

VIRTUAL INSTITUTE FOR SOCIAL RESEARCH: POWERED BY DISCOVERY ENVIRONMENT, IRODS AND DATAVERSE

MIKE CONWAY

DICE Group

JONATHAN CRABTREE

Odum Institute

MATTHEW KRAUSE

DICE Group

DON SIZEMORE

Odum Institute

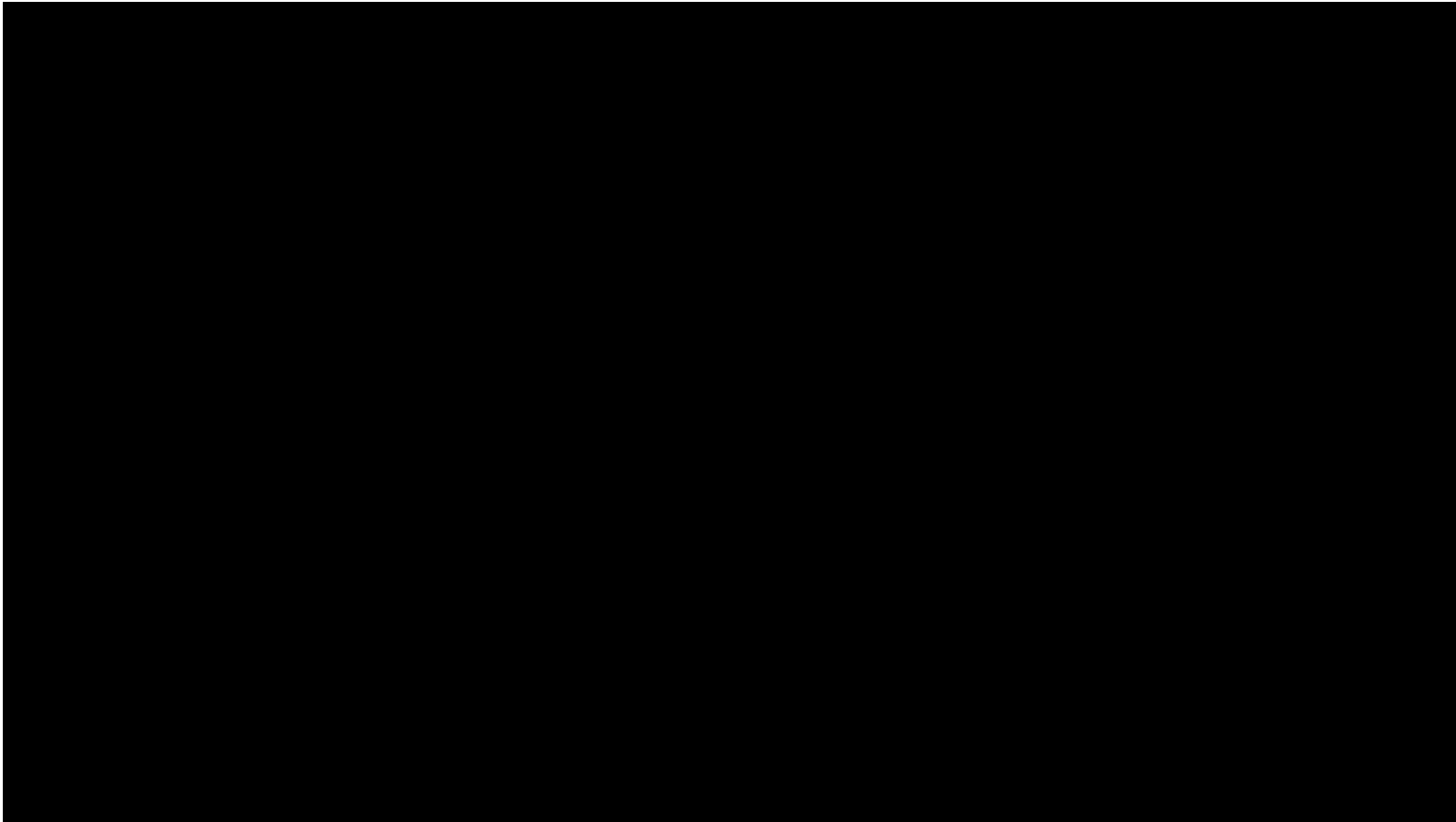
THE ODUM INSTITUTE FOR RESEARCH IN SOCIAL SCIENCE
DAVIS LIBRARY, 2ND FLOOR, CB# 3355
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
WWW.ODUM.UNC.EDU

iRODS
USER GROUP MEETING 2016

OVERVIEW

- What is VISR
- VISR components
- Discovery Environment
- VISR use cases and next steps
- DFC/Odum efforts to adopt DE
- Integrating mobile devices
 - Matt & Ceasar

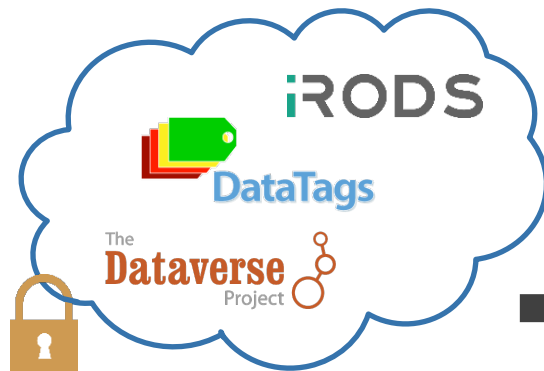
VISR/DISCOVERY ENVIRONMENT DRAFT OVERVIEW VIDEO



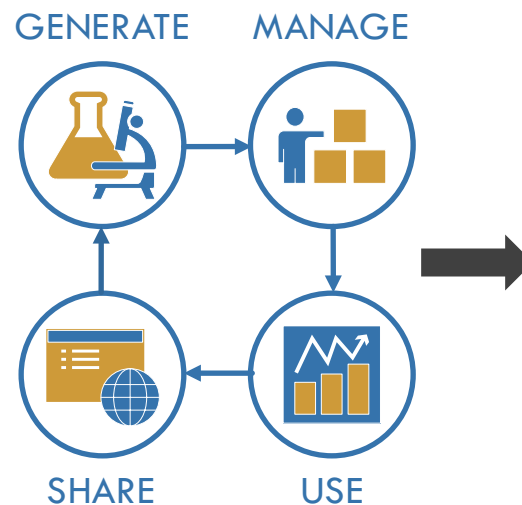
UNC
THE ODUM INSTITUTE

VIRTUAL INSTITUTE FOR SOCIAL RESEARCH (VISR)

**A platform for
services & tools...**



**...to sustain the
data lifecycle...**

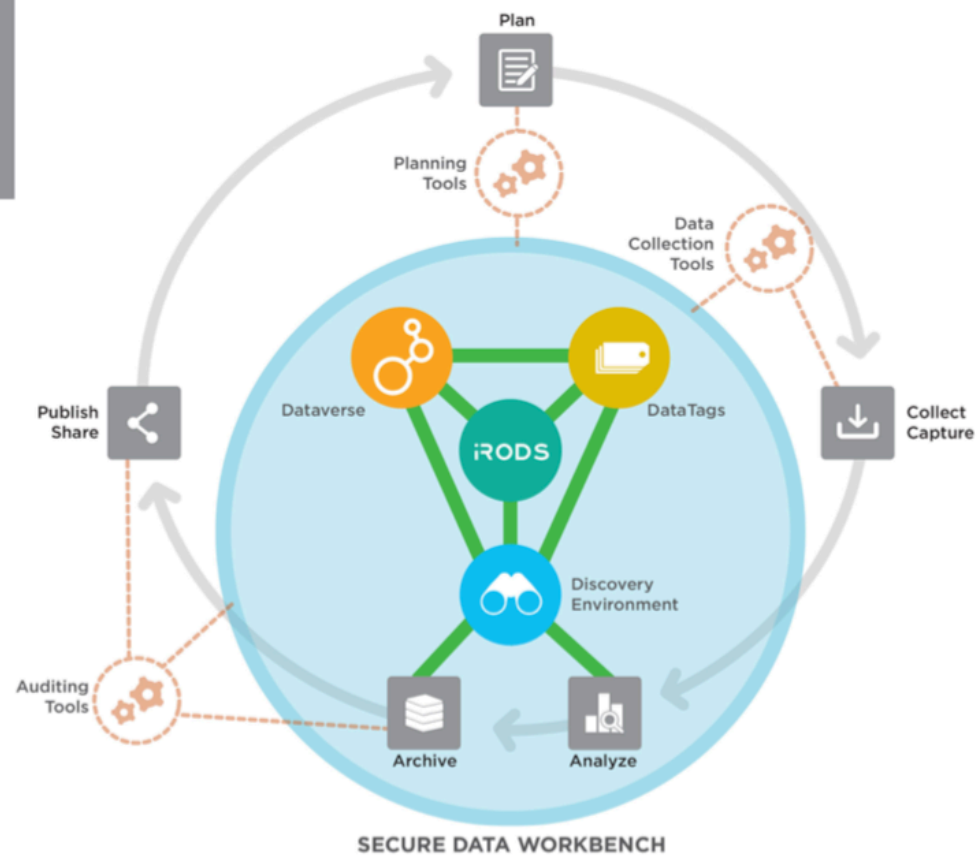


**...and enable
better science.**

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

VIRTUAL INSTITUTE FOR SOCIAL RESEARCH (VISR)

Research
Data
Curation
Lifecycle



COUPLING SYSTEMS FOR SOCIAL SCIENTISTS

- CyVerse Discovery Environment (DE)
- iRODS
- Dataverse

Powered by



iRODS

A Web Application for Publishing, Citing,
Analyzing and Preserving Research Data

The
**Dataverse
Network**TM
Project



UNC
THE ODUM INSTITUTE

ODUM INSTITUTE DATAVERSE

<http://arc.irss.unc.edu/dvn/>

POWERED BY THE **Dataverse Network** PROJECT
7.5.6.3-odum-02

Odum Institute Dataverse Network

Search this Dataverse Network

Search

Advanced Search Tips

The Odum Institute Dataverse Network provides access to data collections curated by the Odum Institute as well as collections owned by other institutions and individual scholars. You can search across or browse any of these dataverses listed below. You may also create your own branded dataverse to manage and provide access to your data. Learn more about the Dataverse Network Project at <http://dataverse.org/> or view our [video tutorials](#).

Many files have access restrictions. For more information, click [Odum File Restrictions](#).

Dataverses

51 Dataverses

A **Dataverse** is a container for research data studies, customized and managed by its owner.

BROWSE DATAVERSES BY FILTER BY TYPE

Odum Institute Archive Dataverse (1)

Scholar Dataverses (14)

Institution Dataverses (36)

Journal Dataverses (1)

RECENTLY RELEASED DATAVERSES

Betty Nance-Floyd

Jan 8, 2016

Studies

31,881 Studies, 146,162 Files, 528,113 Downloads

A **study** is a container for a research data set. It includes cataloging information, data files and complementary files.

RECENTLY RELEASED STUDIES

Emory University Research Data Management Faculty Survey by Akers, Katherine G.; Doty, Jennifer


Mar 1, 2016

Determinants of Exclusive Breastfeeding in Rural Kenya, 2013 by Gewa, Constance A.

Feb 26, 2016

Monmouth University New Hampshire GOP Primary Poll, Number 88 by Monmouth University Polling Institute

Feb 22, 2016

 **UNC**
THE ODUM INSTITUTE

ODUM INSTITUTE DATAVERSE

31,896
PUBLISHED STUDIES



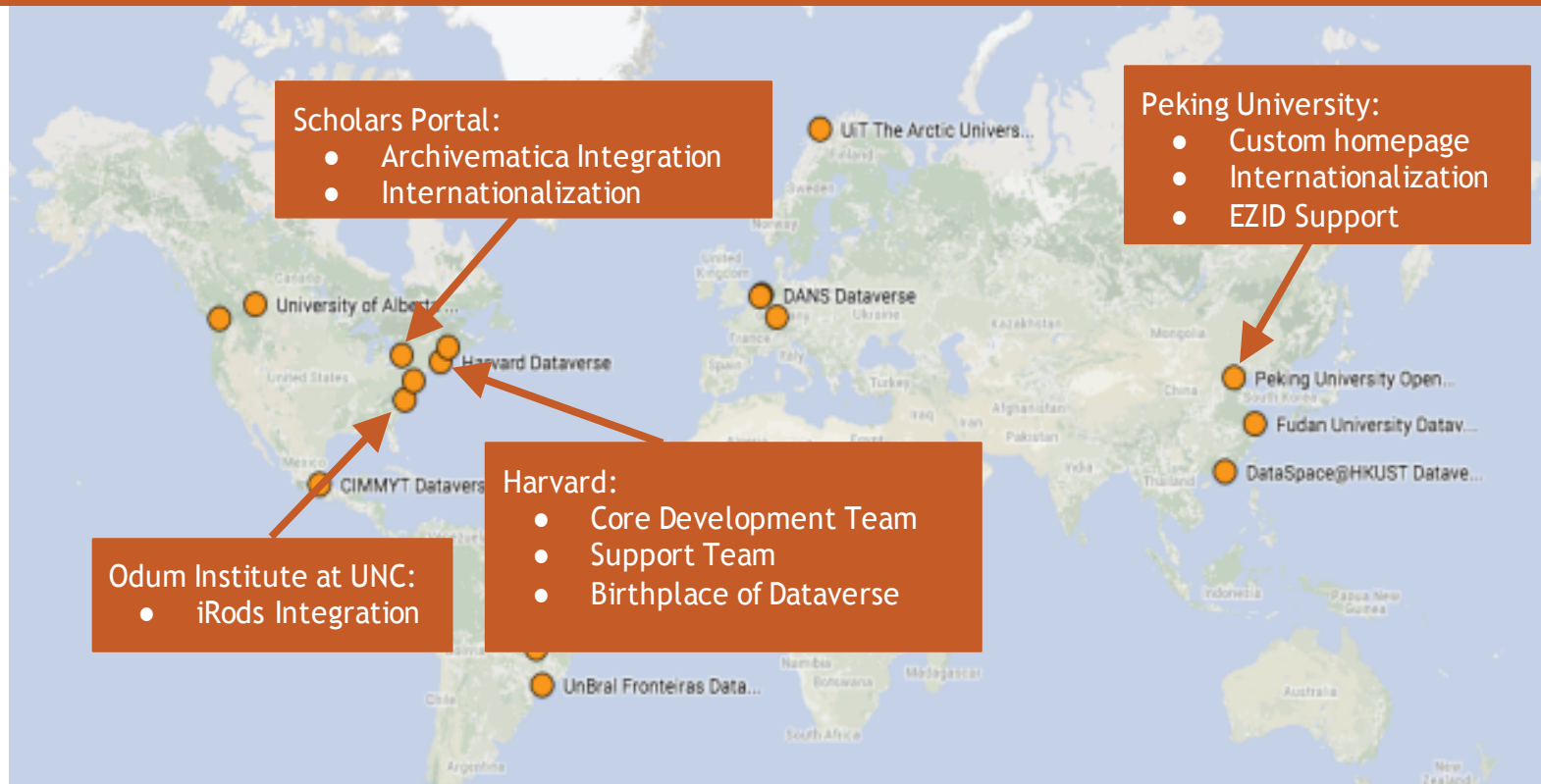
146,272
AVAILABLE FILES



534,728
FILE DOWNLOADS



GROWING DATAVERSE COMMUNITY



VISR DE IN ALPHA PHASE TESTING

Discovery Environment

35

Data

Apps

Analyses

Analyses

Analyses Edit Refresh Show All Filter by Name or App

Name	Owner	App	Start Date	End Date	Status
Odum_R_Demo_anal...	testde1@irss.unc...	Odum R Demo	2016 Apr 1 13:43...	2016 Apr 1 13:45...	Completed
Odum_R_Demo_anal...	testde1@irss.unc...	Odum R Demo	2016 Apr 1 13:37...	2016 Apr 1 13:38...	Failed
CrabtreeBagit_analysi...	testde1@irss.unc...	CrabtreeBagit	2016 Apr 1 13:32...	2016 Apr 1 13:33...	Completed
Odum_R_Demo_anal...	testde1@irss.unc...	Odum R Demo	2016 Apr 1 13:29...	2016 Apr 1 13:31...	Completed
CrabtreeBagit_analysi...	testde1@irss.unc...	CrabtreeBagit	2016 Apr 1 10:55...	2016 Apr 1 10:56...	Completed
CrabtreeBagit_analysi...	testde1@irss.unc...	CrabtreeBagit	2016 Apr 1 10:48...	2016 Apr 1 10:56...	Completed
CrabtreeBagit_analysi...	testde1@irss.unc...	CrabtreeBagit	2016 Apr 1 10:45...	2016 Apr 1 10:46...	Completed

Displaying 1 - 7 of 7

0 item(s)

Apps

Apps Workflow Share Search Apps

Category

Public Apps (1)

- Beta (1)
- Odum (0)

Workspace (2)

- Apps under development (2)
- Favorite Apps (0)
- My public apps (0)
- Shared with me (0)
- High-Performance Computing (-1)

Workspace >> Apps under development

Name	Integrated by	Rating
CrabtreeBagit	testde1 testd...	☆☆☆☆☆ (0)
Odum R Demo	testde1 testd...	☆☆☆☆☆ (0)

Data: Crabtree

Upload File Edit Download Share Refresh

Navigation

Crabtree

testde1

- analyses
- BioProject_odumTest
- Crabtree
- irsynctest
- OdumRdemo
- test2fromDE
- Community Data
- Shared With Me
- Trash
- Favorites

Viewing: /OdumDEZone/home/testde1/Crabtree

Name	Last Modified	Size
AnalysisResults	2016 Apr 1 13:36:46	
CrabtreeBagit_analysis1-2016-04-01-17-33-02.4	2016 Apr 1 13:33:07	
assign3.cta	2016 Apr 1 13:34:50	16.52 KB
assign5.RData	2016 Apr 1 13:26:41	7.49 KB
Crabtree3.R	2016 Apr 1 13:43:03	812 bytes
Crabtree5.R	2016 Apr 1 13:28:02	3.61 KB
Slide1.gif	2016 Mar 31 20:56:52	38.49 KB

Displaying 1 - 7 of 7

0 item(s)

Details

Select a file or folder to view its details

Feedback

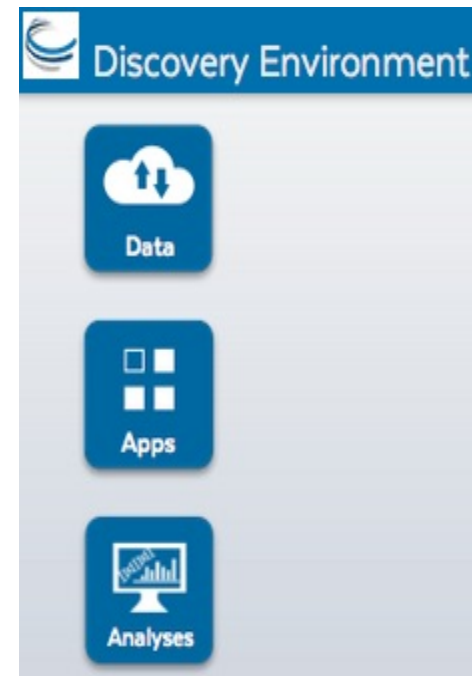
Data: Crabtree

Apps

Analyses

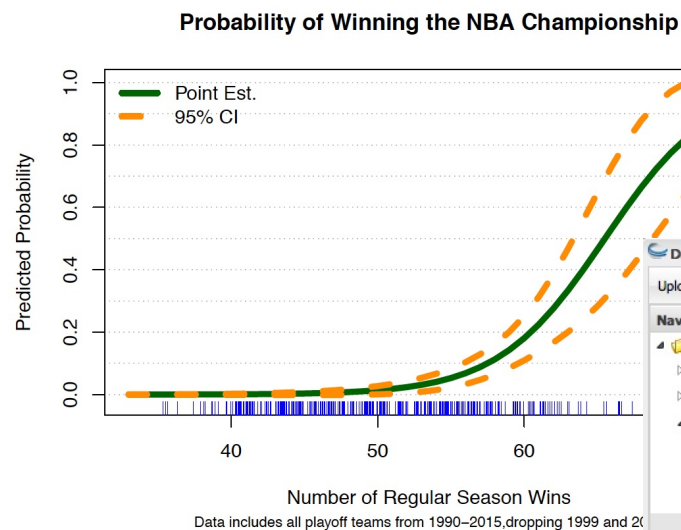
DATA MANAGEMENT, ANALYSIS, AND SHARING

- Expandable with DOCKER tools
- Create custom workflows and applications
- Upload, share and manage data
- Share and exchange custom workflows
- Scheduling via CONDOR
- Powered by iRODS data management
- Scalable analysis with simple to use interface
- Data accessible with any iRODS interface
- Ability to add and track custom metadata



CURRENT ODUM DEMONSTRATIONS

- R Statistical Analysis
- Library of Congress BagIt Packager



Apps

Apps ▾ Workflow ▾ Share ▾ Search Apps

Category << **Workspace >> Apps under development**

Name	Integrated by	Rating
CrabtreeBagIt	testde1 testd...	☆☆☆☆☆ (0)
Odum R Demo	testde1 testd...	☆☆☆☆☆ (0)

Public Apps (1)

- Beta (1)
- Odum (0)

Workspace (2)

- Apps under development (2)
- Favorite Apps (0)
- My public apps (0)
- Shared with me (0)

Data: mynewbag

Upload ▾ File ▾ Edit ▾ Download ▾ Share ▾ Refresh

Navigation

- testde1
 - analyses
 - BioProject_odumTest
 - Crabtree
 - AnalysisResults
 - CrabtreeBagIt_analysis1-2016-04-01-17-33-02.4
 - logs
 - mynewbag
 - irsynctest
 - OdumRdemo
 - test2fromDE
 - Community Data
 - Shared With Me
 - Trash
 - Favorites

mynewbag

Viewing: /OdumDEZone/home/testde1/Crabtree/CrabtreeBagIt_an

Name	Last Modified
data	2016 Apr 1 13:33:08
bag-info.txt	2016 Apr 1 13:33:15
bagit.txt	2016 Apr 1 13:33:15
manifest-md5.txt	2016 Apr 1 13:33:11
tagmanifest-md5.txt	2016 Apr 1 13:33:10

Details

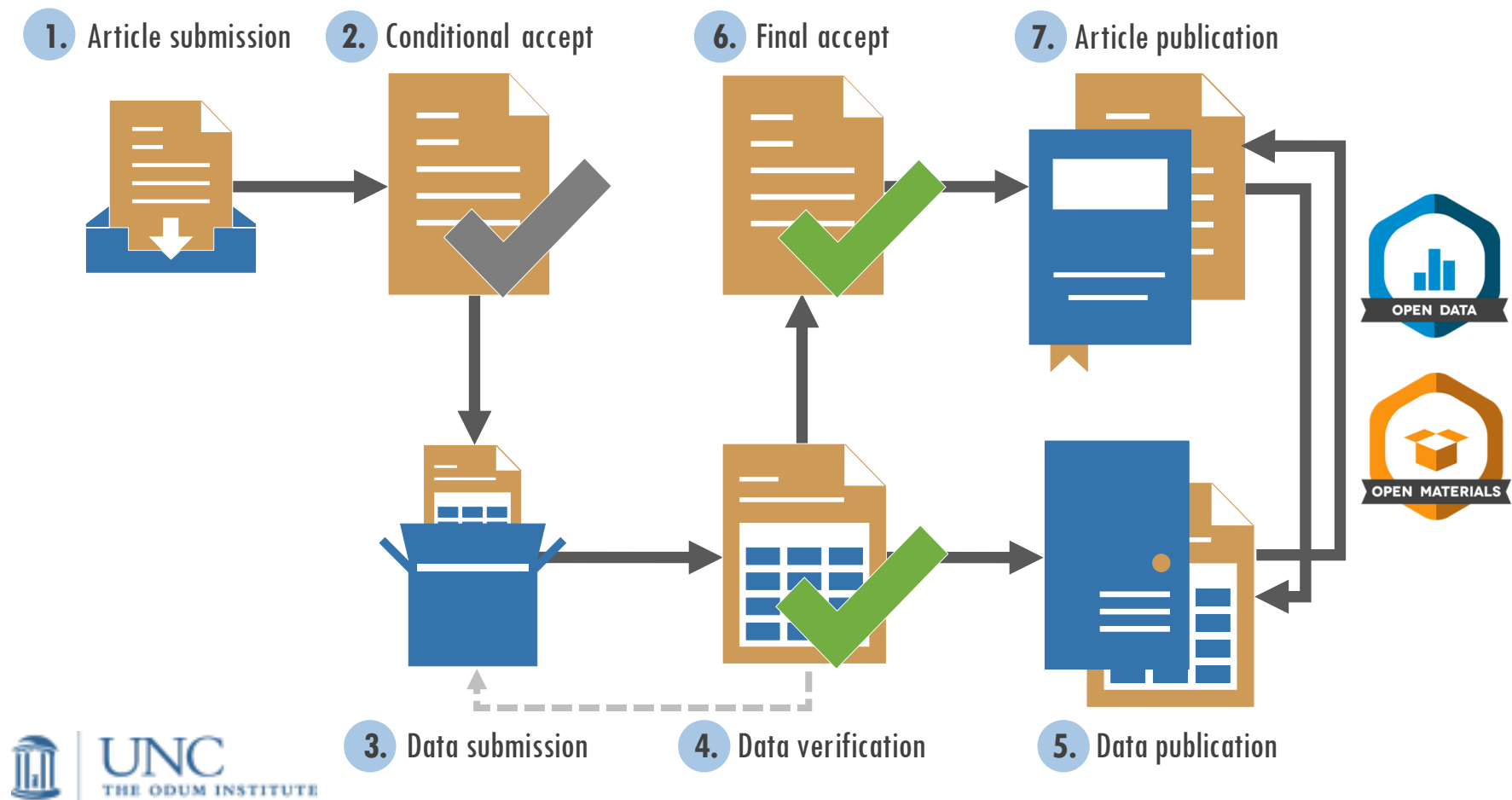
Select a file or folder to view its details

Displaying 1 - 5 of 5 0 item(s)

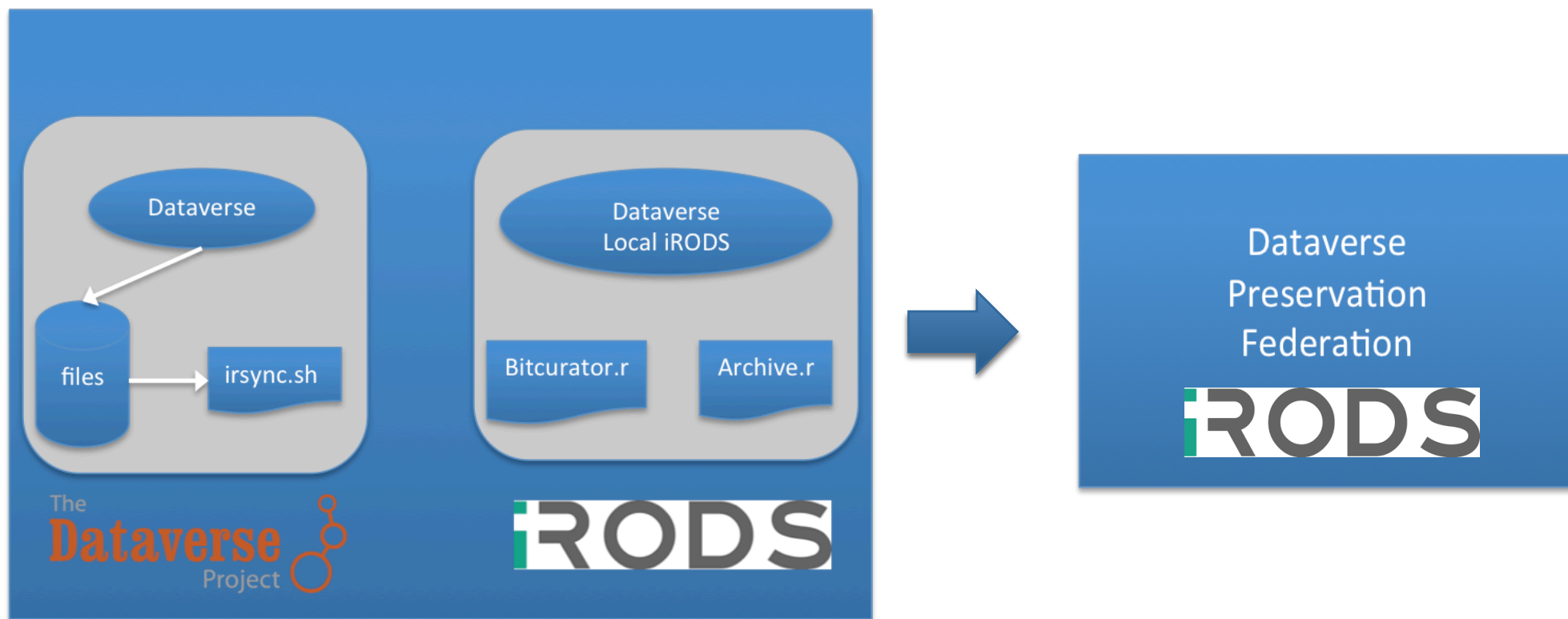
SOCIAL SCIENCE COMMUNITY BENEFITS FROM VISR

- Scalable resources
- Workflow management
- Job scheduling
- Community based data sharing
- Rules based data management
- Automated metadata generation
- Streamlined publication of data
- Secure environment to share sensitive data
- Shared applications and code

REPLICATION DATA VERIFICATION



AUTOMATED ARCHIVE PRESERVATION ACTIONS

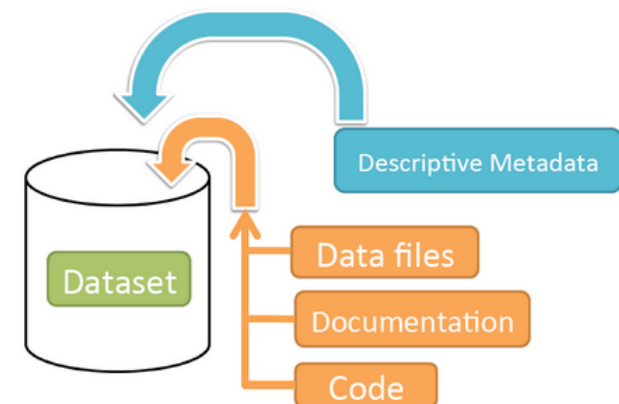


HIGH PRIORITY NEXT STEPS

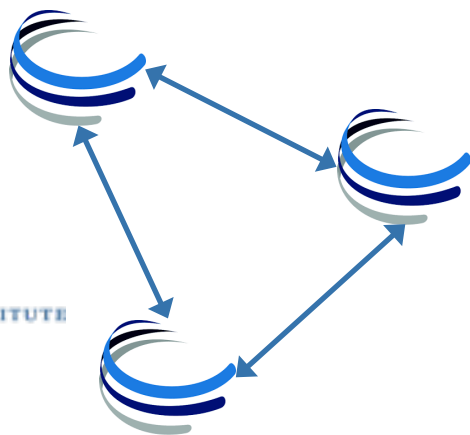
- Publish to Dataverse
- InCommon Authentication
- Handle Sensitive Data
- Federation between DE's

The
Dataverse
Project

Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.



InCommon®

CURRENT DISCOVERY ENVIRONMENT EFFORTS

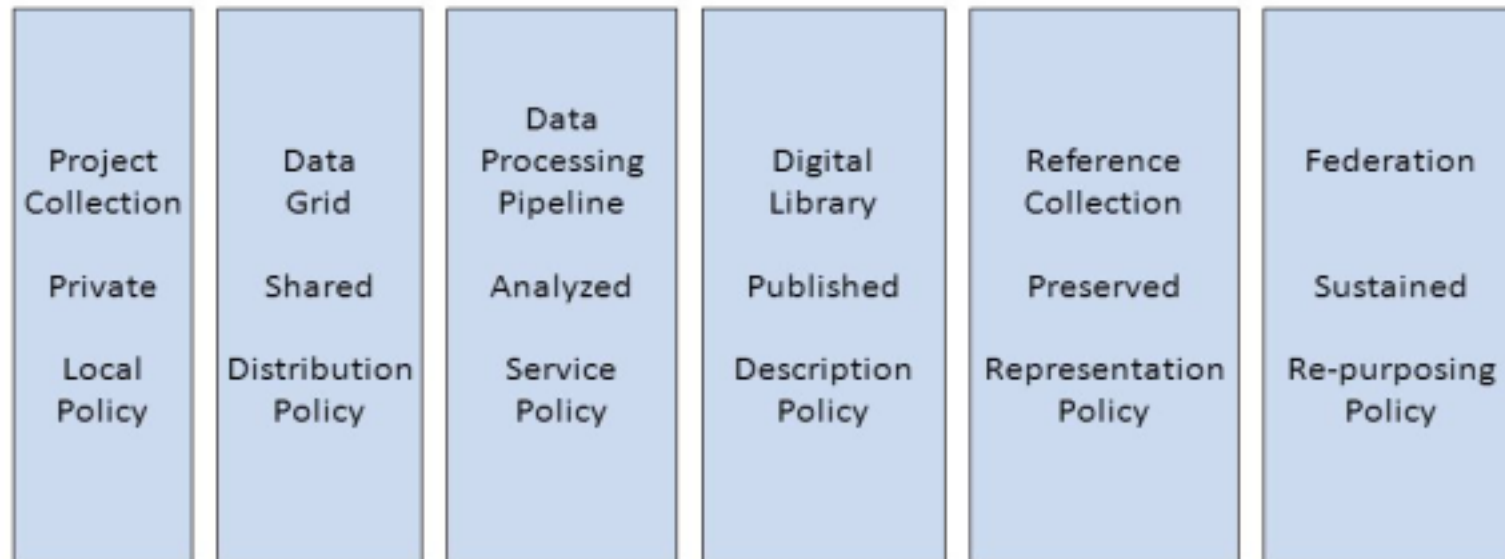
- iRODS on its way to industrial grade
- CyVerse adds a seasoned, production-ready middleware layer and user environment for data driven research

CURRENT DISCOVERY ENVIRONMENT EFFORTS

- Odum, CyVerse, and DFC developers have successfully brought DE up outside of CyVerse.
- <https://github.com/angrygoat/DE>
- Working towards a CyVerse 'community DE' release that we intend to focus on, and contribute to.

Community-based Collection Life Cycle

The driving purpose changes at each stage of the data life cycle



Stages correspond to addition of new policies for a broader community
Virtualize the stages of the collection life cycle through policy evolution

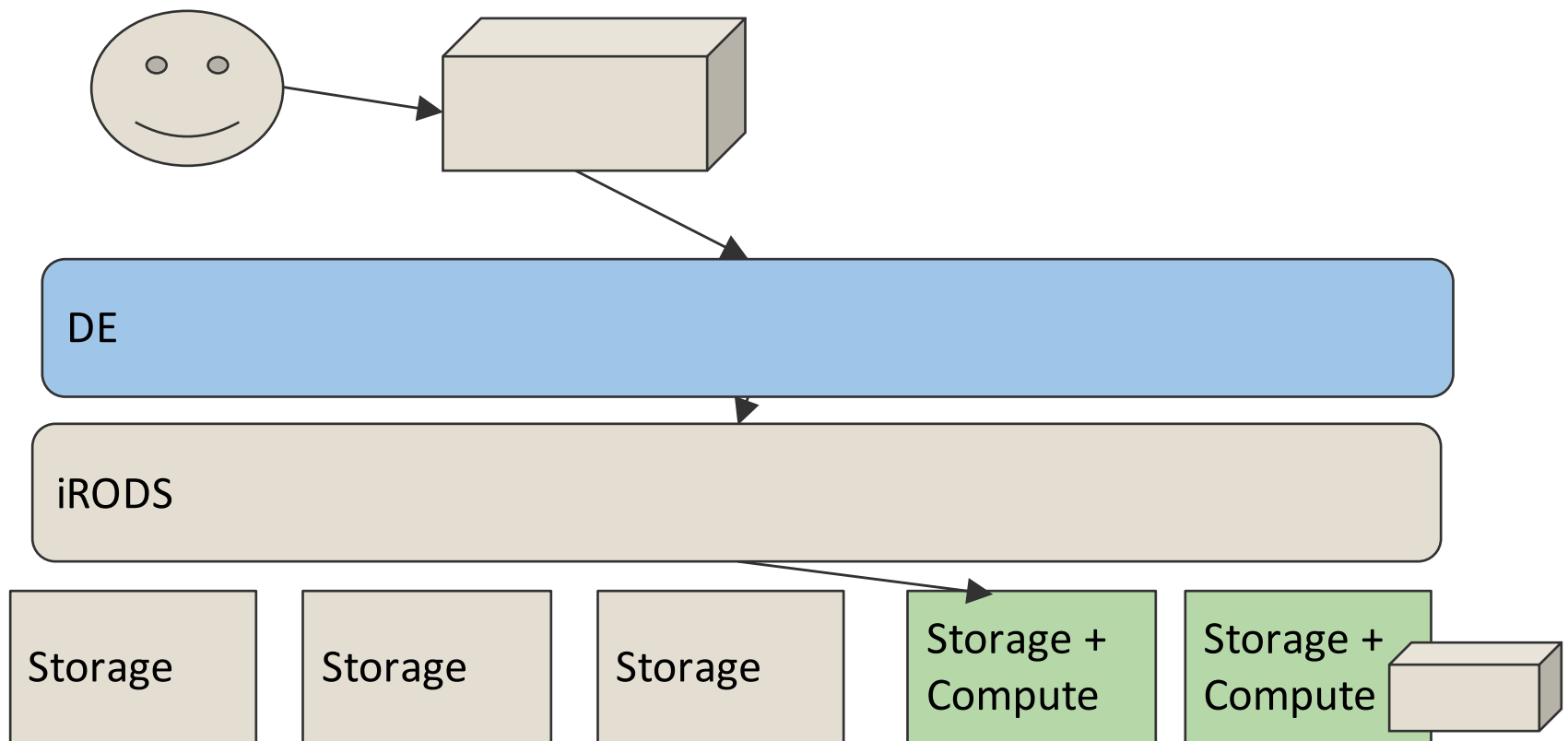
CURRENT STATE

- Developed a 'reference implementation' that includes default environment suitable for testing, prototyping.
- Successfully integrated social science, hydrology, and ocean science researchers through policies and computational tool provisioning.
- Right after User Group Meeting we are mutually doing production deployments

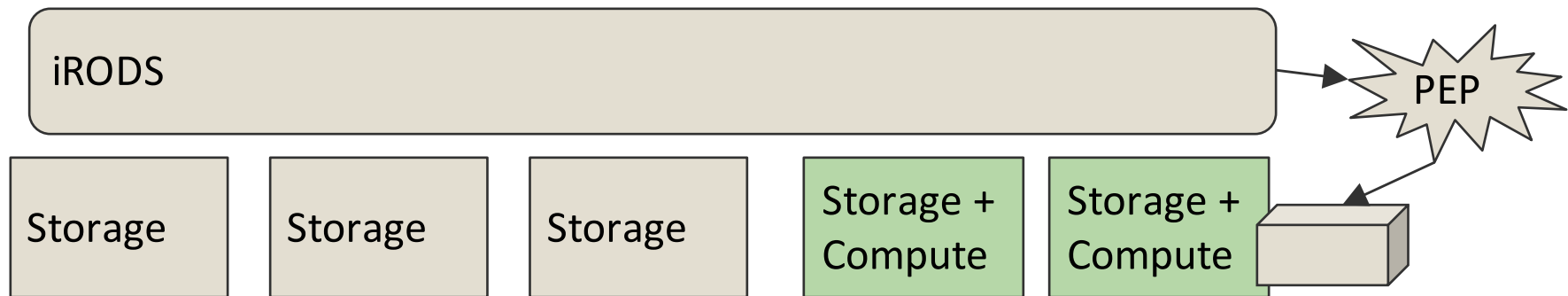
CURRENT DEVELOPMENTS/ENHANCEMENTS

- Adding a pluggable publish mechanism, capable of using 'apps' framework, that allows collections in iRODS to be formatted and disseminated using multiple vectors. Current targets are DataONE and Dataverse.
- Adding capability to publish grid data as a DataONE member node.

COMPUTATIONAL RESOURCES



COMPUTATIONAL RESOURCES



- Metadata extraction, NLP
- Subsetting
- Indexing
- Publishing

ACKNOWLEDGEMENTS AND THANKS TO MANY

- RENCI
- CyVerse
- DFC
- iRODS Consortium
- Don Sizemore
- Mike Conway
- Lisa Stillwell
- Reagan Moore
- Nirav Merchant
- Dennis Roberts
- Kasha Mammone
- Helen Tibbo
- Cal Lee
- Akio Sone
- Rodney Hodson
- Craig Willis



National Science Foundation Cooperative Agreement:
OCI-0940841

LINKS

- <http://dx.doi.org/10.7910/DVN/ITNETO>
- <https://github.com/DICE-UNC/dfc-dataverse-integration>
- <https://github.com/DICE-UNC/indexing-irods>
- <https://github.com/akio-sone/dvn/tree/Odum-Ext>
- https://github.com/donsizemore/odum/tree/master/irods_rules
- <http://datafed.org>
- <http://www.odum.unc.edu/odum/home2.jsp>
- <http://Dataverse.org>
- <http://nationaldataservice.org>

THANKS!

Jonathan Crabtree
Jonathan_Crabtree@unc.edu



Our Website
www.odum.unc.edu



Follow us on [Twitter](#)
[@Odum_Institute](#)



Subscribe to our [Channel](#)
The Odum Institute

