# iRODS Impact on Science and Data Management

iRODS UGM 2017

Ashok Krishnamurthy ,Kira Bradford, Michael Conway, Michael Shoffner, Justin James

## iRODS impact on data management for Scienctific domains: 2 Use Cases

#### • BRAIN-I

• A unified computation framework for analysis, storage, and visualization of 3D microscopy data of the brain

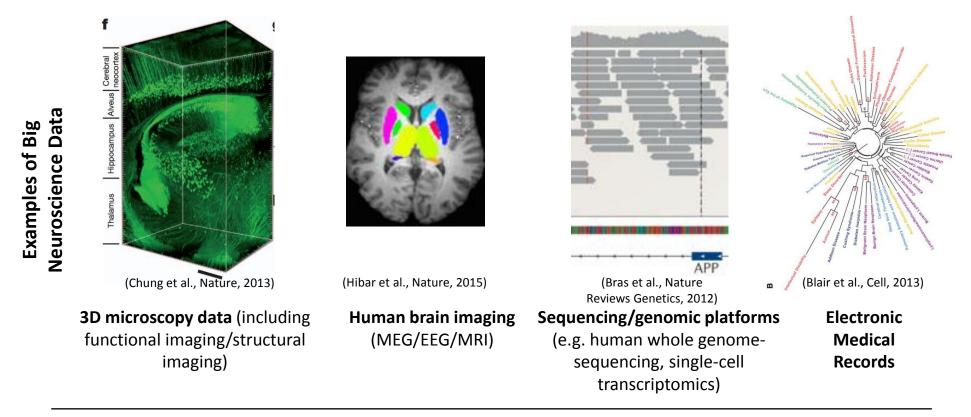
#### • SC2I

• Clinical decision support tools to improve medical outcomes in acute care



BRAIN-I: A unified computational framework for analysis, storage, and visualization of 3D brain microscopy data

#### Big Data Problems in Neuroscience



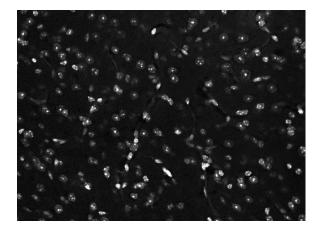
Confidentiality of human data

4

Computational infrastructure for storage, sharing and analysis of 3D microscopy images



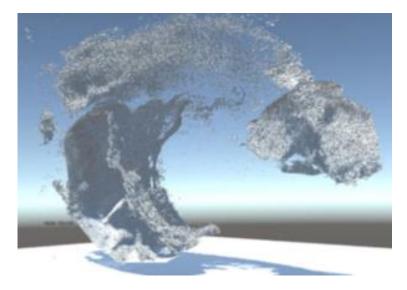
Novel segmentation tools to trace brain structure

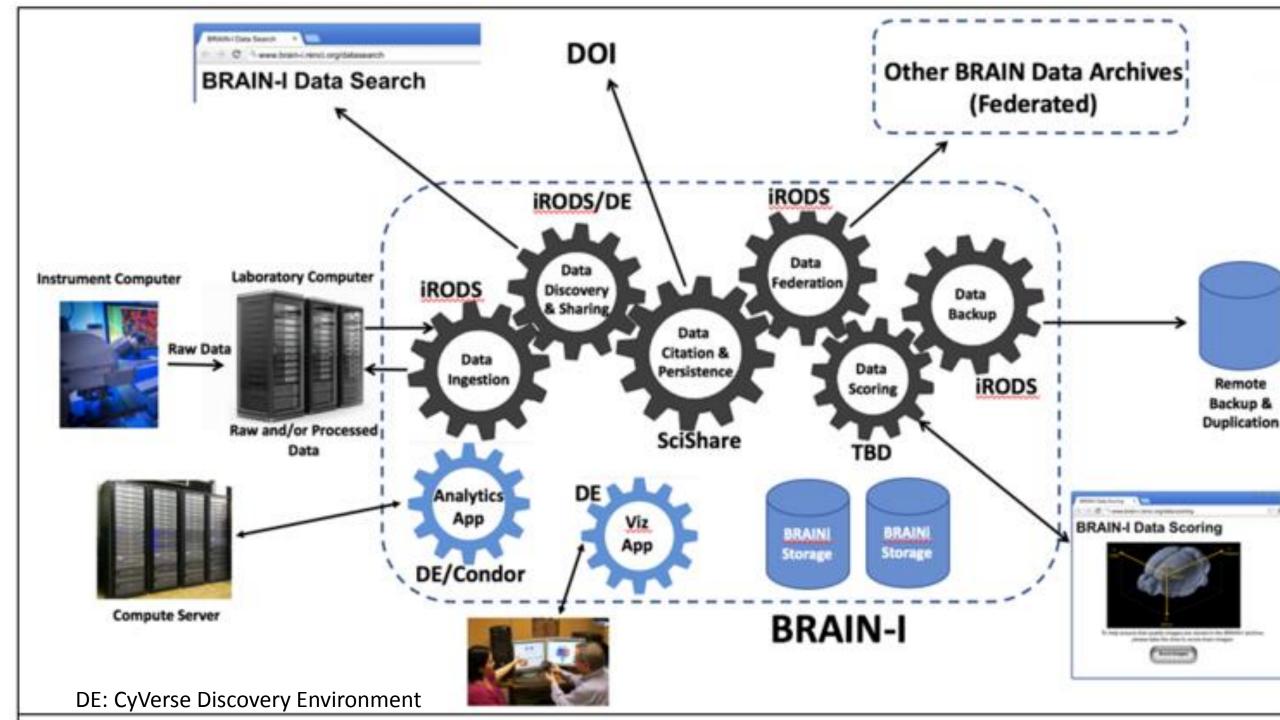


**BRAIN-I** 

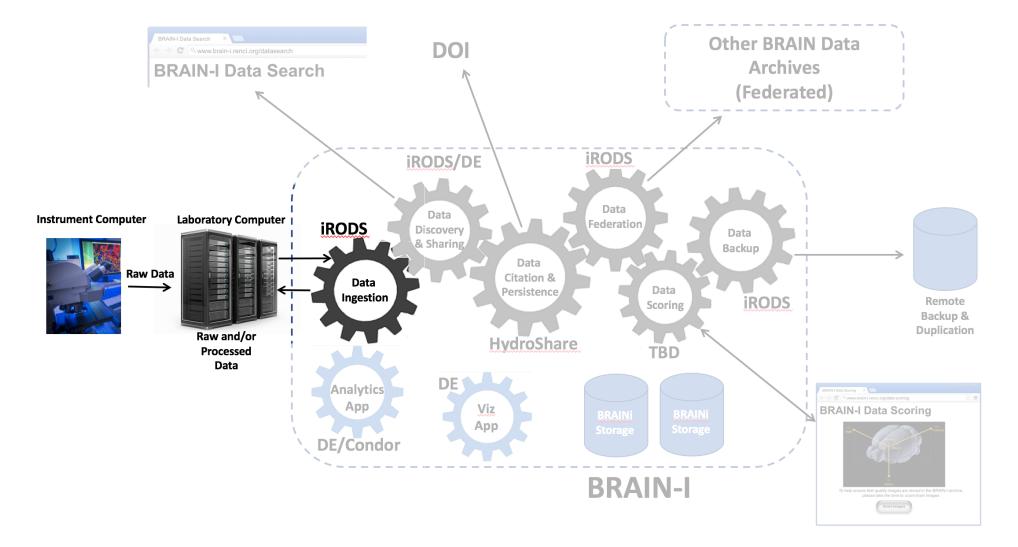
Funded by the National Science Foundation

Visualization of 3D brain images using immersive environments

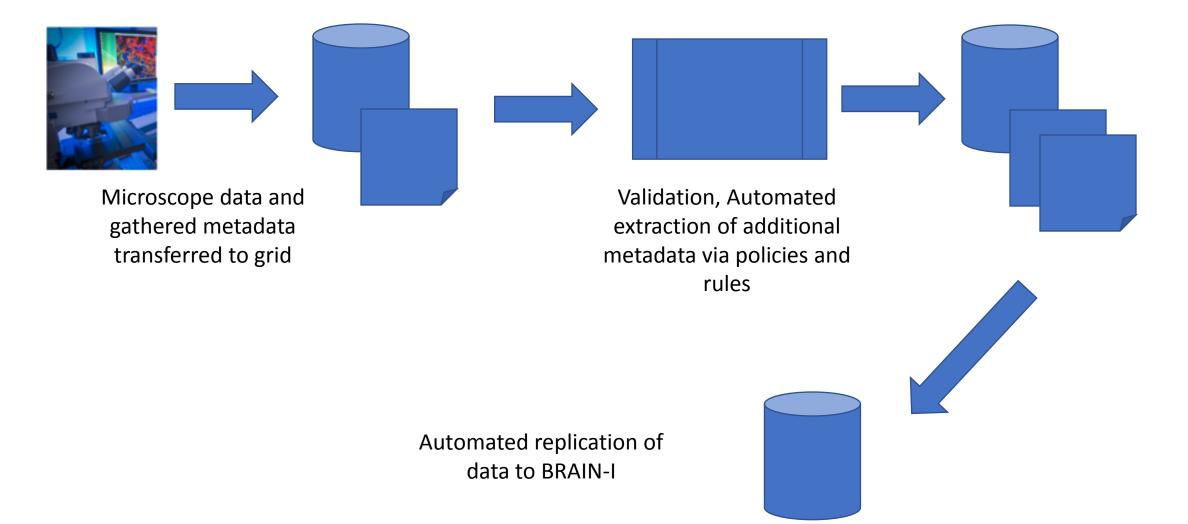




## Data Ingestion



### **Data Accession Sequence**



## Data Ingestion – Standards and Identifiers

× – D iDrop File M	Manager							
		5 <b>1</b>		ols Settings Exper	<b>K</b>			
				ots settings Exper	imenc			
	Browse /dsai_renci	/home/testde2/BRAI	N-I					
Local Directory ExperimentID: A15-0417-L2				 Mast modified				
vmlinuz	Lab/PI:	440		Mon Nov 14 16:54:42 ES	T 2016			
▶ sys	, Experimental Purpose	"Demo2			Thu Mar 16 14:07:10 EDT 2017 Tue Mar 28 16:31:52 EDT 2017			
▶ root				Tue Mar 28 16:31:52 EDT				
▶ srv ▶ dev		0	<u>Cancel</u>	QK Fri Dec 02 12:55:56 EST 2	2016			
		5005.VMak	1.8 GB	Tue Apr 11 16:40:13 EDT	2017			
▶ USF	examples.desktop		8.8 KB	Fri Dec 02 12:46:41 EST 2	2016			
▶ lib64 ▶ bin	FileTemplate.class		2.4 KB	Fri Dec 02 15:30:53 EST 2	2016	16		
▶ docker	hello.txt		19 B	Mon Mar 20 14:11:05 ED	T 2017			
▶ boot	hi		6 B	Tue Apr 11 16:37:10 EDT	2017	=		
► cdrom initrd.ima	hs_err_pid10758.log		68.0 KB	Fri Dec 02 13:14:31 EST 2	2016			
r sbin	hs_err_pid6878.log		72.3 KB	Fri Dec 02 12:47:05 EST 2	2016			
▶ home	hs_err_pid7536.log		68.1 KB	Fri Dec 02 13:06:41 EST 2	2016			
▶ lost+found ▶ lib	hs_err_pid8827.log		DiDrop - Tools Di	alog	6			
	indri-5.9.tar.gz		<ul> <li>Experiment Tools</li> </ul>		117			
▶ media	raw_news_for_ipo_test.html		Create an experiment		17			
initrd.img.old	▼BRAIN-I		Set up a preparation Add image and channels					
▶ Var	►A15-0417-L1		goo image and chainin	213	017			
					-			
Current File: Total Progress:					_	use		
Host:irc	ods.data2discovery.org	Zone:dsai_renci U			nsfers <u>See deta</u>	ils		

Data Capture on Instrument

- Desktop 'agent' that can manage accession of instrument data to the lab data grid
- Provision metadata for experiments via templates
- Interrogation of instrument for additional metadata

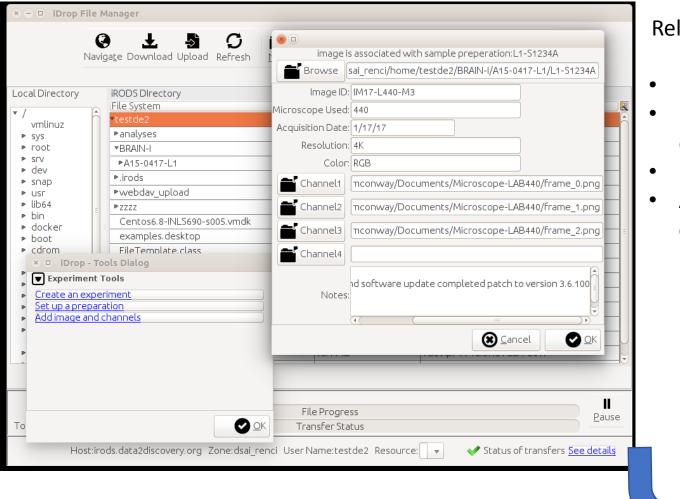
## Data Ingestion – Standards and Identifiers

		× D				
_ocal Directory	iRODS Directory	Sam	nple is a	associated with experin	nent: <mark>A15-0417-L1</mark>	
• /	File System	Browse	/dsai	_renci/home/testde2/B	RAIN-I/A15-0417-L1	
vmlinuz	testde2	Sample	Type	WT -		
▶ sys	▶ analyses					
▶ root ▶ srv	▼BRAIN-I		· .	L1-S1234A		
▶ dev	►A15-0417-L1	Clearing Pro	otocol:	ACP7		
▶ snap	▶.irods	Sacrifice	Date:	1/12/17	]	
▶ usr ▶ lib64	▶webdav_upload	Clearing	Date:	1/12/17		
▶ bin		Primary Antibody			ĺ	
docker	Centos6.8-INLS690-s005.vmdk			<u> </u>	J	
▶ boot ▶ cdrom	examples.desktop FileTemplate.class	Secondary Antibody			J	
× □ iDrop - 1		Date in Cle	- 1	<u></u>		
Experiment Tools		Primary Antibody		PAB1		
		Primary Anubody	osed:			
<u>Create an experiment</u> Set up a preparation				SAB2		
Add image and channels		Secondary Antibody				
Þ	-					
	-	Geno	otype:	G52-MC		
-	-	E	3reed:	XB887223		
		Caging S	Status:			
_	-		Sex:	Male 💌		
		File Age (w				
То		Tran Age (w		PR1		
				NX33		
Host:i	rods.data2discovery.org Zone:dsai_renc	i User Na In-vivo Treat	:ment:			
		Tire - Daiah		Ω		
		Time Point:		J		
		Notes:				

Data Capture on Instrument

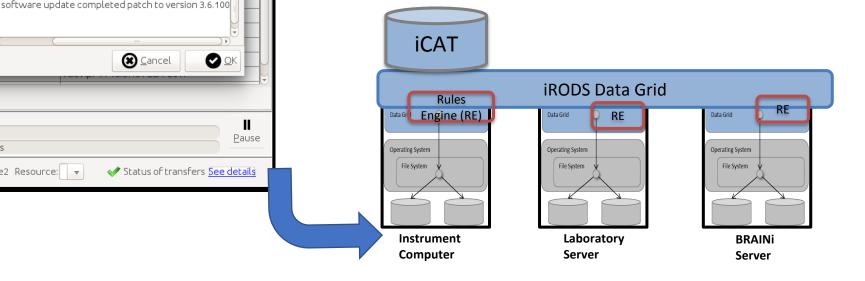
- Adding a prepared test specimen to the experiment
- Common metadata is populated automatically from the template

## Data Ingestion – Standards and Identifiers

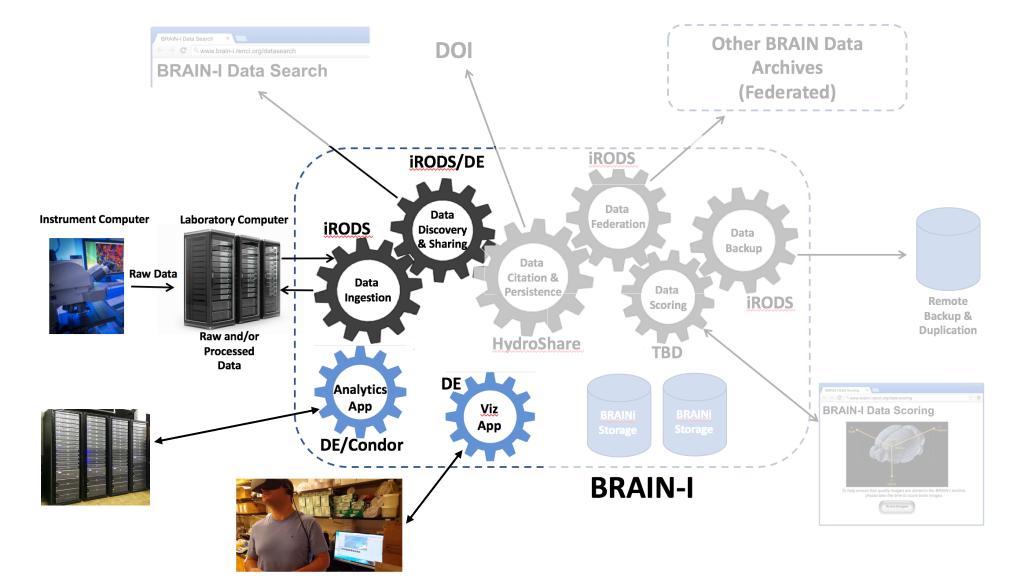


Reliable (hands off) accessioning of curated instrument data

- Image channels identified and linked to sample
- Reliable, auditable accessioning of large files to lab data grid
- Error tracking, reliability
- Ability to schedule multiple accession actions to run overnight

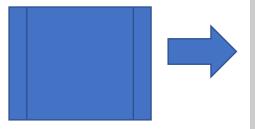


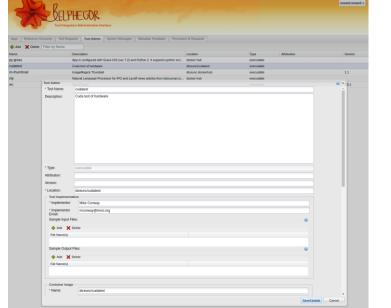
## **Analysis and Visualization Tools**



## Analysis and Visualization Tools

Package any app or algorithm as a Docker image





Data Category Workspace >> Apps Public Apps (1) Name Beta (1) 🕕 🐺 Cudatest testde2 testd (0) - 👳 Workspace (2) 🔊 🕾 thumbnaile testde2 test Apps under ( Favorite Apps (0 My public apps (0) Shared with me (1) Sthumbnaile Analysis Name:thumbnailer analysis Section 1 Folder containing png and jpg images Select a folde Browse Width 250 \* Launch Analysis

Discovery Environment

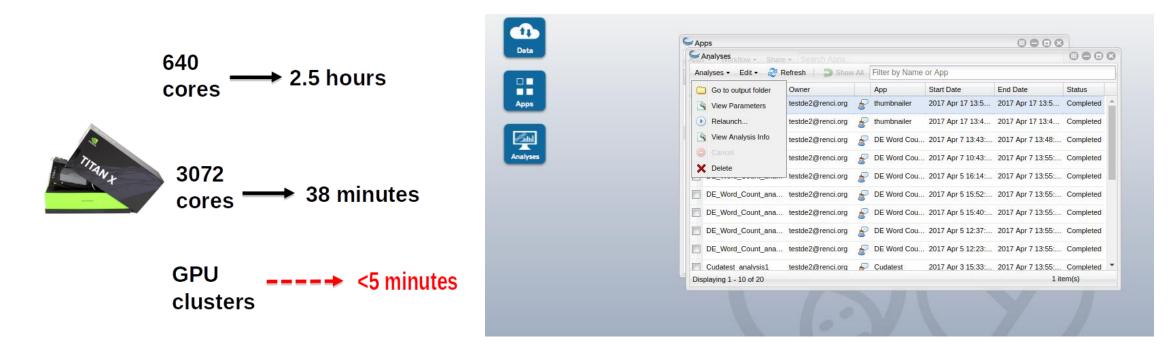
Have an administrator add the app as a 'Tool'

Users can create a GUI to launch the tool, and share these GUI Apps with others



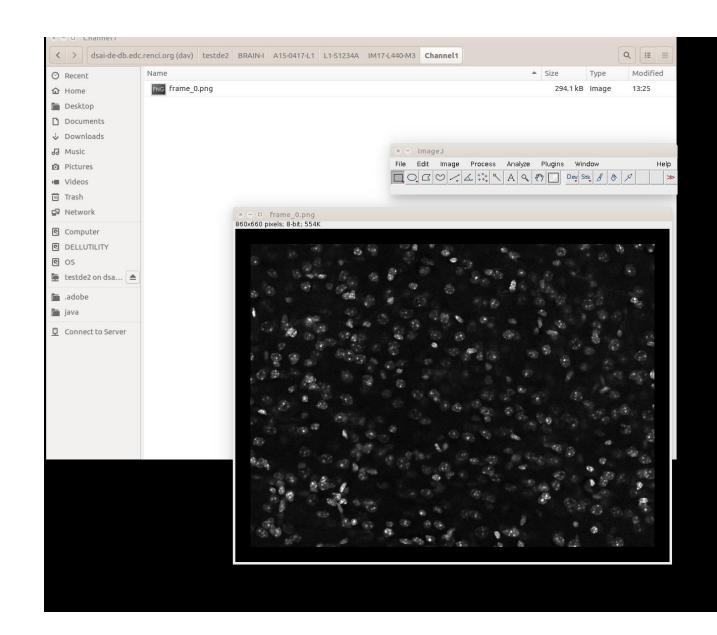
Data replicated to GPU compute resource

Dockerized analysis routed to GPU machine automatically Analysis products, provenance metadata, parameters appear in the grid when complete

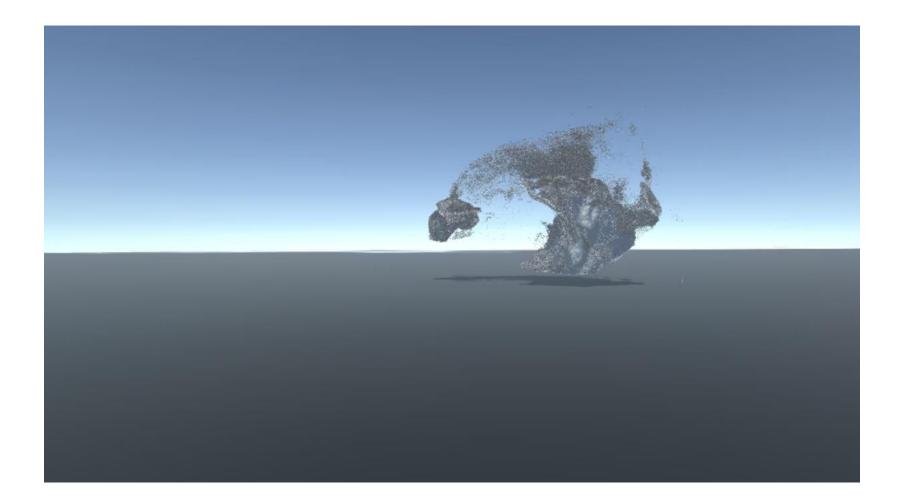


#### Easy desktop/web access for researchers

- Data grid integrates with desktops and common domain tools.
- Here we are viewing BRAIN-I data on a desktop using off-the-shelf image tools such as ImageJ
- Plan to add access via Jupyter notebooks very soon

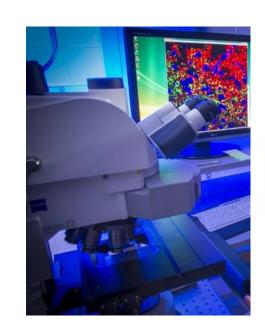


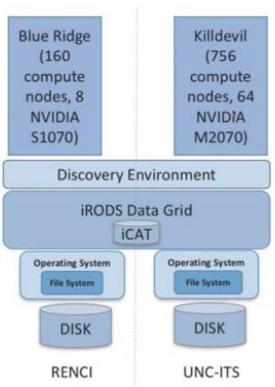
## Using Oculus for 3-D Visualization

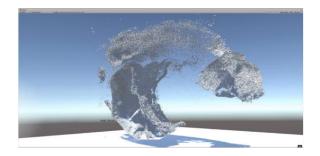


## iRODS helps BRAIN-I gets cyberinfrastructure out of the way of science

- Easy, reliable data management and tracking from microscope to publication
- Intuitive environment for computation and data sharing
- Policy based data management, secure and auditable









## Surgical Critical Care Initiative (SC2i)

## SC2i: Surgical Critical Care Initiative Precision Medicine for Acute Care



- Goal of SC2i: To create clinical decision support tools that focus on best choices for each patient based on data collected from studies at civilian and military research hospitals.
- Partners:
  - Uniformed Services University of the Health Sciences
  - Walter Reed National Military Medical Center
  - Naval Medical Research Center
  - Duke University School of Medicine
    - RENCI is a sub-contractor to Duke
  - Emory University School of Medicine
  - Decision Q
  - Henry M Jackson Foundation for the Advancement of Military Medicine

## Central Data Repository (CDR) in SC2i

- Data from all institutions is saved in a Central Data Repository for analysis and visualization.
- RENCI is primarily responsible for architecting, implementing and maintaining the CDR
- The CDR is a secure system in AWS GovCloud

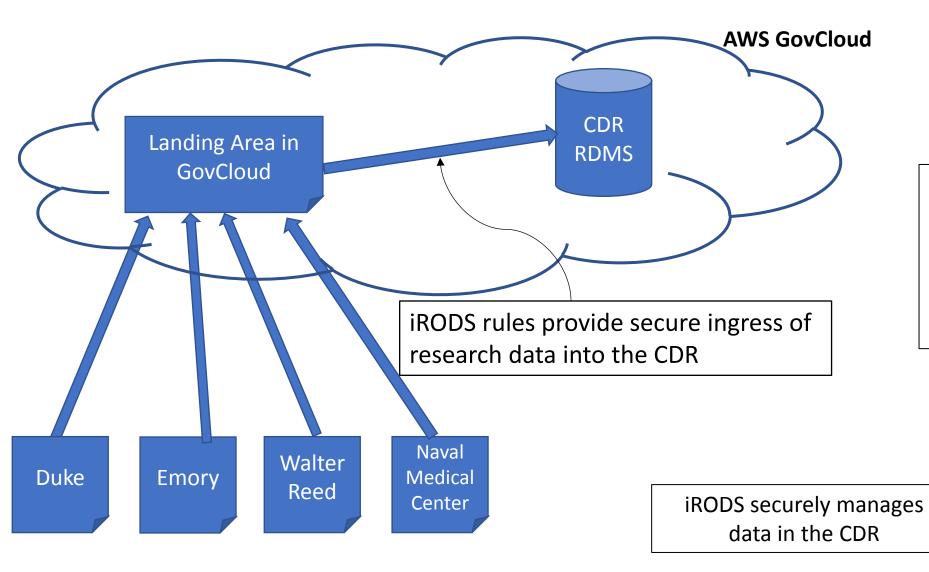


- GovCloud is a FedRAMP compliant region within Amazon Web Services (AWS)
- Provides secure/compliant infrastructure for government customers
- CDR runs on GovCloud infrastructure



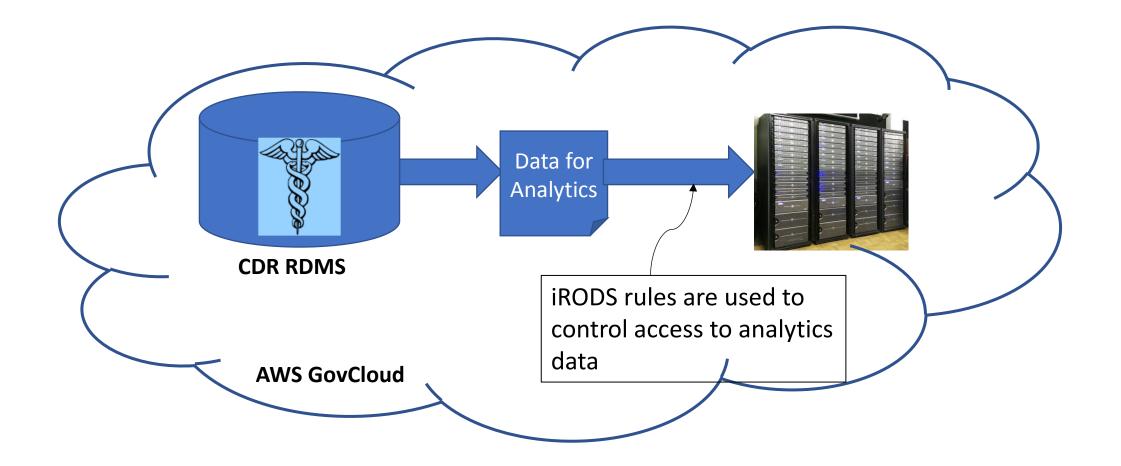
<u>FedRAMP</u> is government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services

## Data Upload and Ingest Using iRODS



iRODS's configurable access control, customizable rules and policies, and secure user management features fulfill security and privacy requirements

## Data ETL for Analytics using iRODS



#### Contact

## Ashok Krishnamurthy Deputy Director RENCI

ashok@renci.org