

User Group Meeting

April 2026



OKINAWA INSTITUTE OF SCIENCE AND TECHNOLOGY GRADUATE UNIVERSITY

沖縄科学技術大学院大学



IT'S THE

~~-All-~~ **Storage**
~~-All-~~ **The Time**

Edition!



THE Cluster Name
NEW Nominations





An Apology

We had 3 storage outages this spring:

- Installing the new Bucket storage
- Configuration changes to Bucket
- Networking and configuration changes to Deigo, Saion, Flash and Work

This was at least one too many





Connection Protocols

NFS
Standard
Mature
Old Bucket



- Worked fine with old Bucket
- It's already there - Need no changes to cluster nodes

Lustre
High Perf
Resilient
Flash, Work



- Need to update Flash, Work
- Needs driver updates on *all* compute and other nodes





Connection Protocols



But – New storage is in Data Center 2

- No Infiniband, only ethernet
- Longer latency, less bandwidth
- 👉 some reads (large directories) would sometimes time out



- 👉
 - Lustre doesn't behave this way
 - Take the downtime, do the updates
- 👉 Now it works as expected + better performance, new features



Question:

Fine-grained access control is not one of the new things ...

Bucket and Flash can *potentially* do lots of new things:

- Per-user quotas and/or total file numbers
- "subprojects" with quotas and limits
- Email notifications
- Better, faster usage data
- ???

What would you like to have?

`ask-scda@oist.jp`

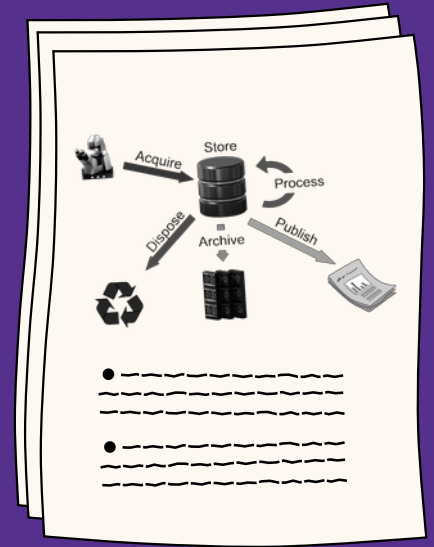


Research Data Management

Handle your data
from acquisition to disposal

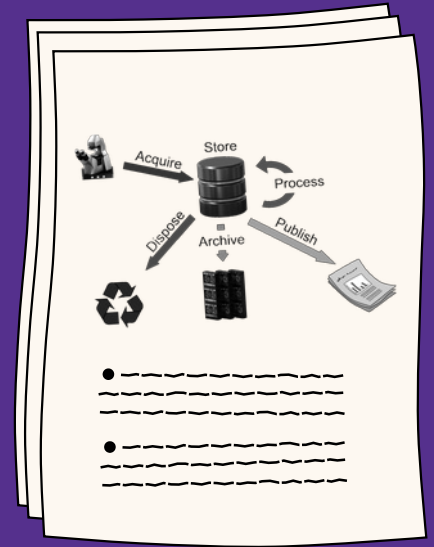
- What is it, how much is there
- Where is it from, where is it now
- What is it for, who is responsible
- How is it processed, how do you read it
- Who should have access, what is the licence
- Where is the documentation

...

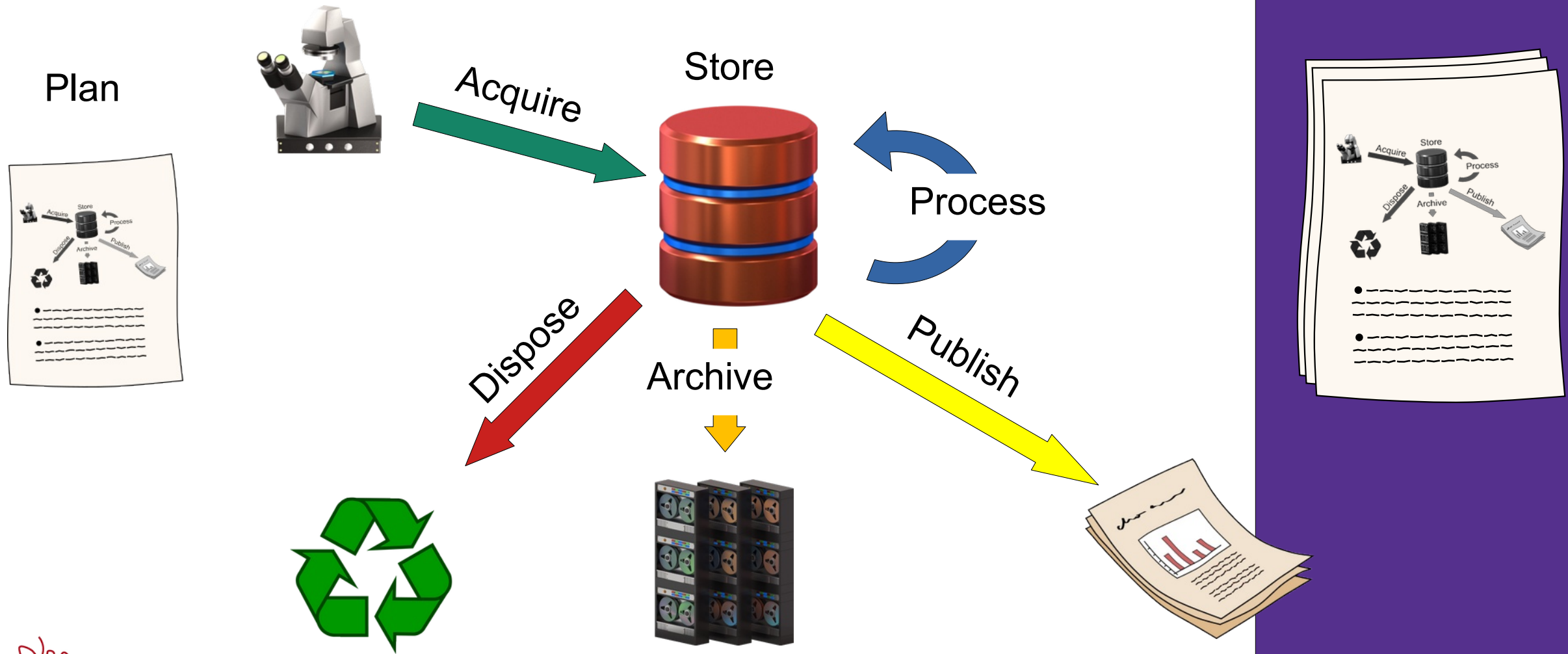


Data Lifecycle Management

- Specify Requirements
 - Storage, transfer, data security
- Data integrity and reusability
 - Data is tracked, validated, referenced, credited
- Compliance and funding
 - Requirement for grants, human and animal research, legal requirements

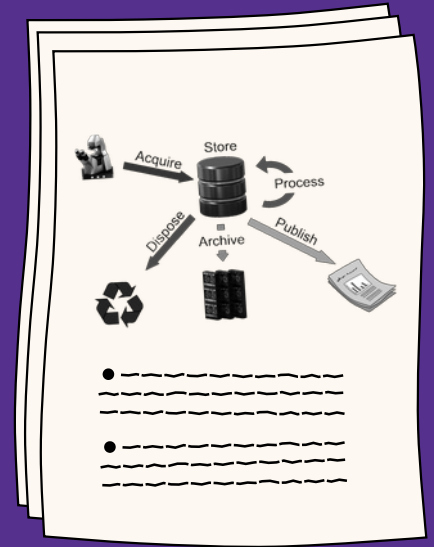


Data Lifecycle Management



Data Management Plan

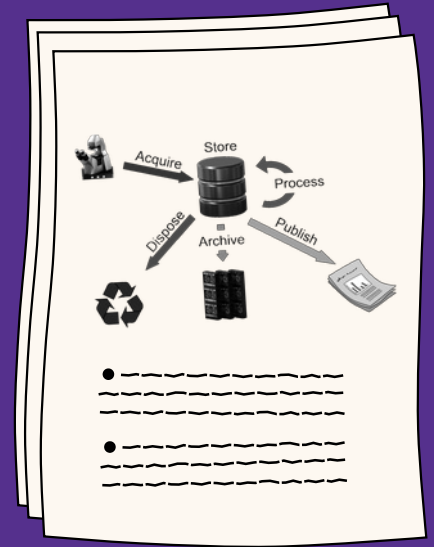
- What data, how much data?
Storage, bandwidth, cost
- For how long?
No data is stored indefinitely
- Tracking, labeling?
Data annotation, references crucial
- Where will it go?
What to keep, what to dispose of
- Who is responsible?



Manage Unit Storage

As a PI:

- Establish ground rules for the storage
 - The best time is when unit starts. The next best time is now.
 - Aim to enable data management
- Organize periodic clean-up, archiving
 - Easier to keep clean over time
- Manage access (students, etc)



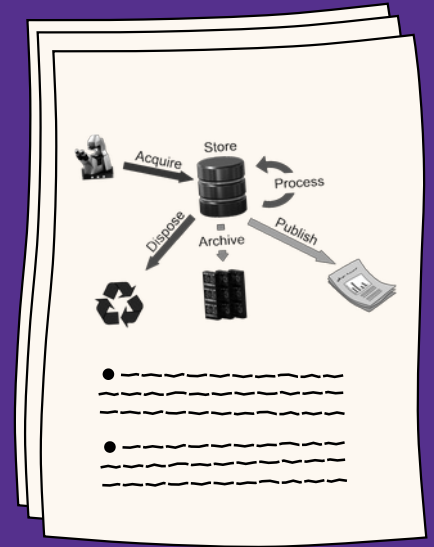
Manage Unit Storage

As a PI:

- Establish ground rules for the storage
- Organize periodic clean-up, archiving
- Manage access (students, etc)

As a Member:

- Keep your data organized
 - Document your data and processes
- Delete unneeded files
 - Build files, temp files, bad analysis, useless audio/video...



Backup Data



Why a new backup system?

The tape backup was designed for ~10PB storage

- At 7PB, backups take ~3 days
- ~15PB, we're out of time and space

Backups are meant for *disaster recovery*

And human mistakes

- Only business-critical data needs off-site backups

For disaster recovery:

- Only critical subset can be restored
- Need good backup organization

For file restoration

- Usually only most recent version
- Need good backup organization

Backup Data

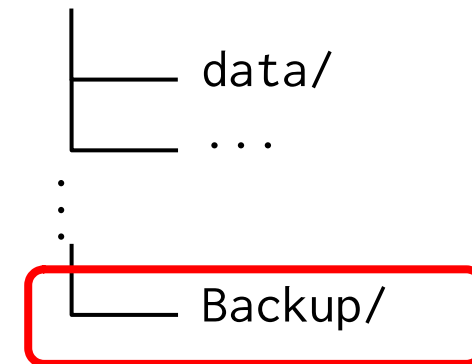


Backup Folder

- In every units top folder
- 100TB of Local Backup
- Weekly Remote Snapshots
 - 5 snapshots or 6 months
- Just* copy your data over

* terms and conditions apply. Subject to change. Not valid when invalid.

/bucket/UnitU/



Backup Data →



Backup Folder

- In every units top folder
- 100TB of Local Backup
- Weekly Remote Snapshots
 - 5 snapshots or 6 months
- Just* copy your data over

Our old system:

- Only remote snapshots (hard to get files back)
- Only 2 snapshots and 6 months

* terms and conditions apply. Subject to change. Not valid when invalid.

Backup Data →



Backup Folder

- In every units top folder
- 100TB of Local Backup
- Weekly Remote Snapshots
 - 5 snapshots or 6 months
- Just* copy your data over

As long as the data is in Backup/ there is **always** at least one snapshot!

* terms and conditions apply. Subject to change. Not valid when invalid.

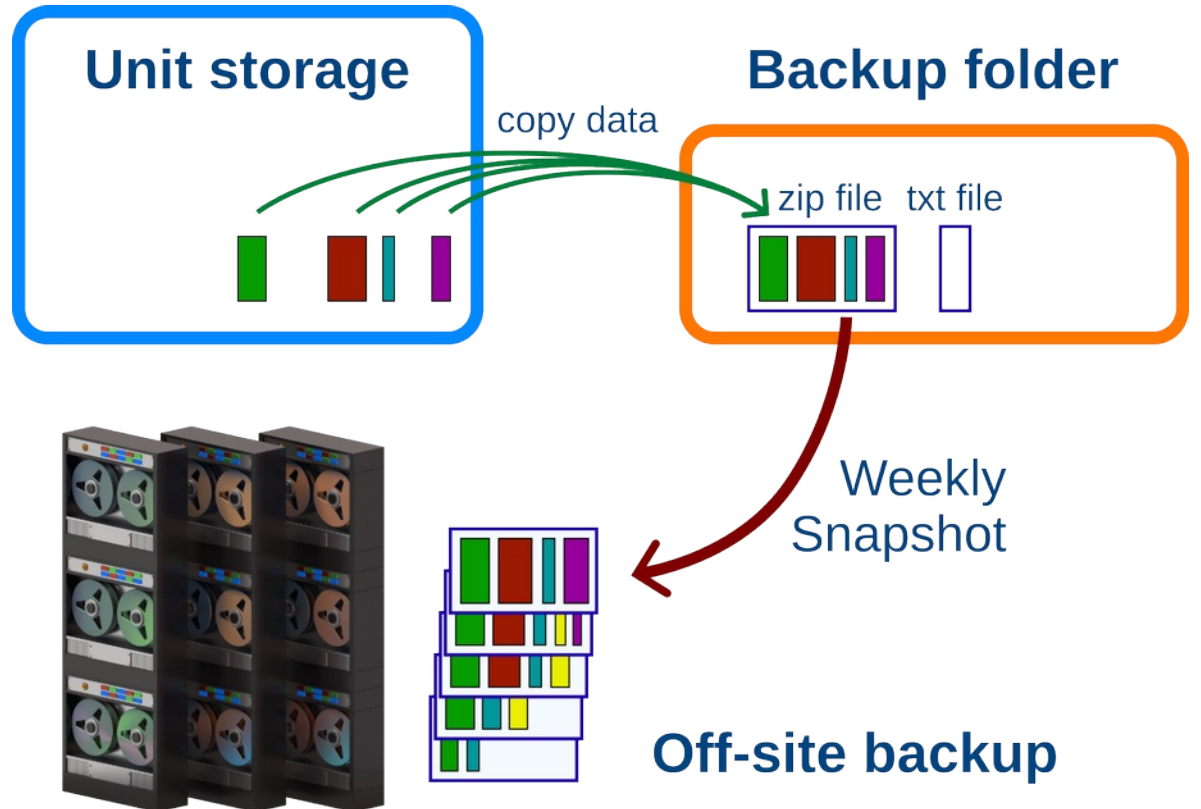
Backup Data



Backup Folder

- In every units top folder
- 100TB of Local Backup
- Weekly Remote Snapshots
 - 5 snapshots or 6 months
- Just* copy your data over

* terms and conditions apply. Subject to change. Not valid when invalid.

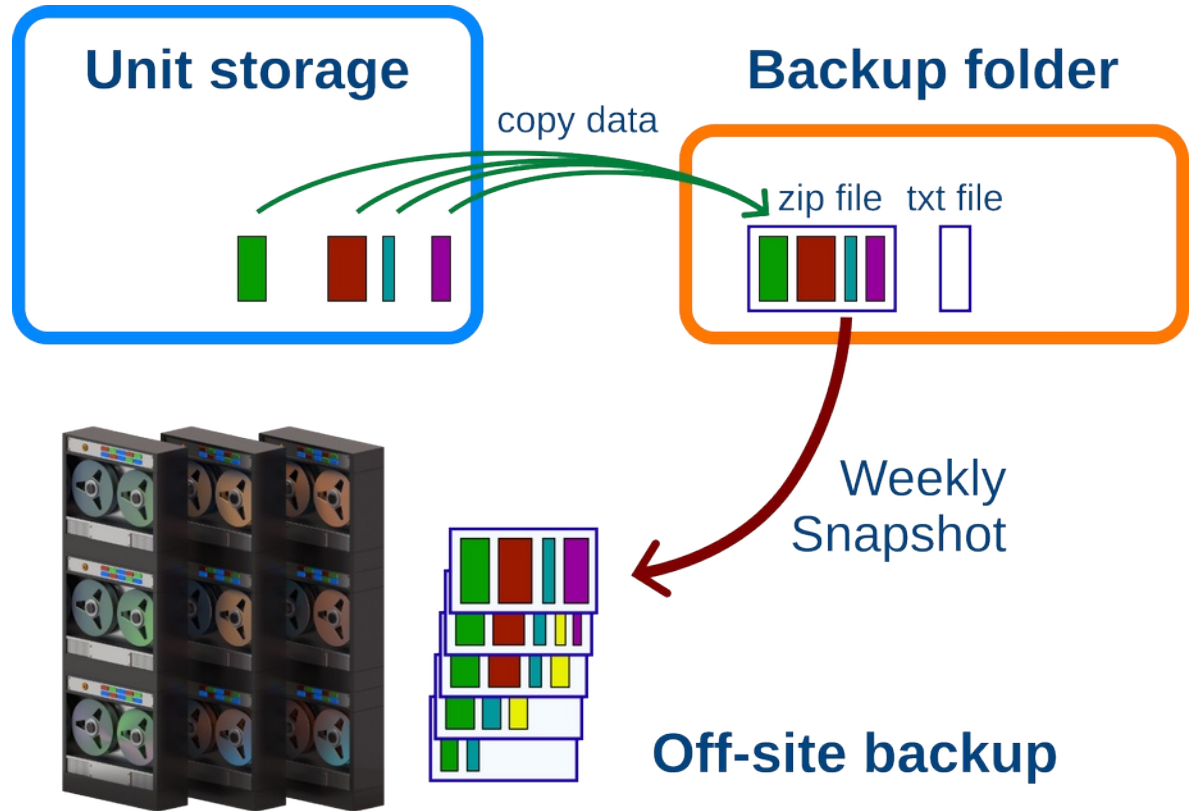


Backup Data



How to Use:

- Put your data in Backup/
- *Only* archive files:
 - zip, tar, tgz, ...
 - archive directly into Backup/
- Add a separate description file

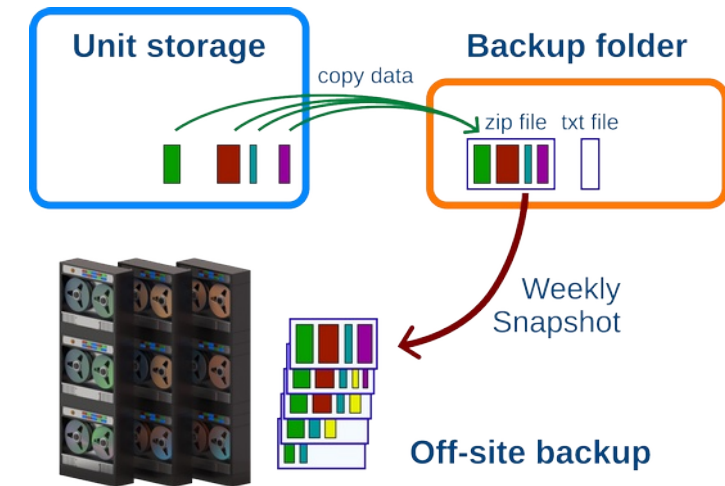


Backup Data



Create (or update) a zip archive

```
# go to your bucket storage  
$ cd /bucket/MyunitU  
  
# create the backup archive  
$ zip -r Backup/Janne_neuron.zip Janne/neuron
```

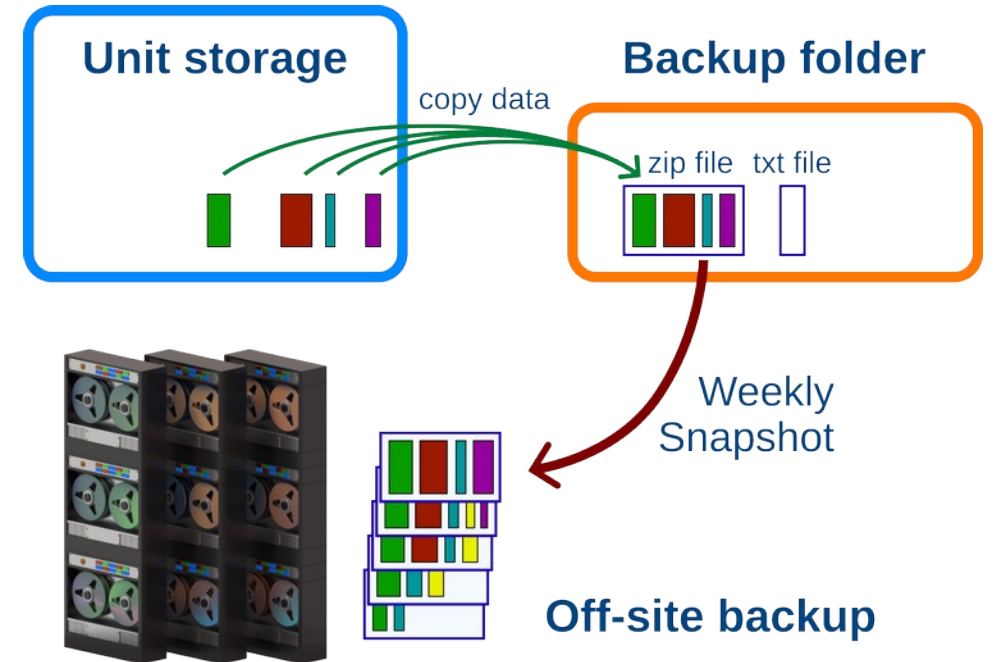


Backup Data



For a new version:

- overwrite the previous archive
- keep the same name
- You can (soon, maybe) schedule a regular backup with a Slurm job



Backup Data



Description file

- Same name as archive, ".txt" ending
- Contains:
 - **Name:** User name
 - **Project:** Project name
 - Free-form text description

The purpose is for
future you and your PI to
know what the backup contains!

Archive:

janne_neuron.**zip**

Text file:

janne_neuron.**txt**

```
Name: jan-moren  
Project: neuron spike rebound model
```

```
Model code (NEST) in git. Branch  
"publish" is version for the paper  
...
```

Backup Summary



1. Archive file: Backup/janne_neuron.zip

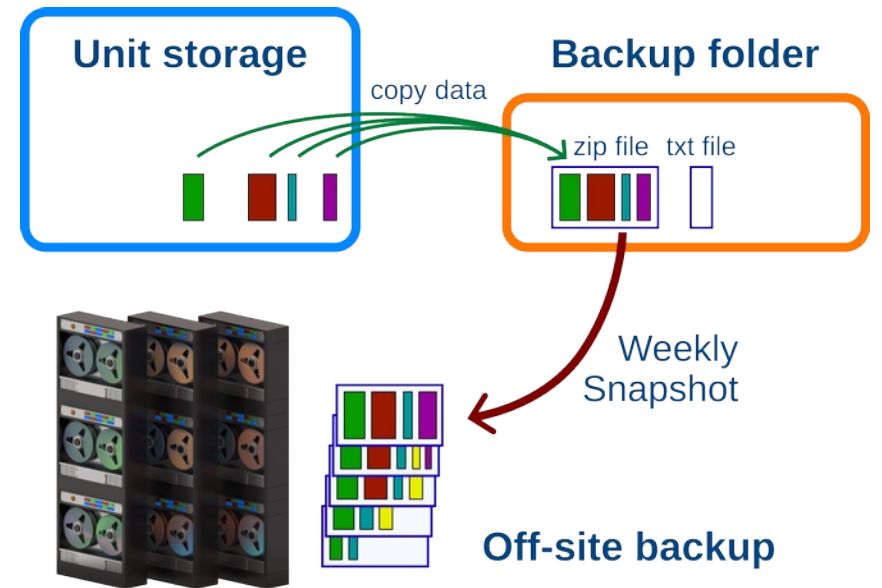
```
$ zip -r Backup/Janne_neuron.zip Janne/neuron
```

2. Description file: Backup/janne_neuron.txt

```
$ nano Backup/Janne_neuron.txt
```

3. Update the backup:

```
$ zip -r Backup/Janne_neuron.zip Janne/neuron
```

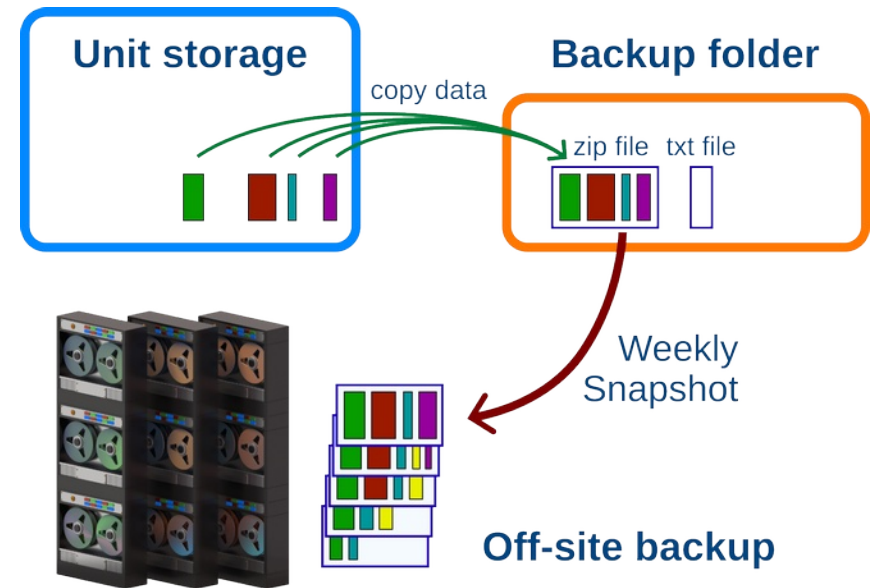


Backup Summary



1. **Archive file:** Backup/janne_neuron.zip
2. **Description:** Backup/janne_neuron.txt

Other files than these will not get a remote snapshot!
Do not put other data here.



Backup Summary



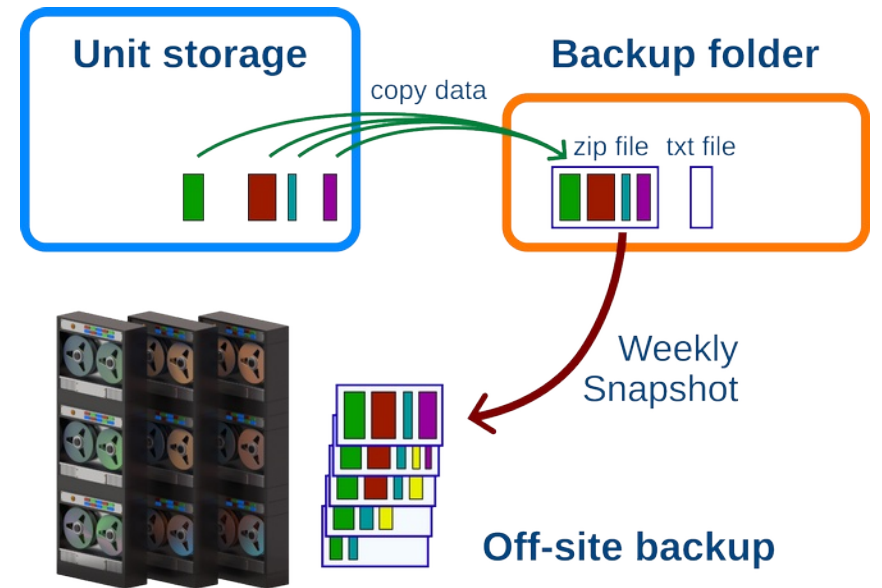
1. **Archive file:** Backup/janne_neuron.zip
2. **Description:** Backup/janne_neuron.txt

Changed archives are copied to tape weekly

Up to 5 versions are stored offsite

Up to 6 months old.

Always at least 1 copy



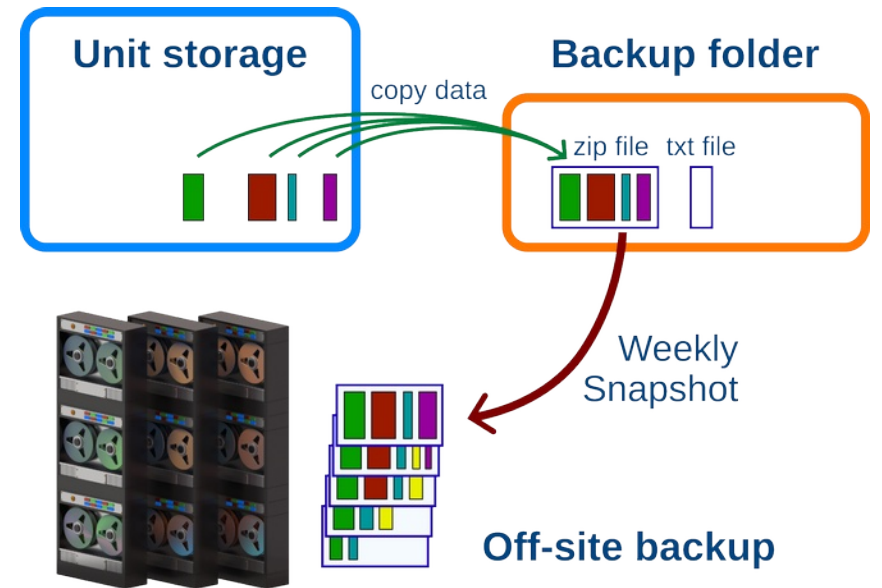
Data Recovery

You've lost a file – what to do?

Your most recent backup is in Backup/
→ get your file back directly

For older versions contact us
and tell us:

- Name and path of the file
- Name of the archive
- The date you want





Name Our New Cluster!

Adan

Grows all along the coast with its yellow fruit



Goya

The veggie we all know and love!



Kaji

Okinawan for "wind", our all-season companion



Kuina

The famous and cute Okinawa Rail



Kuwa

Mulberries, all over the OIST campus





Voting is Open Now!

Open until May 15

Don't delay!

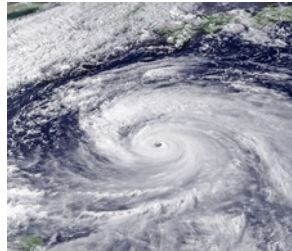
Adan



Goya



Kaji



Kuina



Kuwa



<https://forms.cloud.microsoft/r/c2Hxd8vEpb>





Software!

Geneious Prime: The vendor has discontinued floating licences – you need to buy it yourself.

Slack: Faculty Assembly voted to cancel contract. Last day is **December 11**

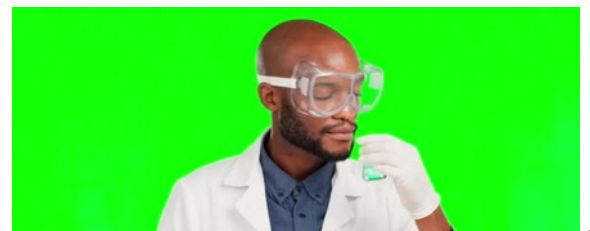
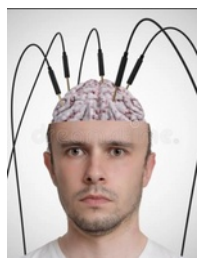
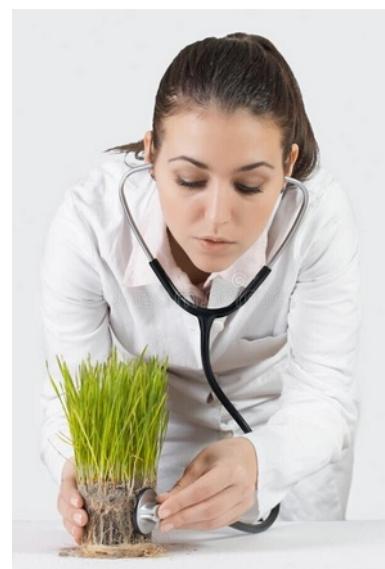
SPSS and Amira: Current version is the last one we will provide centrally.

Recently Updated Software:

Amira	- 2025
COMSOL	- 6.4
LabView	- 2026Q1
Lumerical	- 2026R1
Mathematica	- 14.3
IBM SPSS	- 31
MATLAB	- R2025b
Mnova	- v17
OriginPro	- 2026



Bioinformatics!



Rebooting the Bioinfo user group...



We created modules, hosted databases, set up RStudio server, etc., it was great but...

People did their time at OIST and now I am alone.

Let's do some bioinformatics together once a month and start a new dynamic!

We can also evolve in something broader than bioinformatics if there is appetite.



Bioinformatics!



Bioinformatics Open Hours:

Tuesdays	10-12	B648
Thursdays	15-17	L5EF02

Training

Introduction to R:
Replicating Publication-Ready Figures with ggplot2
May 29, 10-12
Room B503

Introduction to R Programming
Back again in September!

