



Cyverse Discovery Environment:

UNC's Implementation of the Community Edition

Don Sizemore, Mike Conway, Tony Edgin

THE ODUM INSTITUTE

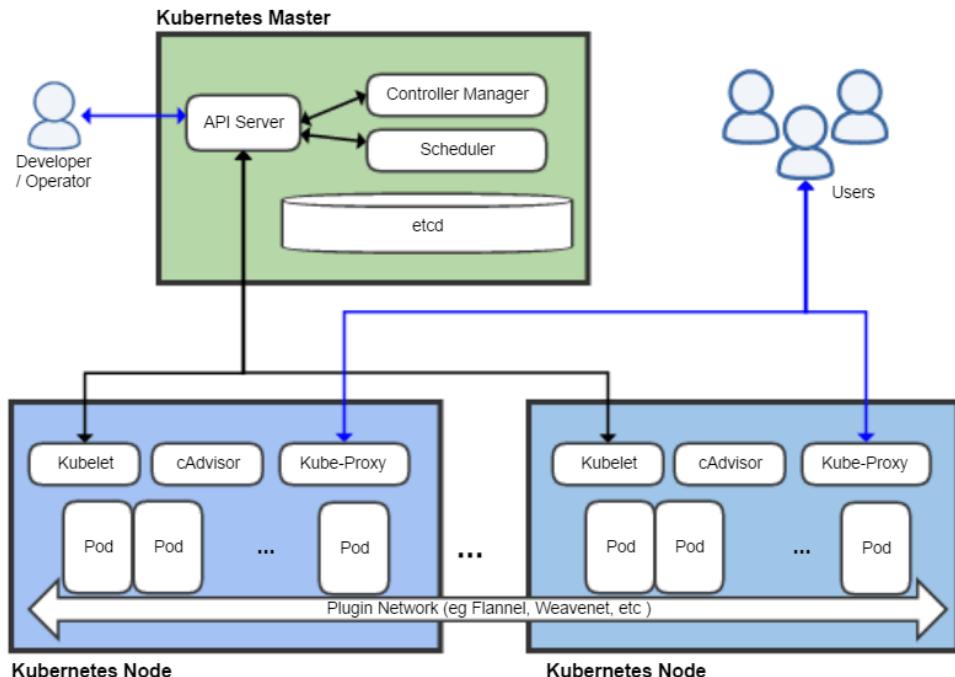
FOR RESEARCH IN SOCIAL SCIENCE

228 DAVIS LIBRARY, CB# 3355

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NC 27599-3355

WWW.ODUM.UNC.EDU

Architecture / Overview



Source: Wikipedia



iRODS

Data: dls

Upload · File · Edit · Download · Share · Metadata · Refresh · Trash

Navigation

- dls
 - analyses
 - Community Data
 - Shared With Me
 - Trash
 - Favorites

Viewing: /BrainArchiveZone/home/dls

Name	Last Modified	Size	More
analyses	2018 Mar 1 08:58:50		More
01_create_experiment.png	2018 Mar 1 09:53:49	15.29 KB	More
<input checked="" type="checkbox"/> dataset-nba.csv	2018 Apr 10 10:05:57	12.41 KB	More
Demo-nba.Rcode.R	2018 Apr 10 10:05:57	2.86 KB	More
test2.txt	2018 Feb 2 08:29:10	6 bytes	More
test3.txt	2018 Apr 9 14:32:09	36 bytes	More
test.txt	2018 Feb 2 08:29:11	5 bytes	More

Details

Last Modified: 2018 Apr 10
Date Submitted: 2018 Apr 10
Permissions: own
Share: Begin sharing
Size: 12.41 KB
Type: text/csv
Info-Type: Select
md5 Checksum: sha2:C3...
Send to:
Tags: Search

Displaying 1 - 9 of 9 1 item(s)

Apps

Apps · Workflow · Share · Refresh · Search Apps · Manage Tools · Switch View

Categories

- My Apps
- Topic
- Operation
- HPC

Sort By: Name

Apps under development

Favorite Apps	Apps under development	My public apps	Shared with me
 Dataverse Publisher	 "Don Sizemore" dls	 Rscript	

dataset-nba.csv

dataset-nba.csv

Save · Refresh · Wrap Text · Line Numbers · Not Editable

year	team	wins	champ	eastern	topall	topconf
2015	Atlanta Hawks	60	0	1	0	0
2015	Cleveland Cavaliers	53	0	1	0	0
2015	Chicago Bulls	50	0	1	0	0
2015	Toronto Raptors	49	0	1	0	0
2015	Washington Wizards	46	0	1	0	0
2015	Milwaukee Bucks	41	0	1	0	0
2015	Boston Celtics	40	0	1	0	0
2015	Brooklyn Nets	38	0	1	0	0
2015	Golden State Warriors	67	1	0	1	1
2015	Houston Rockets	56	0	0	0	0
2015	LA Clippers	56	0	0	0	0
2015	Memphis Grizzlies	55	0	0	0	0
2015	San Antonio Spurs	55	0	0	0	0
2015	Portland Trail Blazers	51	0	0	0	0
2015	Dallas Mavericks	50	0	0	0	0
2015	New Orleans Pelicans	45	0	0	0	0
2014	Indiana Pacers	56	0	1	0	0
2014	Miami Heat	54	0	1	0	0
2014	Toronto Raptors	48	0	1	0	0
2014	Milwaukee Bucks	48	0	1	0	0
2014	Washington Wizards	44	0	1	0	0
2014	Brooklyn Nets	44	0	1	0	0
2014	Charlotte Bobcats	43	0	1	0	0
2014	Atlanta Hawks	38	0	1	0	0
2014	San Antonio Spurs	62	1	0	1	1
2014	Oklahoma City Thunder	59	0	0	0	0
2014	LA Clippers	57	0	0	0	0
2014	Houston Rockets	54	0	0	0	0
2014	Portland Trail Blazers	54	0	0	0	0

Page Size (KB) | 1 of 2 | > | >> |

Analyses

Analyses · Edit · Refresh · Share · All · Search by Name or App

Name	Owner	App	Start Date	End Date	Status
DE_Word_Count_a...	dls@renci.org	DE Word C...	2018 Apr 18 1...	2018 Apr 18 1...	Failed
Rscript_analysis1	dls@renci.org	Rscript	2018 Apr 10 1...	2018 Apr 10 1...	Completed
Rscript_analysis1	dls@renci.org	Rscript	2018 Apr 10 1...	2018 Apr 10 1...	Failed
R_analysis1	dls@renci.org	R	2018 Apr 10 1...	2018 Apr 10 1...	Failed
R_analysis1	dls@renci.org	R	2018 Apr 10 1...	2018 Apr 10 1...	Failed
DE_Word_Count_a...	dls@renci.org	DE Word C...	2018 Apr 9 14...	2018 Apr 9 14...	Completed
DE_Word_Count_a...	dls@renci.org	DE Word C...	2018 Apr 9 10...	2018 Apr 9 10...	Completed
DE_Word_Count_a...	dls@renci.org	DE Word C...	2018 Mar 27 1...	2018 Mar 27 1...	Completed
DE_Word_Count_a...	dls@renci.org	DE Word C...	2018 Mar 26 1...	2018 Mar 26 1...	Completed
DE_Word_Count_a...	dls@renci.org	DE Word C...	2018 Mar 15...	2018 Mar 15...	Failed
DE_Word_Count_a...	dls@renci.org	DE Word C...	2018 Mar 10...	2018 Mar 10...	Failed

Displaying 1 - 11 of 11 0 item(s)

Computation: “*something of a Black Art*”

Name	Description	Location	Type	Attribution	Version
wc	Word Count		executable		0.0.1
python	Python 2.7 with no networking, a 1GB RAM limit, and a 10% cpu share. Entrypoint is python.	/usr/local/bin	executable		1.0.0
swc	Sarah's Word Count		executable		1.0.0
foosh	Foosh		executable		1.0.0
R base			executable		3.4.4
Dataverse Publisher	Publishes a file to a given dataset.		executable		1

Create App

Save Preview Command Line Order

App Items

New app

Tool used:
enter 3 or more characters

* App name:
New app

* App description:
Enter a description for your new app (<255 characters)

Files/Folders

Multiple Input Files
Add Delete
Files

Input File

File Browse

Input Folder

Folder Browse

Text/Numerical Input

Info Text

Single-line Text

Text

Multi-line Text

Text

Checkbox

Environment Variable

\$ENV_VAR

Integer

10

Decimal

1.0

Lists

List Option

Integer List

1

Decimal List

1.0

Command line view

App Items

Section

Section ...

Tool used:

enter 3 or more characters

* App name:

New app

* App description:

Enter a description for your new app (<255 characters)

Section 1

Multiple Input Files Label - 1:

Add Delete

Name

Select multiple input files. Tip: You can also drag and drop files from the Data window.

Details: Multiple Input Files Label - 1

Multiple Input Files Selector label: Multiple Input Files Label - 1

Argument option: Enter argument option

Make this field required.

Exclude this item if nothing is entered.

Tool tip text: Enter tool tip here

Type of information contained in these files: Unspecified

Do not pass this argument to command line

Repeat "Argument option" on the command line before each filename

Apps

Apps Workflow Share Refresh Search Apps Manage Tools Switch View

Categories: My Apps Topic Operation HPC

Sort By: Name

Favorite Apps Apps under development My public apps Shared with me

Dataverse Publisher "Don Sizemore" dis de 5 stars (0) lock

Rscript Manage Sharing Selected Apps Name Dataverse Publisher

Who has access

Name Permissions

jonl

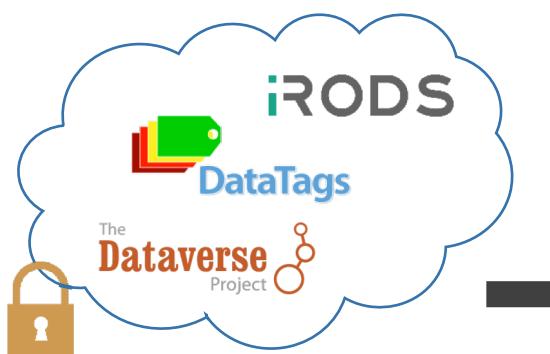
Choose Collaborators

jonc Username: ***jon*** ***jon***@email.unc.edu

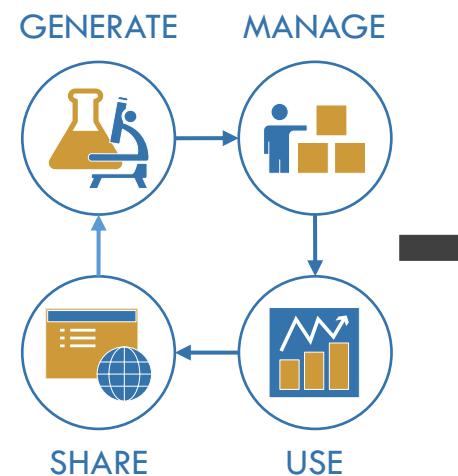
Done Cancel

Use Case: Virtual Institute for Social Research (ViSR)

A platform for
services & tools....



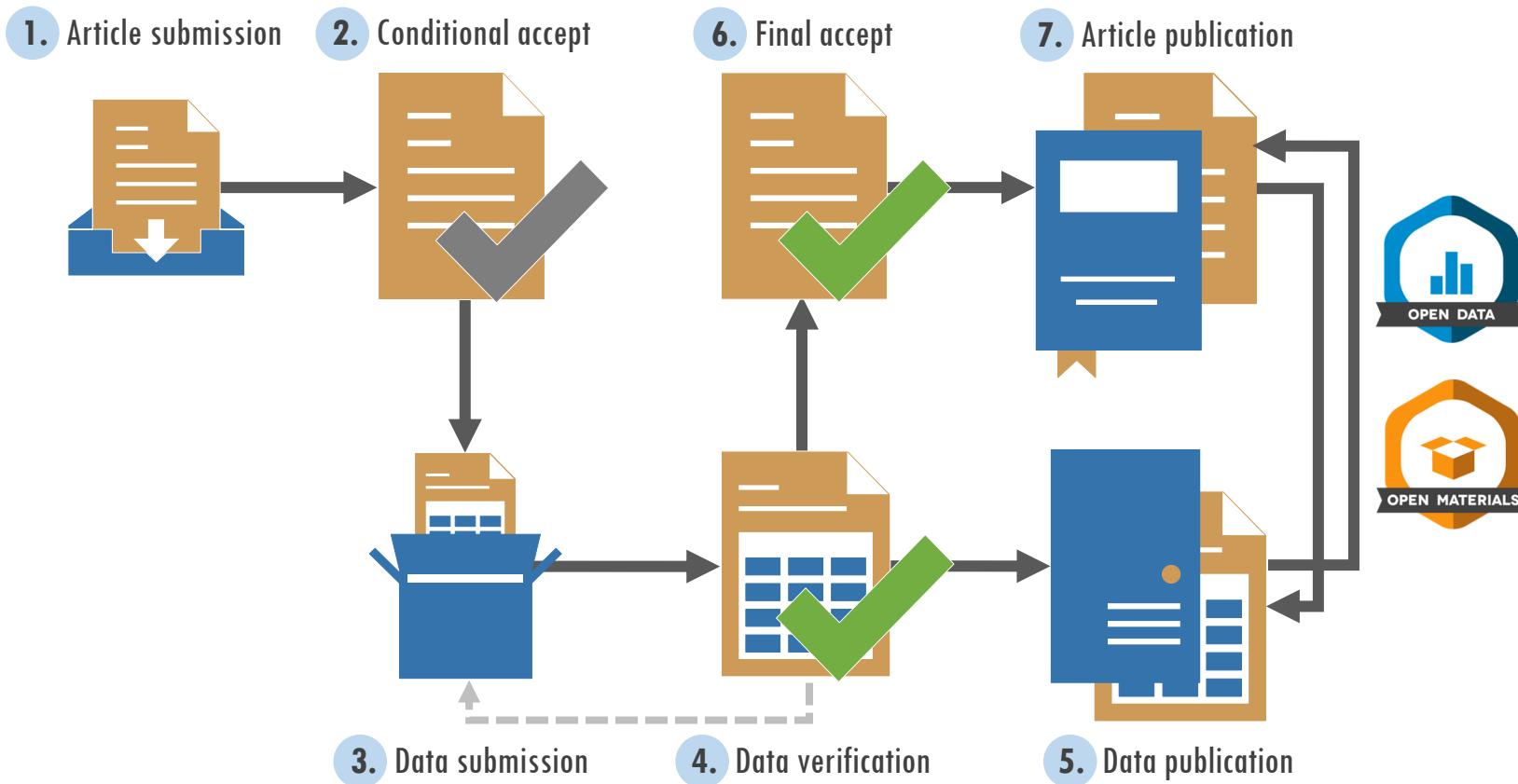
...to sustain the
data lifecycle...



...and enable
better science.

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

Use Case: Replication / Verification





Insights from the DE Evolution:

What can we learn from CyVerse's design

Don Sizemore, Mike Conway, Tony Edgin

THE ODUM INSTITUTE

FOR RESEARCH IN SOCIAL SCIENCE

228 DAVIS LIBRARY, CB# 3355

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NC 27599-3355

WWW.ODUM.UNC.EDU

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.

iRODS

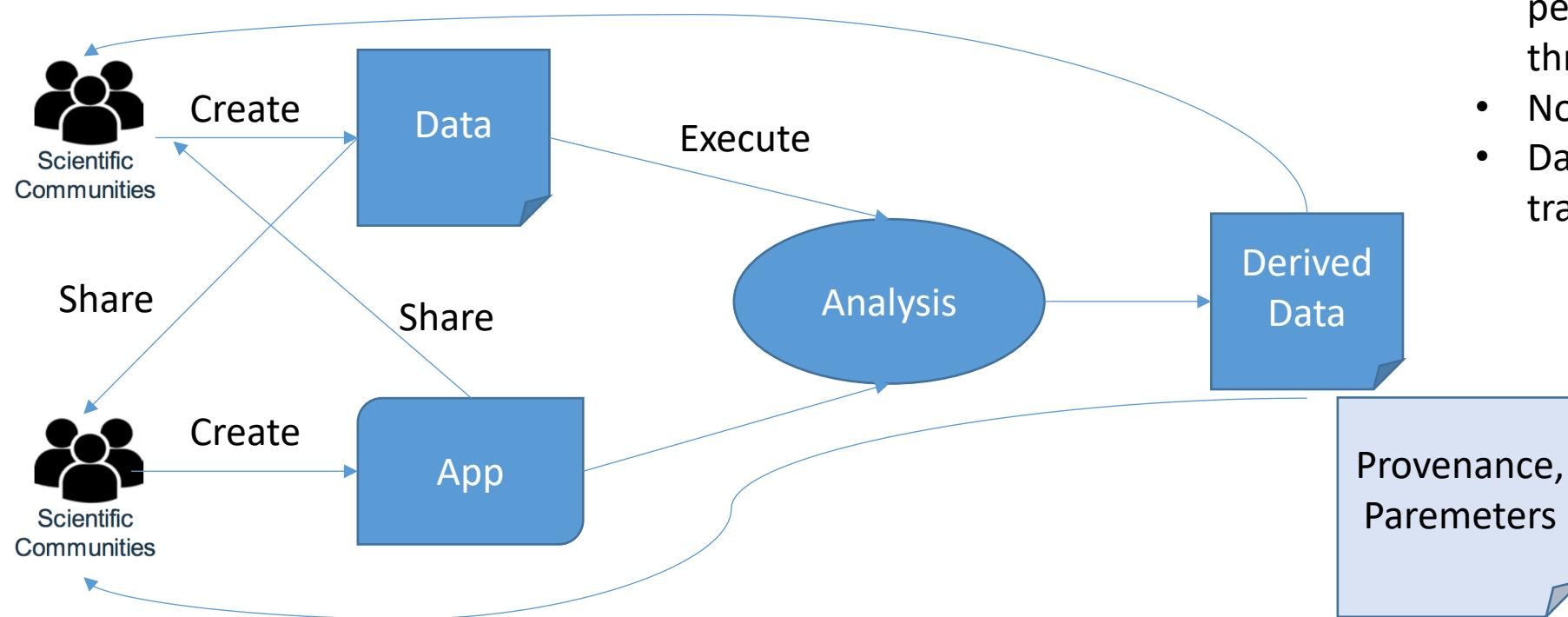
The Middle Bits

Site Specific Deployment

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

THE ODUM INSTITUTE

FOR RESEARCH IN SOCIAL SCIENCE

228 DAVIS LIBRARY, CB# 3355

UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NC 27599-3355

WWW.ODUM.UNC.EDU