

DFC Milestones



Name
Affiliation



DFC October 2014 NSF Review Slide 1



Milestones: Hydrology		Quarter							
		Year 4				Year 5			
1	Metadata for workflows: Metadata for documenting the hydrology workflows to support sharing through systems like SEAD and HydroShare are required.	T	T	D	D	D	F	F	F
2	HydroShare interoperability: Create the ability to index automatically workflow resources through HydroShare, like what we have done for SEAD.	T	T	D	D	D	D	F	F
3	Federation with CUAHSI Water Data Center: Expose data available through the CUAHSI Water Data Center through the DFC.	H	H	H	H	F	F	F	F
4	Training for workflow development. Document the process of creating workflows as training material to support adoption of the approach.	T	T	D	D	D	F	F	
5	Workflow for the SWAT hydrologic model: Time permitting, expand our workflows for additional models with the tentative plan of the SWAT hydrologic model.					T	T	D	D
8	➤ Demonstrations				S				S


Legend: D: develop & test; F: feedback & improve; H: help users; T: Theory & prototype; S: demonstrate
 

Milestones: SciON; Marine & Seismic		Quarter							
		Year 4				Year 5			
1	Hardening of Sensor Stream Capture	D	D	D	D	I	I	I	I
2	Create micro-services for handling multiple types of waveforms		D	D	D				
3	Create micro-services for handling other non-waveform data		D	D	D	D	D	D	
4	Creation and deposition of Archival Sensor Packages into NetCDF resources (eg THREDDS Server)			D	D	D	D		
5	Promote applications using asynchronous federation subscription, notification, playback		D	D	H	D	D	D	H
6	Provide oceanographic data as available	I	I			I	I	I	I
7	Demonstrate				S				S

Legend: D: develop & test; F: feedback & improve; H: help users; T: Theory & prototype; S: demonstrate ; I: Ingest Data




DFC DataFed FEDERATION CONSENSUS NSF
 DFC October 2014 NSF Review Slide 3

Milestones: Engineering		Quarter							
		Year 4				Year 5			
1	Build interface to enable users to edit the format registry ontology			D	D	D	F	F	
2	Expand format registry to include more scientific domains and extend metadata ontology where appropriate	D	D	D	D	D	D	F	F
3	Leverage more file format metadata in the Format Registry identifier to improve accuracy	D	D	D	F	F			
4	Integrate identification indexer with chained indexing support within iRODS, and explore link with HIVE ontology server	T	T	D	D	D	F		
5	Provide continued production support for CIBER-U, RodsWiki, Format Registry, Format Converter	F	F	F	F	H	H	H	H
6	Modifications to RodsWiki as needed based on feedback and requests from user community	F	F	H	H	H	H	H	H
8	Demonstrations				S				S




Legend: D: develop & test; F: feedback & improve; H: help users; T: Theory & prototype; S: demonstrate

DFC DataFed FEDERATION CONSENSUS NSF
 DFC October 2014 NSF Review Slide 4

Milestones: Social Sci. - Odum/UNC-CH		Quarter							
		Year 4				Year 5			
1	Implement initial security policy for DFC Odum collection (work with DFCO-PS1, PS2)	D	D						
2	Develop automated DFC metadata creation workflows for social science data (work with DFCO-T13)	D	D	D	D	D			
3	Develop the ability to create new studies in Dataverse within automated DFC workflows.			D	D				
4	Allow Dataverse to populate iRODS AVUs	D	D						
5	Develop authentication connections for iRODS/DVN			F	D	D			
6	Develop domain specific micro-services & workflows (work with DFCO-T15)	D	D	F	D	D	D	F	D
7	Domain requirements/refinements for policies & procedures (work with DFCO-PS1, PS2)	D	D	F	D	D	F	D	D
8	Demonstrations				S				S





 Legend: D: develop & test; F: feedback & improve; H: help users; T: Theory & prototype; S: demonstrate
 

Milestones: TDLC; Science of Learning		Quarter							
		Year 4				Year 5			
1	Establish standards for different collections from different scientific domains.	D	D	D	F	F	S	S	S
2	Graphical User Interface Development: Create easy Interface for the Rules Engine (for control of each dataset for IRB and IACUC restrictions).	D	D	F	D	D	F	D	F
3	Enable data analysis on the grid, using each scientific domain's tools as middleware.	D	D	D	D	D	D	D	S
		F	F	F	F	F	F	F	
4	Facilitate Cloud or Cluster Computing and workflows for TDLC and SLC partners.			D	F	F	D	S	S
5	Facilitate support for database and metadata access, and other infrastructure for TDLC partners.	F	D	F	F	D	F	F	F
		D			D				
6	Publish data collections for use in the community.	D	S	D	S	D	D	S	S
7	Develop portals for sharing data training sets for student education (Interface with MOOCS).			D	F	S	S	F	S



 Legend: D: develop & test; F: feedback & improve; H: help users; T: Theory & prototype; S: demonstrate
 

Milestones: iPlant		Quarter							
		Year 4				Year 5			
1	Federation with DFC partners and other institutes (US and UK iPlant Data Store instances)	D	D	F	F	D	D	F	F
2	HIVE metadata mgmt for iPlant Data commons	D	D	F	D	D	F	D	F
3	Metric and understanding of how our data store is being utilized	D	D	D	D	D	D	D	D
		F	F	F	F	F	F	F	F
4	Scalability for iRODS fuse, data distribution via content delivery systems	D	D	F	F			D	D
5	Enhance support for spatial data, LIDAR	F	D	F	F	D			
		D			D				
6	Use of container technology (Docker) for processing data directly on resource servers	T	T	T	D	D	D	F	F
7	Data Carpentry workshops for managing NGS data	H	H	H	H	H	H	H	H
8	Adoption of iPlant web work bench (DE) by DFC	F	D	F	H	H	H	H	H



Legend: D: develop & test; F: feedback & improve; H: help users; T: Theory & prototype; S: demonstrate

DFC October 2014 NSF Review Slide 7

Milestones: Technology Development		Quarter							
		Year 4				Year 5			
1	Hardening of the Asynchronous Federation Model to improve interoperability mechanisms in DFC	D	D	D	F	D	D	F	F
2	Graphical User Interface Development	D	D	F	D	D	F	D	F
3	(TL1) Apply the weak and asynchronous federation models to Integrate new formats, tools, services and resources for S&E partners	D	D	D	D	D	D	D	D
		F	F	F	F	F	F	F	F
4	(TL2) Facilitate Cloud Computing and workflows for S&E partners	D	D	F	F			D	D
5	(TL3) Facilitate support for database and metadata access, and other infrastructure for S&E partners	F	D	F	F	D	F	F	F
		D			D				
6	(TL4) Theory and implementation of Policy Translation	T	T	T		D	D	F	F
7	Outreach and support S & E partners and outside groups in using DFC tools, resources and services.	H	H	H	H	H	H	H	H
8	Demonstrations				S				S

Legend: D: develop & test; F: feedback & improve; H: help users; T: Theory & prototype; S: demonstrate



www.datafed.org

www.irods.org



National Science Foundation Cooperative Agreement: OCI-0940841



Education & Outreach: Now and Next



National Science Foundation Cooperative Agreement OCI-0940841

The Buzz is HIVE!

- **Student Engagement in R&D:** 6 SILS students participated in HIVE projects via the HIVE/DFC and HIVE/DataONE supplements.
- **Cross-Training Students:** RDA summer intern linked the RDA (Research Data Alliance) Metadata Standard Director WG (MSDWG) and the Policy workgroup this summer.
- **HIVE** technology and testing has been integrated into a number of ILS courses at UNC and Drexel covering metadata and the semantic web.
- **Metadata:** SILS conducted original research on DFC metadata workflows to be presented at Dublin Core international metadata conference this week!



DFC October 2014 NSF Review Slide 2



DFC Meets TDLC Summer Fellows Institute 2015

- Two week intensive introduction to TDLC Center research and facilities will include “Big Data” in the iRODS/DFC environment.
- **Goal:** “Basic Training” -- students should have basic set of skills from the four research networks, plus computational modeling, use of the Motion Capture/Brain Dynamics facility and the Data Grid.
- **Format:** morning lectures, afternoon labs, evening tutorials
- Mon-Sat all day and evening
- Week-long project in week 2
 - Students and postdocs cross-train, leave their “comfort zone”
 - Presentations on final day



DFC October 2014 NSF Review Slide 3



Curation is the new Tech!

- New DFC Information Technologists will be educated in curating and digitally preserving data such that it can have a lifespan that outlasts the scientists.
- Adopting microservice bundles from standards such as “Trustworthy Digital Repositories” and adding basic data management plan to future iRODS installations can facilitate this education.
- IT managers will undergo “basic training boot camps” in preservation and microservice bundles, thereby creating a hybrid community of technology/curators.



DFC October 2014 NSF Review Slide 4



Here come the MOOCs!!

- TDLC Scientists are partnering on 2nd round of Coursera MOOC on learning how to learn.
- First offering of course yielded **197,000** students.
 - Coursera predicts 2nd offering could beat its all-time record of 250,000 students.
- Presents enormous audience for DFC to consider providing supplemental educational materials and releasing student data sets.
- DFC might consider partnering with TDLC on a unique Data Science Course.



Learning
How to Learn



DFC October 2014 NSF Review Slide 5



Data Carpentry

1. Based on successful model of Software Carpentry
<http://software-carpentry.org/>
2. DFC partners will focus on creating new modules focused on:
 - a) Large scale data for next gen sequencing (NGS)
 - b) Managing imaging data
 - c) Spatial data management
 - d) Data hygiene for all scientists
 - e) Pragmatic data management for life sciences
3. Initial work is led by volunteers (Tracy Teal and Karen Kranston)
4. <http://datacarpentry.org/>



DFC April 2013 NSF Review Slide 6



Student Support in DFC

- Goals Accomplished:
 - Integration of RODSWiki/MediaWiki ((REU) Jacob Olitsky, Drexel)
 - Developing engineering digital library services ((REU) Alexandru Nedelcu, Daniel Speichert, Drexel)
 - English2Rules project (Sajid Anwar, Vo Nguyen, Arizona State)
 - Micro-service Analysis (REU) (Shane Pusz, UNC)
 - Integration of the HIVE linked-data system with DFC (Le Zhang, Chelcie Rowell, Rajavi Koushyar, UNC)
 - Policy Surveys and Analysis (Jewel Ward, Deborah Maron, UNC)
 - RHESS Workflow (Brian Miles, UNC)
 - VIC Workflow (Bakinam Essawy, Benjamin Felton, Uva)
 - VIC Workflow (Mirza Billah, USC)
 - DFC PhoneGap ((REU) Jack Harrison, UNC)
 - GUI Design ((REU) Marcos Prieto, UNC)



DFC October 2014 NSF Review Slide 7



Present Outreach

DataNet Partnerships – technology federation

- DataONE
- SEAD
- TerraPopulus

NSF funded Projects

- CUAHSI
- HydroShare
- WSSI
- XSEDE
- Data Bridge
- EarthCube

Federal Agencies

- NASA
- NOAA/NCDC
- EPA

Standards & Consortia

- Research Data Alliance
- National Data Science Consortium
- iRODS Consortium

University Outreach

- SILS Lifetime Library
- Carolina Digital Library
- ADCIRC Modeling and Forecasting
- Genomics Data Grid

Commercial & Industrial Outreach

- Direct Data Network
- EMC



DFC October 2014 NSF Review Slide 8



Future Outreach

- Continue with Present Outreach efforts
- New Outreach Activities
 - Extend Commercial Outreach
 - Extend Federal Outreach
 - Start State-level Outreach
 - Educational Outreach
 - MOOC
 - MOOD
 - University curriculum activities
 - Reach out to other federally funded projects for data and service federation opportunities



DFC October 2014 NSF Review Slide 9



DFC **DataNet**
FEDERATION
CONSORTIUM

www.datafed.org

www.irods.org



National Science Foundation Cooperative Agreement: OCI-0940841