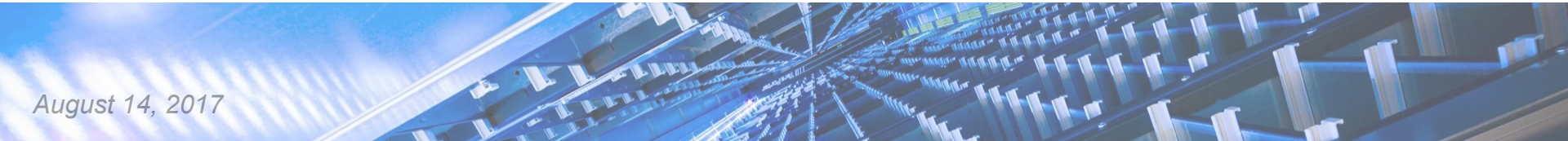
NATIONAL DATA SERVICE

THE DATA EXPLORATION LAB

Container Analysis Environments Workshop

August 14-16th, 2017
National Center for Supercomputing Applications
Champaign, Illinois

Craig Willis
willis8@illinois.edu
NCSA/NDS



How did we get here?

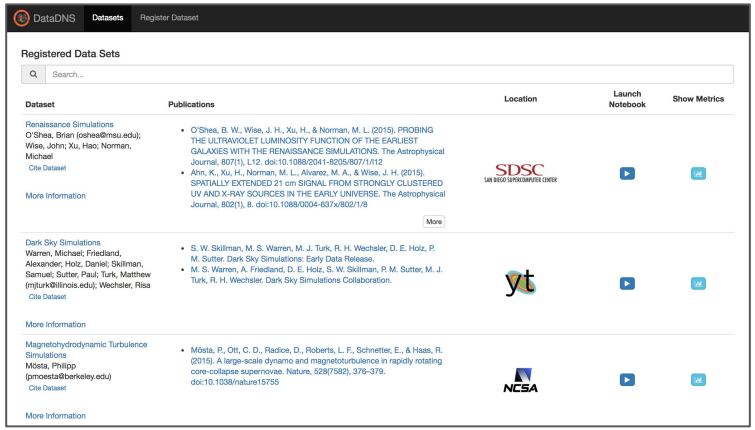
- 7th NDS Consortium Workshop (April 2017)
  - Container Analysis Environments interest group
  - BD2K KnowEng, Labs Workbench, SciServer, Whole Tale...
- Practice & Experience in Advanced Research Computing (PEARC'17)
  - CyberGIS, NDS, BIDS, Minnesota Supercomputing Institute
- Different groups solving similar problems for research communities
- DXL and NDS offered to support a small workshop

The National Data Service and NDS Consortium

Vision: *National initiative toward enabling scientists to broadly publish, find, and reuse digital data collections.*

- NDS core organization:
  - Director and dedicated staff at SDSC/NCSA
  - Executive, steering and technical advisory committees
  - Provides support for NDS pilots (storage, compute, human resources)
- NDS consortium:
  - Consortium of service providers, publishers, libraries and tool-builders
  - Addressing interoperability of data management tools and services
- Intended to complement Research Data Alliance (RDA) activities.

DataDNS: Discovery, access, and in-place analysis for large scale data



- Emerging vision from NDSC
- Data analysis and compute discovery engine.
- Connecting services that provide in-place access to data analysis environments.
- Defines common interface for heterogeneous environments.

Workshop themes

- Users/researchers and why they are using container-based environments (e.g., consistent environments, packaging applications, shifting computation around)
- Interactive vs non-interactive use cases including approaches to workflow management, orchestration, scalability, and moving analysis to HPC environments
- Data/storage focusing on models for managing and exposing data via containers
- Security/permissions
- Challenges/what new use cases are we seeing?

Workshop format

- Today:
  - Presentations + Discussions (~20+10)
- Tuesday/Wednesday
  - Self-organized
  - Breakout groups/deep-dive presentations
- Output: Workshop report
  - Common use cases with examples
  - Best practices/recommendations
  - Challenges/new frontiers/unsolved problems

Communication

- Please confirm that you have access to:
- Slack:
  - https://nationaldataservice.slack.com/ #container-workshop
- Google Drive:
  - https://drive.google.com/drive/folders/0B8i3lba1_nLwVmt4LUxuaTV6NTQ

Upcoming: 8th NDS Consortium Workshop

October 10–12, 2017
Boulder, Colorado
NIST

Goals:

- Specification of value propositions for data service providers as potential goals for the National Data Service.
- Identify areas a national data service can bridge between missing or nascent areas of a data ecosystem.
- Identify synergies and support models between NDS and ecosystem component providers.
- Identify targeted data services relevant to moving away from siloed systems with potential for development with NDS resources and additional support.

http://www.nationaldataservice.org/get_involved/events/NDS8

Candidate topics

Poll: http://slido.com/container-workshop

- Container orchestration/scaling
- Interactive analysis environments
- Batch job support
- Integration with HPC
- Shared storage across containers
- Security/permissions
- Containerizing applications
- Managing images
- Archiving/preservation of images
- Supporting X windows applications
- Supporting licensed software
- Licensed software
- Other ….?