

Organizing for the 21st Century of Research and Discovery: Data Management and the RMACC

Dawn Paschal – Colorado State University

Andrew Johnson - University of Colorado

Chad Hutchens – University of Wyoming

Matthew Ketterling – Colorado School of Mines



RMACC
August 10, 2016

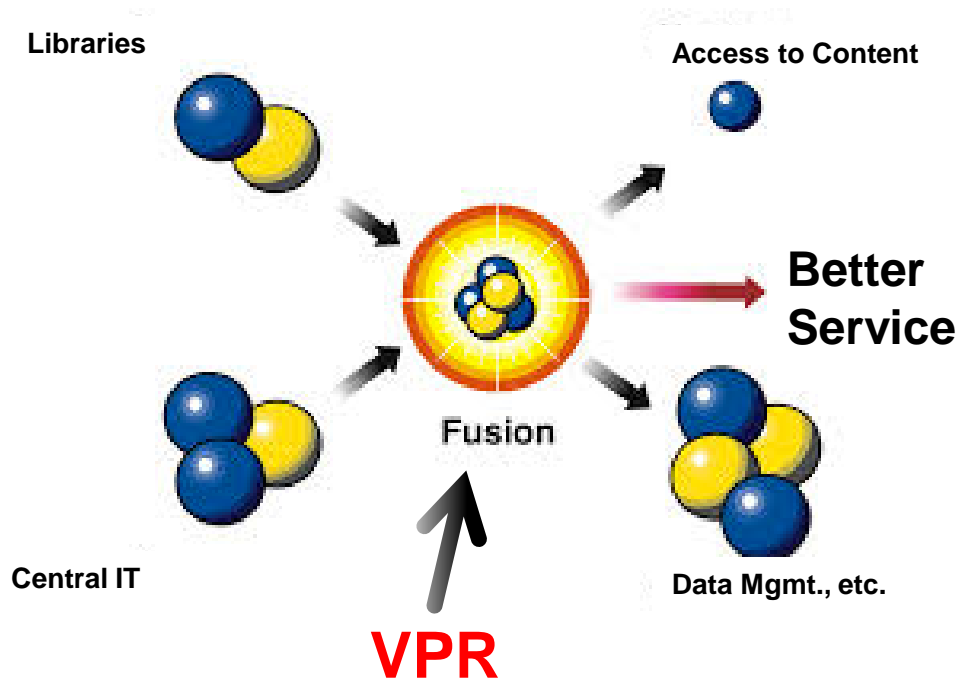
Research Data Management Services at CSU

Dawn Paschal

CSU Libraries

dawn.paschal@colostate.edu

Essential Partnership: CSU IT, Libraries, VPR



CSU ISTeC Research Data Management Committee

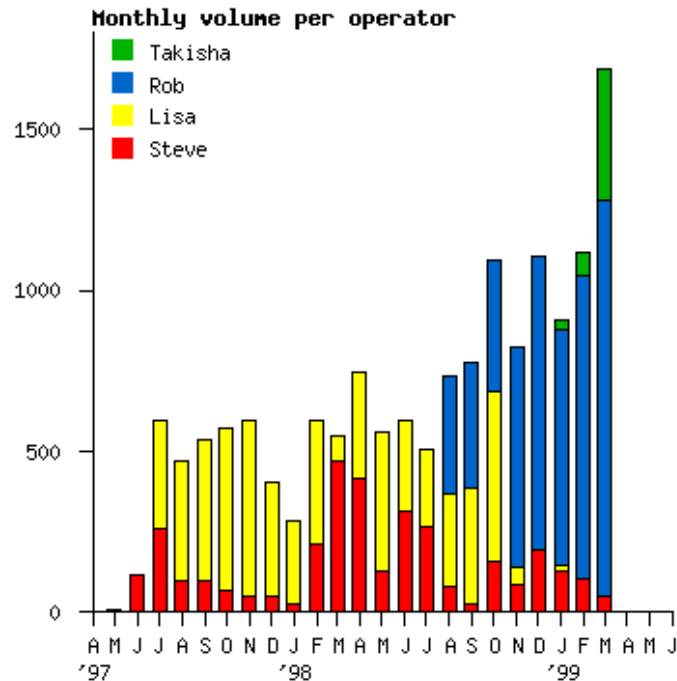
Membership from each College, OVPR, Central IT
Organization/Administration

Key Areas of Development:

- Infrastructure
 - Services
 - Policies
 - Education

Research Data Services

- ▶ Data Management Specialist in the Libraries
- ▶ Consultations (general questions, data management plans, best practices)
- ▶ Customized workshops upon request
- ▶ Support for Globus



Digital Collections of Colorado

- ▶ Shared institutional repository service between CSU, Mines, Mesa, CU-Health Sciences, UC-Colorado Springs (<https://dspace.library.colostate.edu/>)
- ▶ Publication for small (up to 200 GB) and medium (over 200 GB - 10 TB) datasets
 - ▶ Dublin Core metadata, managed access
 - ▶ Persistent URLs
 - ▶ ORCID identifiers
 - ▶ Secure storage
 - ▶ Preservation
 - ▶ No cost to researchers

Hydro-Biogeochemistry Transport of Black Carbon Project - data



This data package was produced by researchers working on the Hydro-Biogeochemistry Transport of Black Carbon Project. This project was supported by National Science Foundation and the Water Institute at Colorado State University. This data package includes a tab-delimited data table and an accompanying tab-delimited file (file name ending in '_var') that denotes header definitions and data types for each column in the data table. Detailed metadata also are contained within an Ecological Metadata Language document (i.e. XML). Black carbon (BC) is a ubiquitous component of the carbon cycle, ...

Author

Cotrufo, Francesca

Date

2015


Collections

- [Data - Colorado State University](#)
- [Data - Natural Resource Ecology Laboratory \(NREL\)](#)
- [Data - Soil and Crop Sciences](#)

URI

<http://hdl.handle.net/10217/170584>

View/Open

 [Data \(27.08Kb\)](#)

Metadata

[Show full item record](#)



- Search DSpace
- This Collection

BROWSE

All of DSpace

Communities & Collections

By Issue Date

Authors

Titles

Subjects

This Collection

By Issue Date

Authors

Titles

Subjects

MY ACCOUNT

[Login](#)

New Campus Research Data Policy

- ▶ Approved by Cabinet April 2016
 - ▶ Reviewed by various campus stakeholders
- ▶ Focuses on stewardship, responsibility, retention of data
- ▶ Requires students to submit data accompanying electronic theses and dissertations (ETDs)



Coming Soon

- ▶ Condo HPC model at CU - Summit
- ▶ Self deposit module for DSpace
- ▶ Support for associating data sets with grant information
- ▶ A 5th Libraries/ISTeC-sponsored Research Data Management Forum

Issues and Challenges

- ▶ Generally, busy faculty, researchers, students lack knowledge of best practices in data management
- ▶ Storage for “working” data sets
- ▶ Populating local repositories with “published” data sets
- ▶ Metadata
- ▶ Preservation

Issues and Challenges

- ▶ Security and privacy
 - ▶ Compliance with federal contracts
- ▶ Sustainability
 - ▶ Difficult given all of these issues
 - ▶ Need more investment in research computing vs. administrative computing



Research Data Services at CU-Boulder

Andrew Johnson
Research Data Librarian
CU-Boulder Libraries

andrew.m.johnson@colorado.edu

Research Data Services (<https://data.colorado.edu>)

- One-on-one data management consulting (DMPs, journal policies, general questions)
- Custom workshops on data management topics for classes, seminars, labs, etc.
- Support for the DMPTool: <https://dmptool.org/>
- Secure data storage via the PetaLibrary:
<https://www.rc.colorado.edu/resources/storage/petalibrary>
- Publishing of small (under 2GB) data sets via CU Scholar:
<http://scholar.colorado.edu/>

Usage of RDS services

- 240 DMPs created with DMPTool (6th most out of 189 institutions using DMPTool)
- Approximately 20 workshops per year (attendance varies widely)
- Approximately 40 consultations per year
- No deposits of data sets in CU Scholar yet (written into dozens of DMPs)
- Over 400 TB PetaLibrary storage used (not all research data)

RDS organizational structure

- Service providers: Personnel in Libraries and Research Computing (1 FTE each)
- Oversight: Senior leadership from Office of the Vice Chancellor for Research, Research Computing, and Libraries
- Research Data Advisory Committee:
 - Feedback on services and help with outreach
 - DMPs for campus grant competitions:
<http://escholarship.umassmed.edu/jeslib/vol5/iss1/1/>
- But, many changes underway...



Intro to the PetaLibrary

University of Wyoming

Advanced Research Computing Center (ARCC)

arcc-info@uwyo.edu | (307) 766-7748

UW Libraries

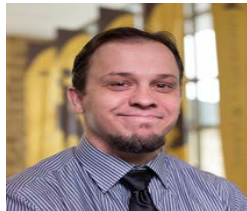
wyodata@uwyo.edu | (307) 766-5560



the ARCC Team



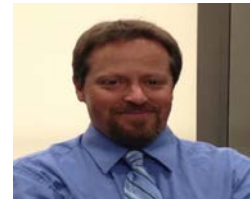
Jared Baker
Application Optimization
and Development



Mike Killean
Storage Architect



Jeff Lang
System Architect
and Administrator



Tim Brewer
End User
Support Manager



Dylan Perkins
Database Analyst



Karen Milberger
Data Ombudsman



Nikhil Shetty
3D Visualization
Expert



Libraries Data Management Group



Larry Schmidt
Head Brinkerhoff
Geology Library



Chad Hutchens
Head of Digital
Collections



Amanda Lehman
Digital Collections
Librarian



Steve Boss
Libraries Information
Technology Department



Ed Fugikawa Digital
Collections Systems
Programmer



Data Mgmt Services provided at Wyoming

- ARCC providing training on HPC applications (workshops, one-on-one)
- DMPTool is available for researchers
- ARCC & UW Libraries partnering on marketing/outreach this Fall...
- Libraries developing Data Repository (public interface) scheduled for release Sept. 15, 2016

UW's petaLibrary - What is it?

- Working data (Data Commons) - ARCC
 - Homes (500GB free storage)
 - Commons (collab working data storage \$100/tb/yr or \$1500/tb/10 yrs)
- Public/shared data (Data Curation) - Libs
 - Publications (finalized, publicly shared data)
 - Archive (under development)
 - Offsite data preservation (also under development)





Public Data Repository

Islandora Data Repository

- Public interface to PetaLibrary
 - Focus on research data only
 - DataCite, Dublin Core metadata
 - ORCID researcher identifiers
 - Using EZID & Fedora API's to create objects & mint DOI's
 - No file size limitations, no cost to researchers
- Target Release Date: Sept. 15, 2016

<https://data.uwyo.edu>



Long-term identifiers made easy



Member
Organization



UW Collaborations

- Research Office, ARCC, UW-IT, Libraries all collaborating
- Began slowly and very informally, has grown over time
- Challenges in communication and knowing who does what
- ARCC is responsible for Working Data and HPC applications
- Libraries responsible for Data Curation, publication, sharing
- Marketing, outreach, and workshops will likely be shared responsibilities



What's left & what's missing?

ARCC & UW Libraries still need to:

- Finalize interface, infrastructure, policies, addt'l DMP templates, etc
- Off-site storage
- Marketing, outreach, workshop development to address lack of knowledge

At UW in General (i.e. amongst students, researchers, and faculty)

- Lack of knowledge on campus about data mgmt practices and tools