Architecture / Overview

Computation: “something of a Black Art”
Publication / Sharing:
Tools => Apps => Workflows
Shared for verification, further research
Use Case: Virtual Institute for Social Research (VISR)

A platform for services & tools...

...to sustain the data lifecycle...

...and enable better science.

- More mileage from every dataset
- More transparency & replicability
- More collaboration
- New insights

Diagram by Jon Crabtree, Odum Institute, UNC-Chapel Hill
Use Case: Replication / Verification

1. Article submission
2. Conditional accept
3. Data submission
4. Data verification
5. Data publication
6. Final accept
7. Article publication

Diagram by Thu-Mai Christian, Odum Institute, UNC-Chapel Hill
Insights from the DE Evolution: What can we learn from CyVerse’s design

Don Sizemore, Mike Conway, Tony Edgin
DRY Principle

• We shouldn't be repeating ourselves. How much can we learn from successful systems?
• iRODS is a set of capabilities with (originally) few opinions on how these raw materials are combined into higher level solutions.
Mining the DE Design!

• Check out the API Doc for their API Gateway at:
  • https://cyverse-de.github.io/api/endpoints/
  • A pretty good overview of the kinds of services built on top of the iRODS stack

• Check out the additional database schema at:
  • https://github.com/cyverse-de/de-db
  • https://github.com/cyverse-de/metadata-db
  • https://github.com/cyverse-de/permissions-db
  • https://github.com/cyverse-de/notifications-db
  • What sorts of additional persistent information is needed outside of iRODS? What choices were made for performance or other optimizations?
FAIR Data and Computation

- Sharing of apps and data
- Discovery of apps and data
- Asynch (and synch coming) execution of high-performance and high-throughput apps
- Notification system
- Data staging, provenance tracking
Interest in Community DE

• Stems back from the days of the DataNet Federation Consortium!
• Lots of work to ease pain of deployment, CI is hard!
• DE challenges scalability of catalog, how to offload search and other activities with iRODS at the center?
• As with Dataverse, how applicable is infrastructure built for a particular domain to other domains?
• What do these experiences tell us about how iRODS fits, what it lacks, what other pieces of the ecosystem work well?
In the next major release:
Interactive GUI apps (Jupyter, Rstudio)

Don Sizemore, Mike Conway, Tony Edgin