CC* Cyber Team: Creating a Community of Regional Data and Workflow Cyberinfrastructure Facilitators

Thomas Hauser
Director of Research Computing
University of Colorado Boulder
thomas.hauser@colorado.edu

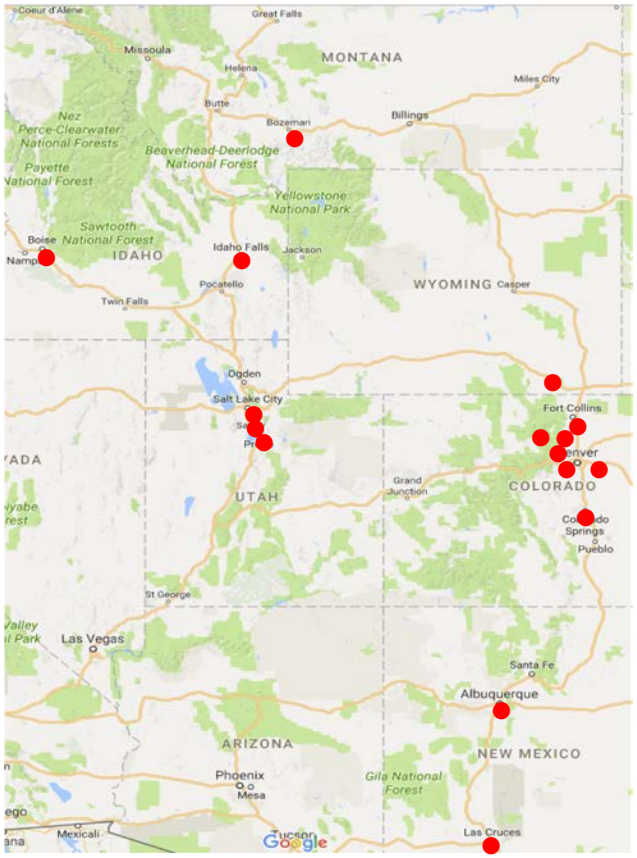
Patrick J. Burns
VP for Information Technology, Dean of Libraries
Colorado State University
patrick.burns@colostate.edu

H.J. Siegel
Professor of Computer Science
Colorado State University
HJ@colostate.edu

Thomas E. Cheatham, III
Director, Research Computing and CHPC. Professor of Medicinal Chemistry
University of Utah
tec3@utah.edu

Project approach

- Engagement with science teams that work with experimental data
- Create shared CI by integrating existing resources:
  - Compute – RMACC Summit
  - Storage – Archival, Repository and Object storage
  - **People - facilitators**
- Community building
- Training and professional development
  - Technical – OpenScience grid
  - Soft skills



- Rocky Mountain Advanced Computing Consortium (RMACC)

Challenges

- Software
  - Containers help us to address a variety of software problems: Biocontainers, Singularity
- Working with protected data
  - Utah contributes experience
- IRB approval of the surveys and focus group interviews
  - Delayed some of the activities
  - Process can be very different at institutions
  - Necessary if you want to publish your work
- Checkpointing
  - Solved for single node computations
- Regional perfSONAR dashboard
- Data mining to improve services
  - Integrate log, user, networking data

Sustainability

- Engagement with science teams
- Compute
  - RMACC Summit limited to 5 years
  - OpenScience grid
  - XSEDE resources
  - Cloud
- Storage
  - PetaLibrary @ CU Boulder
    - Institutionalized funding
    - Media cost for researchers
    - Developing external rates
    - Long term rates – 5-20 years
    - Benchmarking with cloud services
  - Object store @ Utah
- Networking
  - Institutionalized ScienceDMZ and DTNs at some institutions
- People
  - Commitments during submission
- Community
  - RMACC symposium
  - XSEDE campus champions
  - Engagement with under resourced schools
    - Working on pilots in several states
    - Funding for travel through RMACC conference
    - Biggest challenge: staff time

Sustainability @ CU Boulder

- Center for Research Data and Digital Scholarship
  - Libraries are essential partner on researcher engagement
  - Informing and training subject librarians
- Interdisciplinary research data consult hours
  - Data Management, Computing, Statistical Consulting
  - Spring 2017 - once a week
  - Fall 2017 – twice a week
- Re-envision research support @ CU Boulder
- Exploring partnership with graduate school around data

Regional Aspect of the Project

- RMACC symposium
  - New track around data
  - 8th symposium
  - Improve targeting graduate students
- Supporting smaller schools
  - Building partnerships
  - Travel cost funded through RMACC
    - Reinvest SC booth budget to travel to smaller schools
    - 3 pilots this year
      - Collect experiences
      - Evaluate who we do things
  - Challenge: allocation of staff time