Data Integrity

Research Roundtable
February 2024
Florida Atlantic University
Research Integrity Office

Data Integrity vs Data Management

Data Integrity:

The accuracy, reliability, and consistency of data over its entire life-cycle.

Ensures the accuracy, completeness, consistency, and validity of data.

Data Management:

The practice of collecting, organizing, and accessing data to support productivity, efficiency, and decision-making.

Research Data

FAU DOR Policy 10.1.6

- Recorded factual information commonly accepted in the scientific community as necessary to reconstruct, evaluate, and validate research findings and results, regardless of the media on which it may be recorded.
- Examples of Research Data include, but are not limited to, laboratory notebooks, notes of any type including printouts, specimens of any type including organisms, photographs, reagents, digital images, protocols, numbers, graphs, charts, numerical raw experimental results, instrumental outputs from which Research Data can be derived and other deliverables under sponsored agreements.

FLORIDA ATLANTIC UNIVERSITY
DIVISION OF RESEARCH

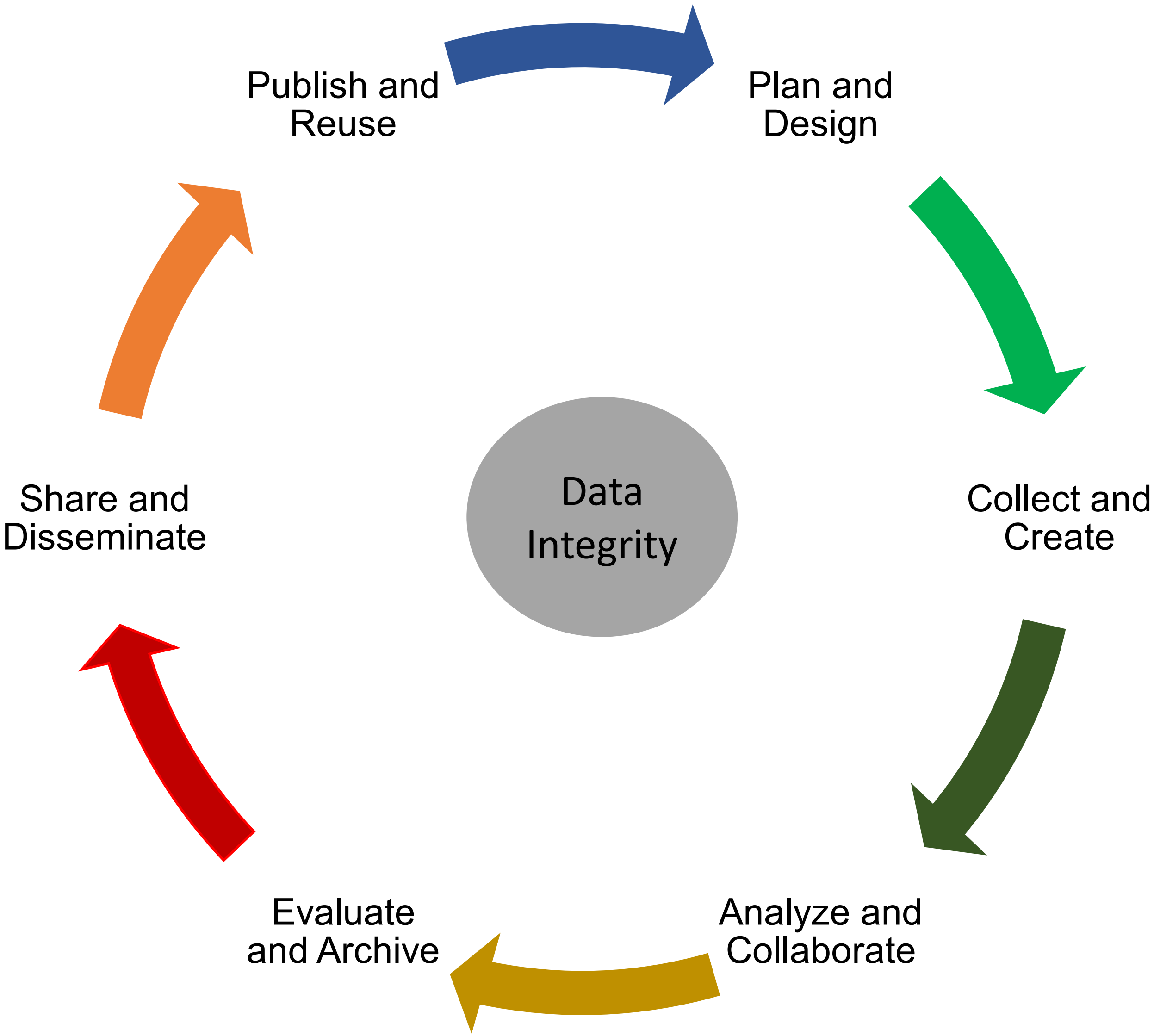
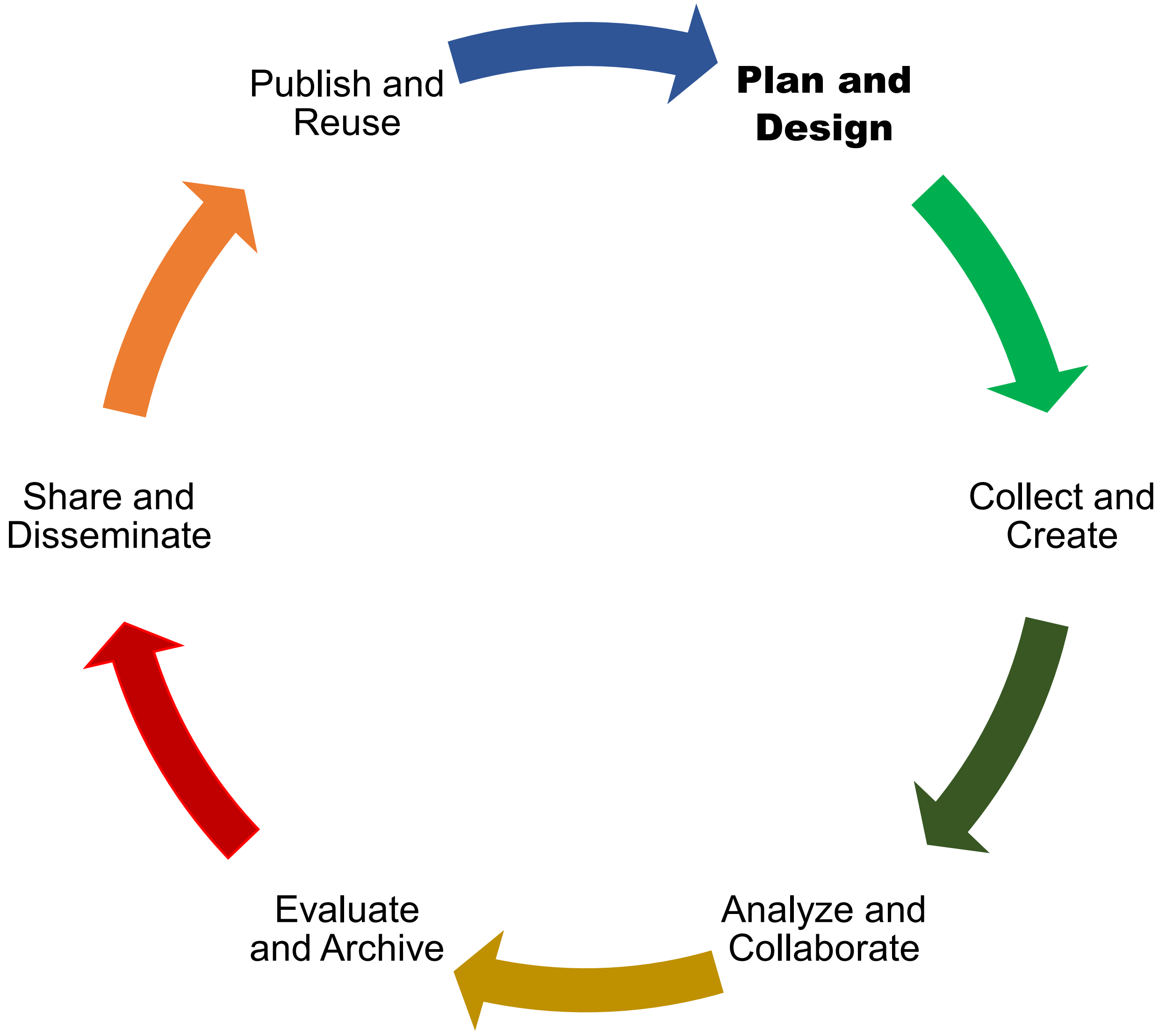


Image adapted from https://datamanagement.hms.harvard.edu/plan-design/biomedical-data-lifecycle

Plan and Design

- Data Management Plans
- Data Policies and Compliance
- Directory Structures
- Roles and Responsibilities
- File Naming Conventions

https://dmptool.org/



Plan and Design

Collect and Create

Analyze and Collaborate

Evaluate and Archive

Share and Disseminate

Publish and Reuse

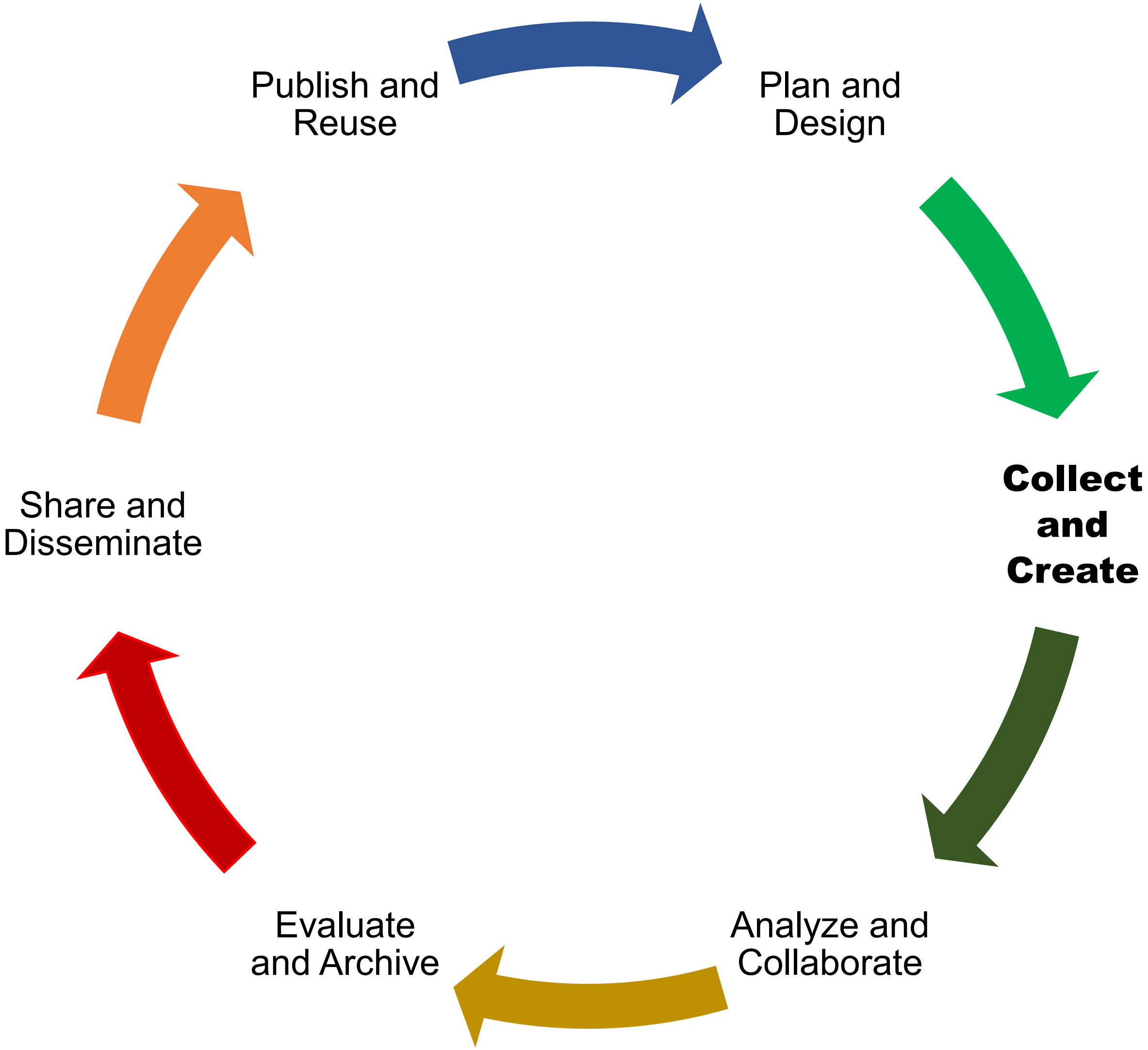
Collect and Create

FAU DOR Policy 10.1.6:

Collecting project data in a consistent and systematic manner.

Requirements for the recording and storage of Research Data and material will vary by discipline.

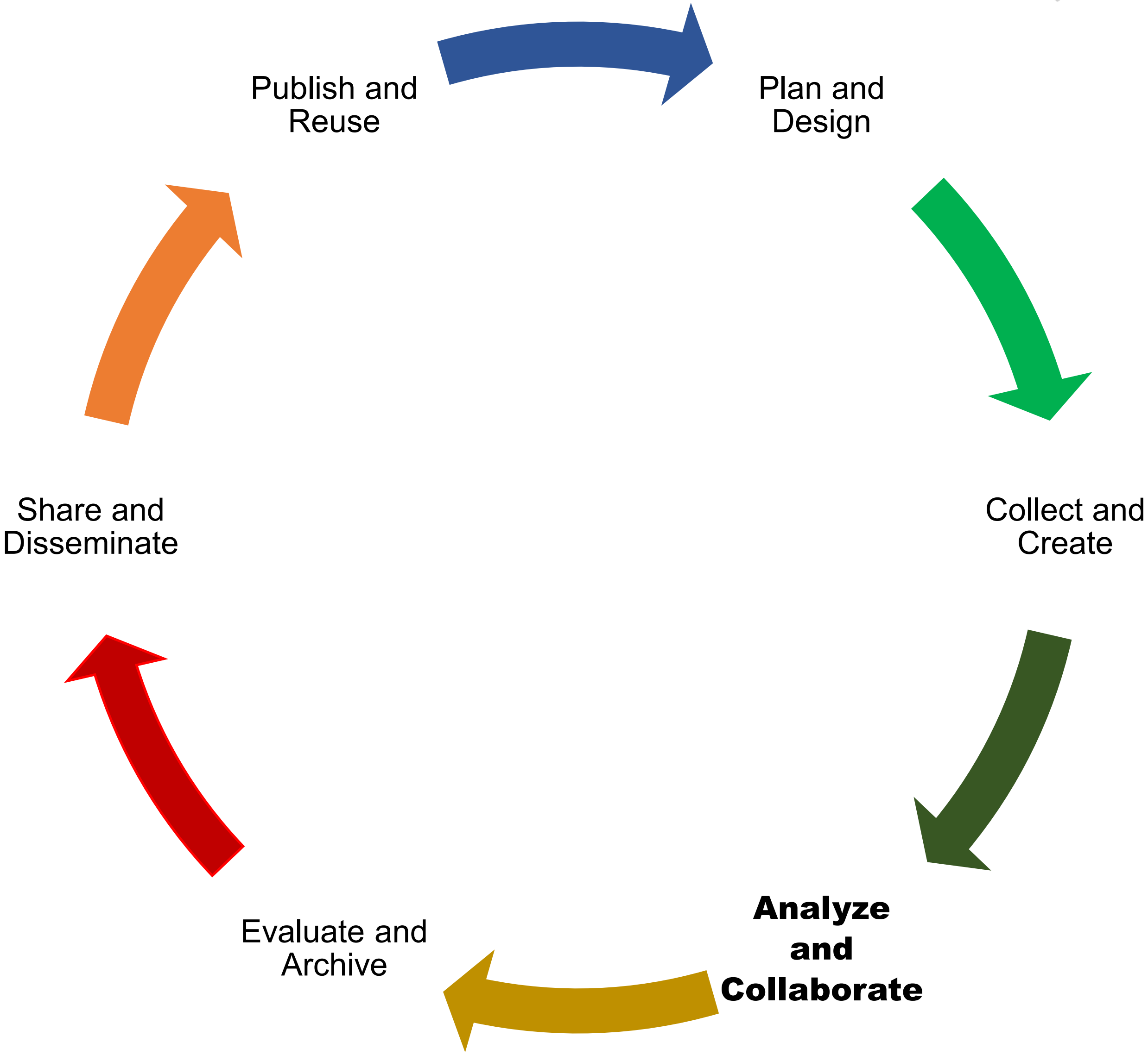
PIs should always adhere to requirements of funding agencies, standards of the applicable industry, professional guidance where available, any principles set out on the College level as well as FAU recommendation as outlined in compliance documents.



FLORIDA ATLANTIC UNIVERSITY
DIVISION OF RESEARCH

Analyze and Collaborate

- Inspecting, cleaning, transforming, and modeling data with the goal of discovering useful information, informing conclusions and supporting decision-making.
- Collaborative tools and software
- Documentation and Metadata
- Reproducibility
- Analysis ready datasets
- Image management
- Version control



Plan and Design

Collect and Create

Analyze and Collaborate

Evaluate and Archive

Share and Disseminate

Publish and Reuse

Evaluate and Archive

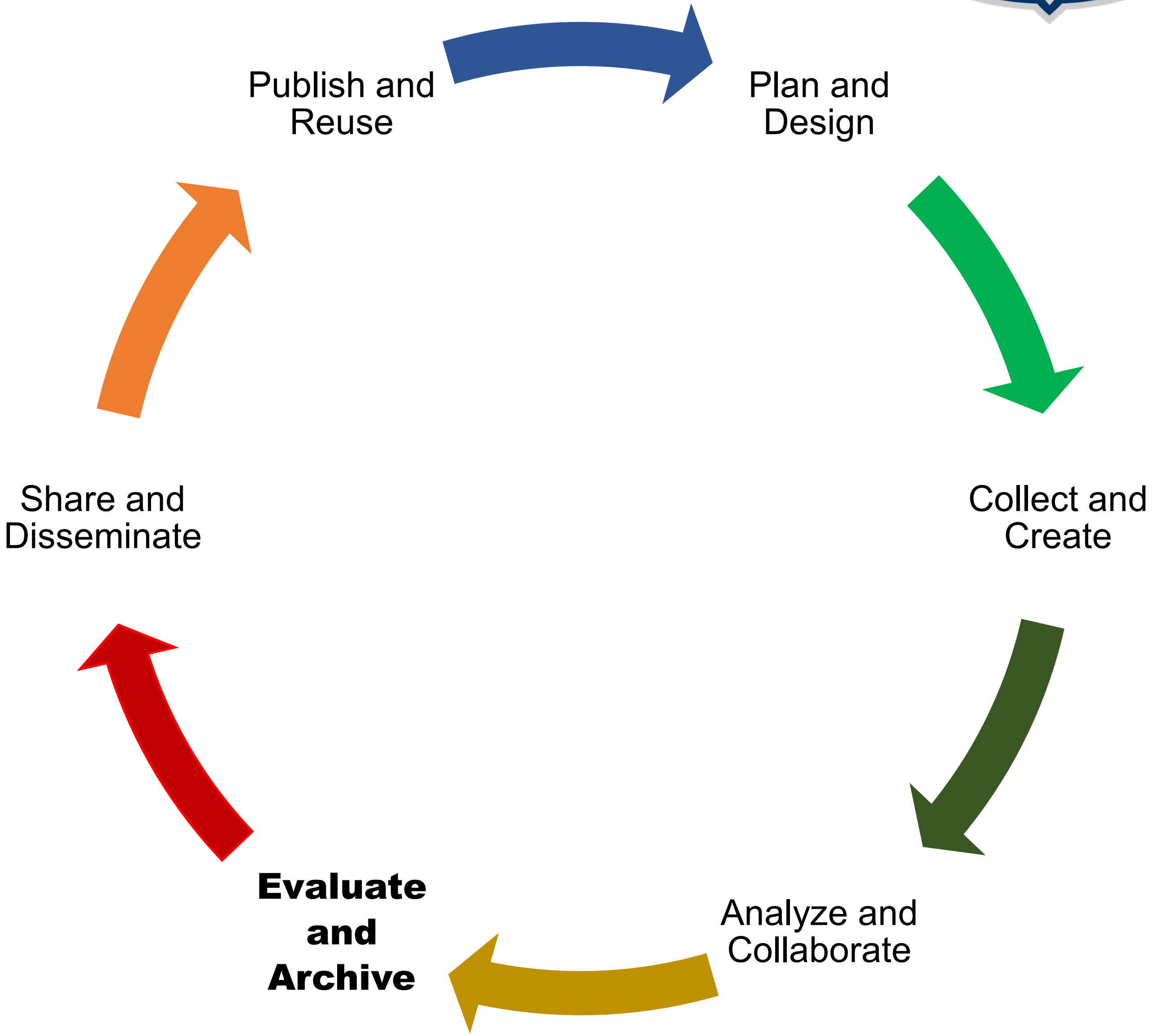
- Security,
- Retention,
- Destruction,
- Archive/ records management

FAU DOR Policy 10.1.6:

Research Data should be stored using a method that permits a complete retrospective audit if necessary.

The PI will have access to the Research Data generated by the project. Any other faculty, staff, student or person involved in the creation of Research Data may have the ability to review that portion of the Research Data that they created.

Report any Research Data integrity breaches to the appropriate FAU oversight entity.



Plan and Design

Collect and Create

Analyze and Collaborate

Evaluate and Archive

Share and Disseminate

Publish and Reuse

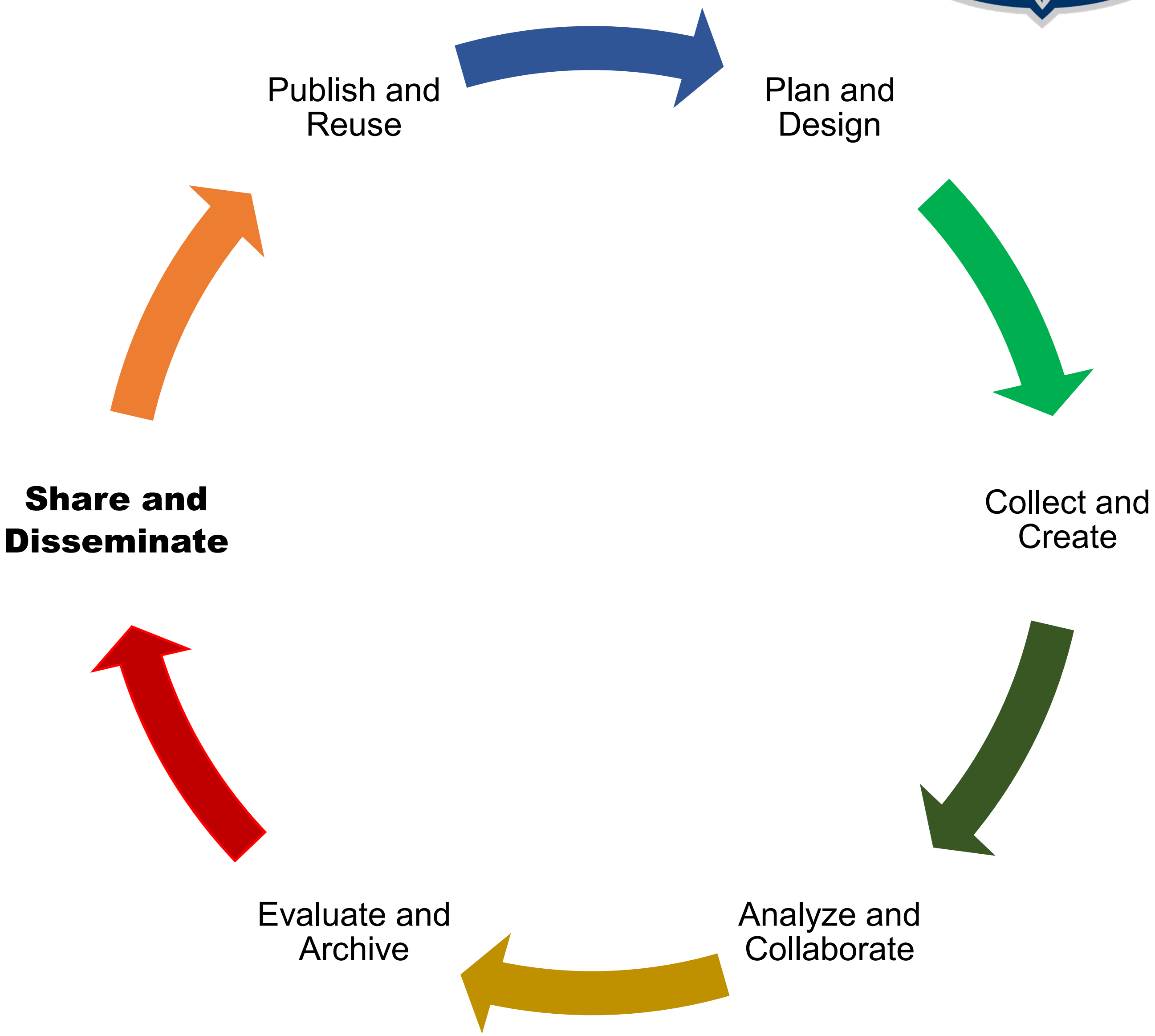
FLORIDA ATLANTIC UNIVERSITY
DIVISION OF RESEARCH

Share and Disseminate

- Data sharing
- Open access
- DUAs
- Intellectual property

FAU DOR Policy 10.1.6:

Details on how, when, to and by whom data will be shared with other researchers and for generalizable knowledge should be detailed in research plans, including but not limited to protocols, data sharing plans, consent documents, and data use agreements.



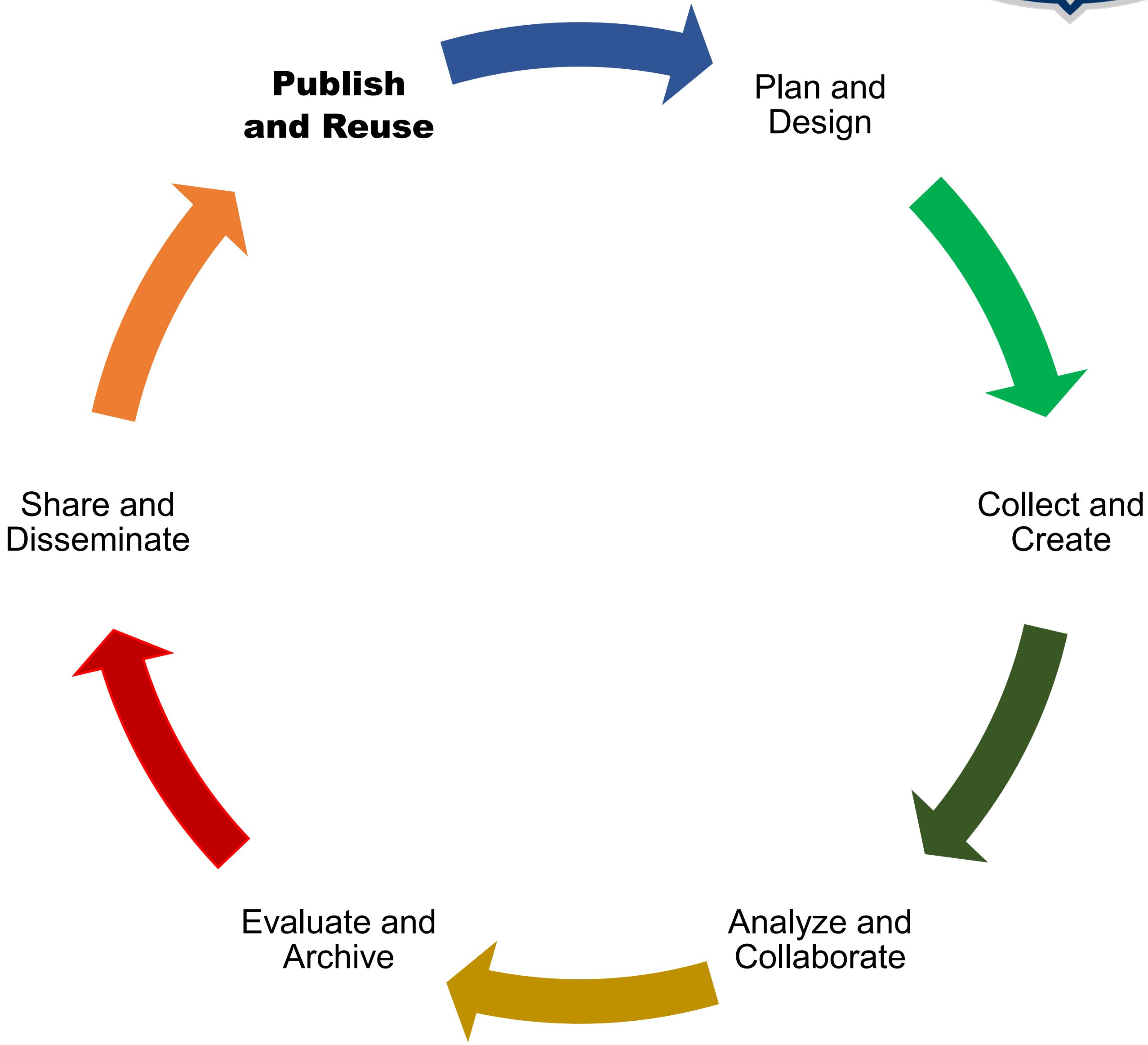
Publish and Reuse

- Scholarly products
- Preprints and publishing
- Data repositories

Data accessibility is the degree to which other researchers, and you yourself can use data.

Data isn't just available, but also usable. Make your data accessible by ensuring it:

- Is in a reliable storage location
- Is retrievable online using standardized protocols
- Has restrictions in place as necessary



Resources

NIH: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

NSF: https://new.nsf.gov/funding/data-management-plan#nsfs-data-sharing-policy-1c8

NOAA: https://nosc.noaa.gov/EDMC/PD.DSP.php

FAU DOR Policies: https://www.fau.edu/research/policies-and-procedures/

FLORIDA ATLANTIC UNIVERSITY
DIVISION OF RESEARCH

Questions and Discussion

FAU Office of Information Technology
https://www.fau.edu/oit/

FAU Human Research Protection Program
researchintegrity@fau.edu

https://www.fau.edu/research-admin/research-integrity/human-subjects-irb/

FAU Responsible Conduct of Research

https://www.fau.edu/research-admin/research-integrity/responsible-conduct-of-research/

FAU Research Computing

RESEARCH COMPUTING AT FLORIDA ATLANTIC UNIVERSITY

OFFICE OF INFORMATION TECHNOLOGY

Services

General

- 10 Gbps Network Uplink to Florida Lambda Rail
- 100 Gbps Core
- Various Intel CPU's
- AMD – EPYC (6,500 cores)
- NVIDIA A100 - ~18
- NVIDIA V100 - 48
- SLURM SCHEDULING
- Open OnDemand
- SSH Console Access

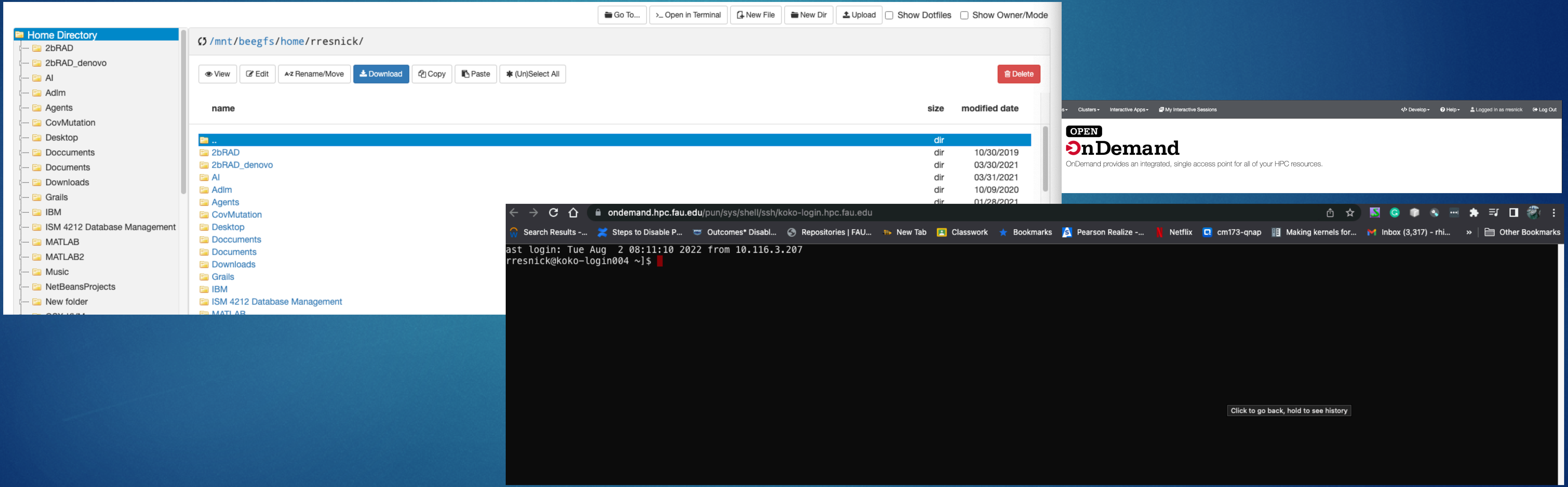
HPC Cluster KOKO

Team

- James Mauser
  Systems Administrator
- Skyler Paulus
  Systems Administrator
- Bahareh Yaseen Saadatmand-Mashhadi
  Research Facilitator, BioInformatics
- Chris Johnson
  Form Development
- Rhian Resnick
  Director

Open OnDemand

- Provides a web UI to the clusters.



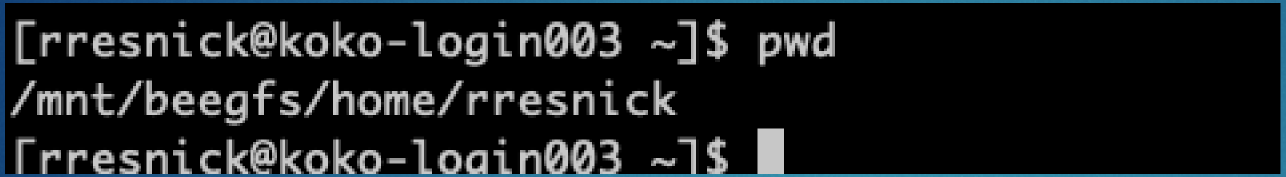
name	size	modified date
..	dir	
2bRAD	dir	10/30/2019
2bRAD_denovo	dir	03/30/2021
AI	dir	03/31/2021
Adlm	dir	10/09/2020
Agents	dir	01/28/2021
CovMutation		
Desktop		
Doccuments		
Documents		
Downloads		
Grails		
IBM		
ISM 4212 Database Management		

OnDemand provides an integrated, single access point for all of your HPC resources.

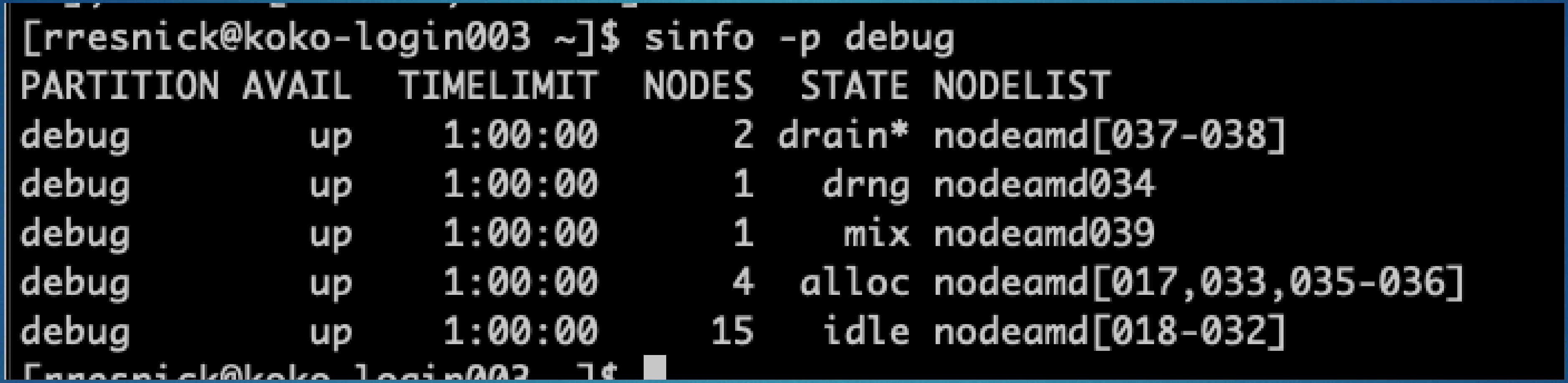
```
ast login: Tue Aug  2 08:11:10 2022 from 10.116.3.207
rresnick@koko-login004 ~]$
```

OFFICE OF INFORMATION TECHNOLOGY

SSH



```
[rresnick@koko-login003 ~]$ pwd
/mnt/beegfs/home/rresnick
[rresnick@koko-login003 ~]$
```



```
[rresnick@koko-login003 ~]$ sinfo -p debug
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
debug        up    1:00:00      2 drain* nodeamd[037-038]
debug        up    1:00:00      1   drng nodeamd034
debug        up    1:00:00      1    mix nodeamd039
debug        up    1:00:00      4  alloc nodeamd[017,033,035-036]
debug        up    1:00:00     15   idle nodeamd[018-032]
[rresnick@koko-login003 ~]$
```

Storage (PowerScale)

- Home & Scratch
  - 60TB NVME
- Archive
  - 822 TB of storage
- Users – 50 GB quota
- Scratch – 300 GB quota deleted after 90 days

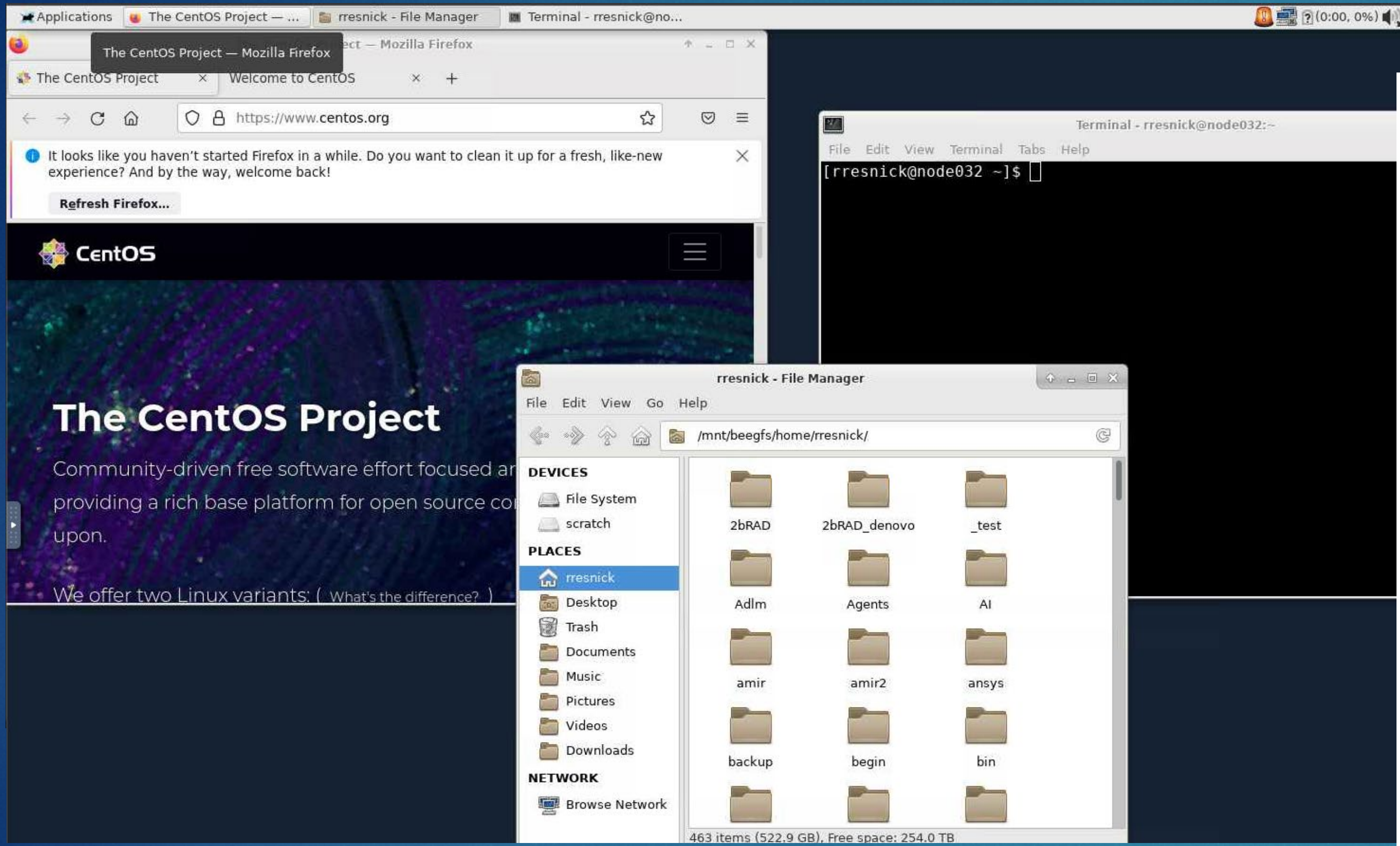
Storage (Cloud Nas)

- Archive
  - 60TB NVME
  - Basically Unlimited Cloud Storage
- Archive
  - No Backups: $13.33 / TB / Month
  - With Backups: $31.50 / TB / Month
  - With Backups and Replication: $63 / TB / Month
- Most users will receive 300 GB of complimentary storage on archive as long as they are connected to a new or funded research activity. (as determined by the Research Computing Team, we are very liberal with this definition, it just protects our storage system from abuse)

Software

- Machine Learning
- Big Data
- Intel Compilers
- MPI
- OpenMP
- HPC, HTC, Graphics
- Install to your home directory
- Install to your group directory
- Install to entire cluster for sharing with others (upon request)

Remote Desktop via OnDemand



It looks like you haven't started Firefox in a while. Do you want to clean it up for a fresh, like-new experience? And by the way, welcome back!

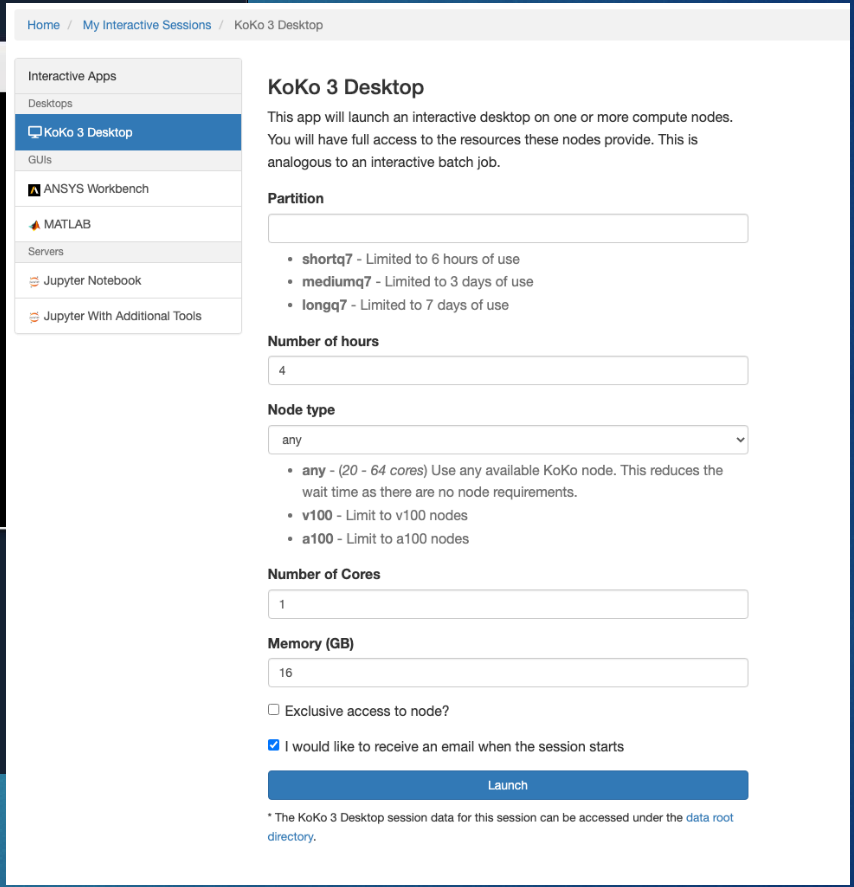
Community-driven free software effort focused ar providing a rich base platform for open source co upon.

We offer two Linux variants: (What's the difference?)

```
[rresnick@node032 ~]$ 
```

/mnt/beegfs/home/rresnick/

463 items (522.9 GB), Free space: 254.0 TB



KoKo 3 Desktop

This app will launch an interactive desktop on one or more compute nodes. You will have full access to the resources these nodes provide. This is analogous to an interactive batch job.

Partition

- **shortq7** - Limited to 6 hours of use
- **mediumq7** - Limited to 3 days of use
- **longq7** - Limited to 7 days of use

Number of hours: 4

Node type: any

- **any** - (*20 - 64 cores*) Use any available KoKo node. This reduces the wait time as there are no node requirements.
- **v100** - Limit to v100 nodes
- **a100** - Limit to a100 nodes

Number of Cores: 1

Memory (GB): 16

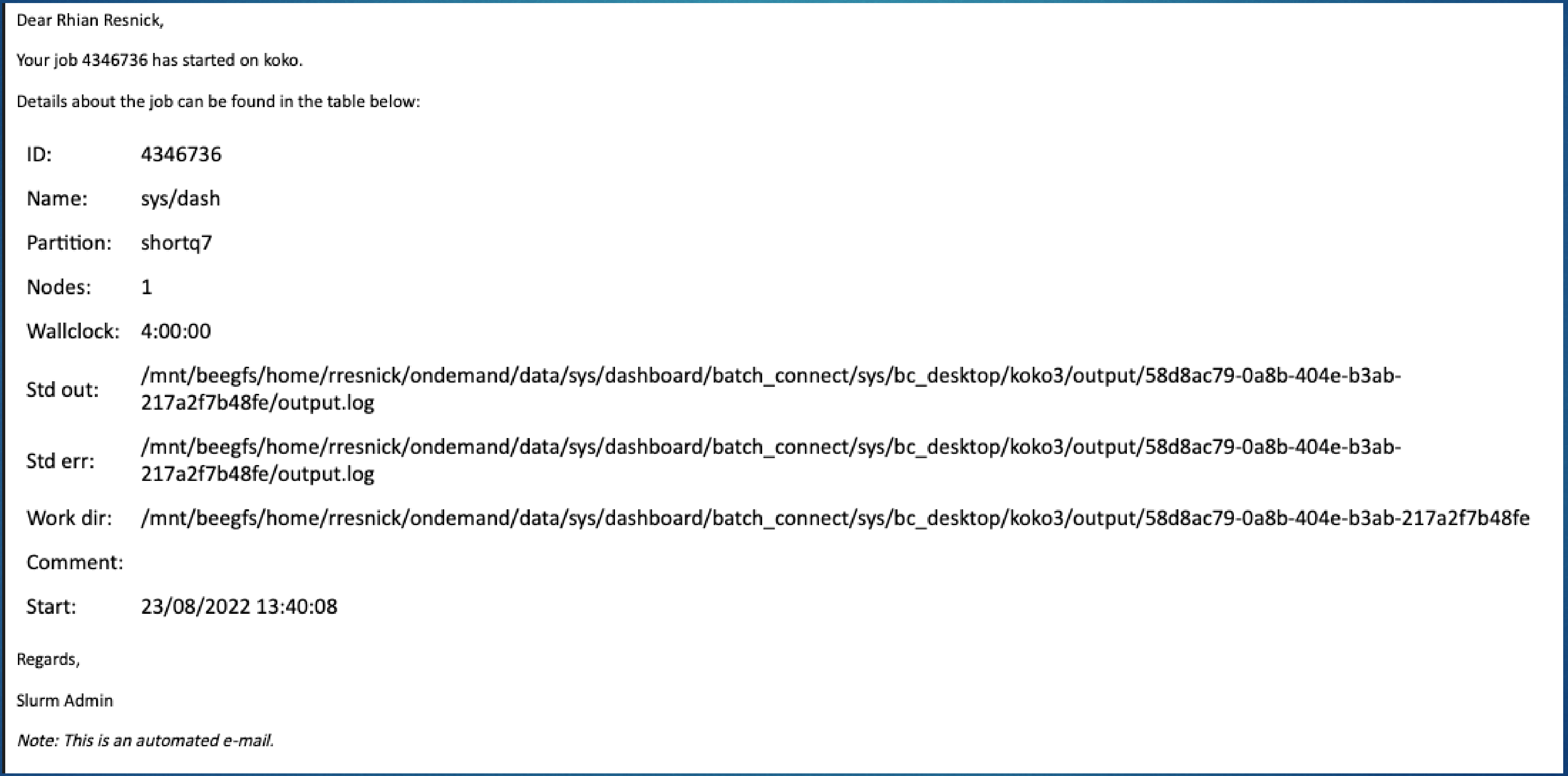
☐ Exclusive access to node?

☑ I would like to receive an email when the session starts

* The KoKo 3 Desktop session data for this session can be accessed under the data root directory.

OFFICE OF INFORMATION TECHNOLOGY

Email Notifications



Dear Rhian Resnick,

Your job 4346736 has started on koko.

Details about the job can be found in the table below:

ID:	4346736
Name:	sys/dash
Partition:	shortq7
Nodes:	1
Wallclock:	4:00:00
Std out:	/mnt/beegfs/home/rresnick/ondemand/data/sys/dashboard/batch_connect/sys/bc_desktop/koko3/output/58d8ac79-0a8b-404e-b3ab-217a2f7b48fe/output.log
Std err:	/mnt/beegfs/home/rresnick/ondemand/data/sys/dashboard/batch_connect/sys/bc_desktop/koko3/output/58d8ac79-0a8b-404e-b3ab-217a2f7b48fe/output.log
Work dir:	/mnt/beegfs/home/rresnick/ondemand/data/sys/dashboard/batch_connect/sys/bc_desktop/koko3/output/58d8ac79-0a8b-404e-b3ab-217a2f7b48fe
Comment:	
Start:	23/08/2022 13:40:08

Regards,

Slurm Admin

Note: This is an automated e-mail.

Costs

- Everything we offer has a free introductory tier to help spring board research.
- But some things have a cost regardless.

1. Secure Virtual Deskop at the cost of $168.50 per core per year (Includes 4GB of memory per core)
2. Storage over 300 GB $160 / TB / Year
3. Prioritized Compute $116 / CPU Core / Year
4. Prioritized GPU $1,024 / GPU Per / Year (this gets interesting when comparing models, so reach out to us. Our model is based on UF's GPU unit model, more information on request)
5. Consultation: $60 per hour
6. Data Center Hosting: $16 per U minimum of 4U

Links and Disclaimers

- Cyber Infrastructure Plan 2017-2022: https://helpdesk.fau.edu/TDClient/2061/Portal/KB/ArticleDet?ID=141481
- Services: https://helpdesk.fau.edu/TDClient/2061/Portal/KB/ArticleDet?ID=142149
  All costs were FAU internal costs, fees are ~20% higher for external entities.
  We are working on new grant funded research pricing. This is still a work in progress.
- Docs: https://hpc.fau.edu