2012-03-20 DfR Participants Call

Attendees

- Jonathan Markow, DuraSpace
- Carissa Smith, DuraSpace
- Susan Parham, Georgia Tech
- Chris Helms, Georgia Tech
- Joan Parker, Moss Landing Marine Labratories
- Sid Byrd, Rice University
- Ying Jin, Rice University
- Gail Steinhart, Cornell University
- Deb Morley, MIT
- Brian Westra, University of Oregon
- Mike Wright, NCAR

Introductions

- Some folks attended requirements workshop in October, some are representatives from institutions that attended, and some are new today:
    - Gail Steinhart, Cornell University
    - Deb Morley, MIT Libraries (sitting in for Steve Gass), looking at various research data service opportunities
    - Susan Parham, Georgia Tech and Chris Helms network administrator
    - Sid Byrd and Ying Jin, Rice University (sitting in for Geneva Henry)
    - Mike Wright (NCAR)
    - Joan Parker, library/it/data management at Moss Landing Marine Labratories
    - Brian Westra, University of Oregon, science and data librarian

Discussion

- Summary of activity since last meeting
    - general feedback from participants was that it was very helpful
    - took discussions from meeting and developed use case stories for various roles
    - then created a list of what the system should do based on those roles and then started to prioritize (and also scope of first phase)
    - first phase of DfR ends at end of calendar year 2012
    - also created a narrative of the scope of items/activities the system will perform this year
    - then developed high level architecture to support the system
    - established partnership with Smithsonian and will be creating a data management and visualization system designed to look at research data that is stored in a Fedora repository and create incentive for researchers to add metadata to data for them to manage, organize, share data
    - DfR is partnering with Smithsonian to include their work for visualizing and managing data
    - have also identified four iterations of agile development for the year that also prioritizes tasks
- Summary of narrative report
    - basing DfR system on DuraCloud (which provides preservation services in the cloud that integrates with public cloud providers Amazon and Rackspace, as well as other providers in the near future)
    - DfR extensions will preserve researcher data which is very vulnerable from the beginning (not institutionally supported, ad hoc)
    - DfR will grab data from very beginning (regardless of where it resides), ask researcher what should get backed up, and give them options to add metadata about the data
    - question from Deb: how does this play out when there are multiple researchers, as well as researchers from multiple organizations
    - answer: DfR will support groups and have a management console where many people can have access to data
    - question from Deb: does local copy of data need to be maintained locally?
    - answer: we must be able to accommodate both scenarios, our initial expectation is that it would be a backup
    - a Fedora repository will be the backend of the Smithsonian application for managing and visualizing data
    - data in DuraCloud, metadata in Fedora (that is backed up to DuraCloud), web interface via Smithsonian application
    - Smithsonian application allows you to look at your data, describe the data, describe the relationships between the data, define your project concept, define associated entities and sub-projects, all of this will allow for the automatic creation of a project web site
    - question: transitory system not for longterm preservation?
    - answer: not sure what choices will be made from institutional level, but data is being stored for longterm, but migration options will also exist
    - question: are there any demos of the Smithsonian application available now or any upcoming?
    - answer: no public demos, just presentations; application is just getting to the stage where it can be demoed; they are trying to make that happen at Open Repositories conference and at some point we will install a version of the interface into the DfR system and we will be in the position to demonstrate it
    - question: future development on repository system and what tools would be enabled
    - answer: development team working on Smithsonian application has a lot of experience creating researcher web applications, so we hope that the application will evolve into a researcher environment in the cloud
    - Internet2/InCommon shibboleth integration that will enable authentication and authorization
    - feedback: intrigued by web site output service and assuming it will tied into datacite for DOIs
    - answer: we will integrate with a number of third party services that support researchers (datacite will be one); also identifiers, data management planning tool (from CDL), create a template plan
- Feedback from researchers is essential
    - researcher input in regards to adding metadata, interacting with the interface, the user interface to the system is greatly needed
    - email Jonathan if you would like to participate
    - feedback/recommendations for moving forward
    - feedback from Gail: take show on the road to scientific conferences to get feedback
    - recommendations welcome to those types of conferences, emails welcome
    - feedback from Joan: what would you do with live datastreams such as those coming from instrumentation directly

- not something we will be able to achieve in the first iteration, because of difference between devices and proprietary software associated with those devices
- question regarding preference of disciplines or metadata structures?
- answer: we would like a variety of disciplines, but will have to limit quantity of data for testing, we will set parameters to data sizes, number of files, etc.; first round of testing will focus on usability of user interface and application
- question: role of curator toward the end of the process, envisioning any other roles of library staff to engage with content from a curatorial perspective
- answer: will vary based on institution, there will be a continuum of partnerships around this
- suggestion: datacurationprofiles.org would be helpful as a resource, they provide a toolkit with questions about metadata and how to characterize data
- send email if you are interested in having researchers and/or other ways to volunteer