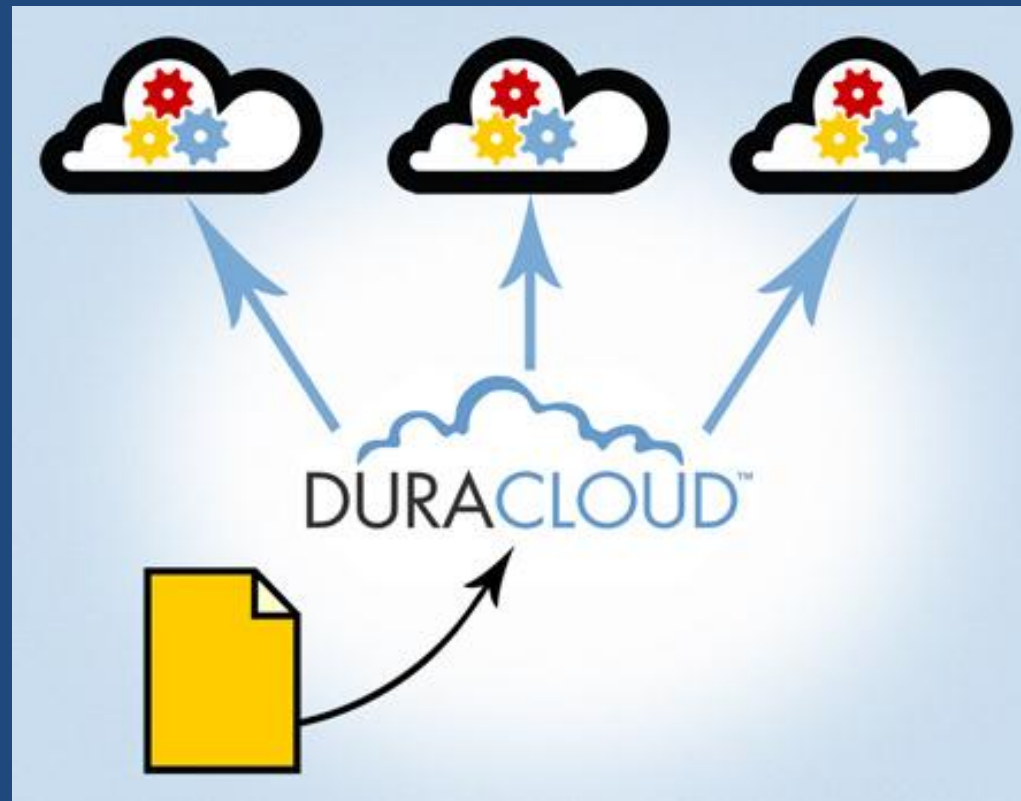
The DuraSpace Mission

We are committed to providing open source technologies and services that promote durable, persistent access to the scholarly record.



What is DuraCloud?

Digital archiving solution based on cloud infrastructure
Across multiple cloud providers



Why DuraCloud DfR?

- New use case for DuraCloud
- Protect vulnerable research data
- Facilitate archiving, access, and preservation
- Facilitate cooperation between researcher and institutional data managers
- Provide services to support the research process

Some DfR Principles

- Open source enterprise software solution
- Capture data close to the source
- Don't interfere with researchers' processes
- Provide incentives, added value for metadata creation
- Easy to use; workflows for collaboration, hand-off to institution

Project Name Change

DuraCloud DTR (“Direct to
Researcher”)



DuraCloud DfR (“DuraCloud for
Research”)

Project Structure

- Advisory Group
- Institutional participants
- Technology team
- Facilitated workshops
- Interviews/focus groups
- UX design and development
- Infrastructure development
- Iterative development
- Prototype testing

Advisory Group

Expertise:

- Researcher support
- Support for institutional data management planning
- Repository software and VRE design, development
- Legal, regulatory compliance

Workshop I Participants and Contributors

- Cornell
- George Washington
- Georgia Tech
- Harvard
- ICPSR
- Johns Hopkins
- MIT
- NCAR/UCAR
- Oregon State
- Rice
- Smithsonian
- U of Oregon
- U of Prince Edward Island
- U of Virginia
- Fluid Project
- DuraSpace

Workshop Outcomes

- Discuss key institutional successes, challenges and priorities in research data management
- Brainstorm requirements and set priorities for DfR
- Discuss project success factors and next steps

Top Five Priorities

#1: Connect the operational and archival phases of the data management lifecycle.

Top Five Priorities

#2: Create simple workflows across the data management lifecycle that automatically capture metadata and provenance.

(...and create incentives for additional metadata creation)

Top Five Priorities

#3: Ensure confidentiality, security, privacy, and predictability of data in the cloud. (Trust and Control)

Top Five Priorities

#4: Automate basic metadata creation and “catalogue” creation.

Top Five Priorities

#5: Create interoperability of operational systems, archiving solutions, and discovery systems used by specific research communities.

Some Other Priorities

- Management of data created beyond the institution
- Controlled sharing of files
- Manage and archive small data sets
- Full life cycle view of storage, backup, replication and archiving.

Next Steps

- Document inclusive list of “user stories” (high level functional spec)
- Define project scope (subset)
- Do architectural design
- Validate approach with advisors, institutional participants
- Work on UX design
- Forge partnerships with organizations doing related work

Architecture

Take advantage of open source reuse wherever possible, e.g.:

- Authentication (Enterprise SSO, Shibboleth)
- Cloud-based repository components
- Encryption, service bus, messaging
- Data management planning tools, persistent identifiers, citations, publishing references,
- etc.

User eXperience

- Reuse existing tools for visualization, manipulation of research data
- Create seamless experience for researcher
- Utilize best practices for interaction design

Partnerships

- Partner with organizations doing similar, complementary work...

Project Schedule

- Iterative development
- Evolving prototypes
- User Focus
- Expanding group of participants
- First release at end of 2012
(Interim releases throughout the year)

Questions?

DuraSpace: www.duraspace.org

DuraCloud: www.duracloud.org

DuraCloud DfR Project:

<https://wiki.duraspace.org/x/ZBfNAQ>

jjmarkow@duraspace.org