Bioinformatics Capacity at *icipe*

Caleb Kibet

Ext: 2098

Office: LC1 18

icipe

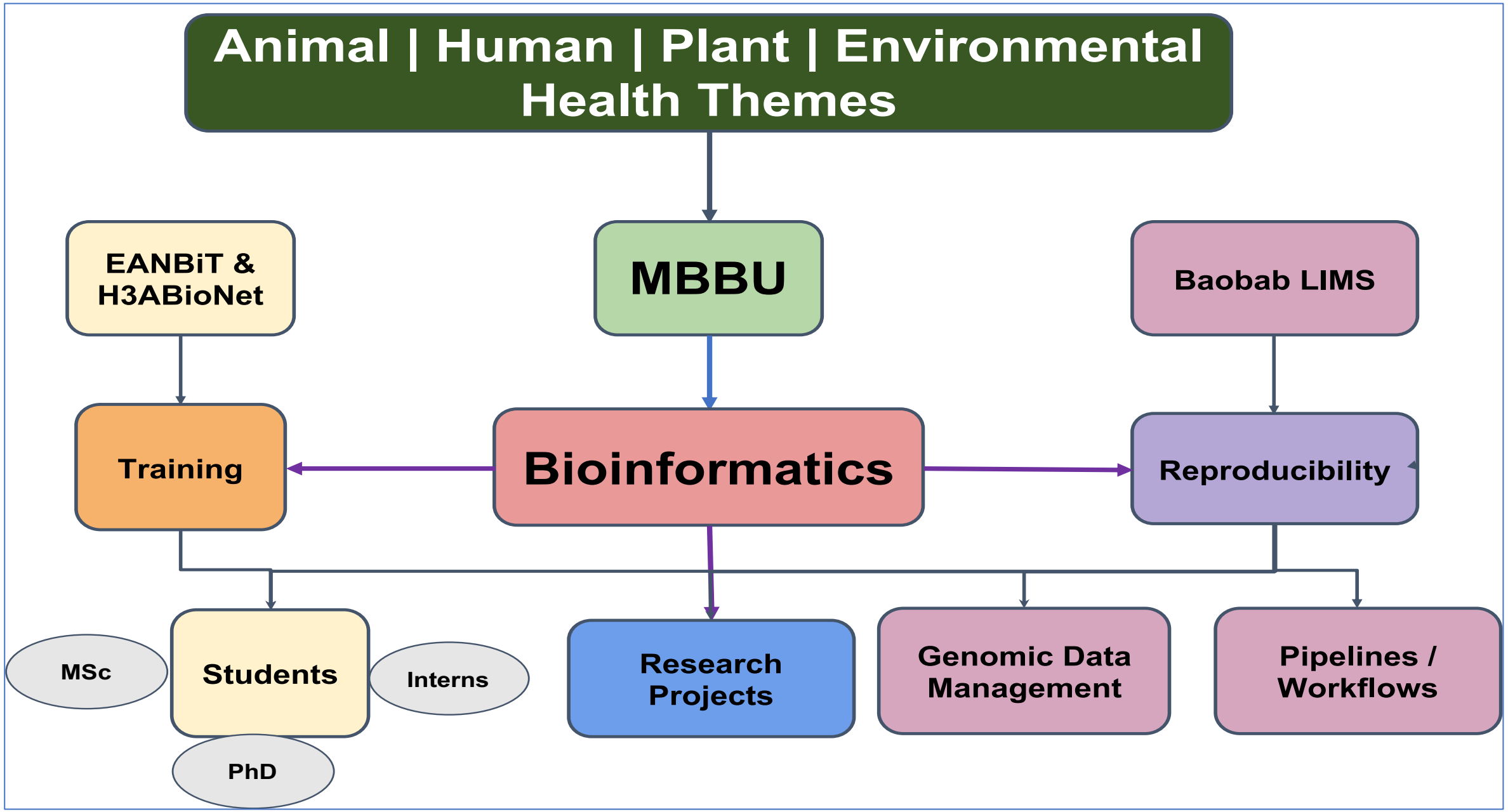
Bioinformatics Strategy

Mission: To lead and support cutting-edge bioinformatics research and Genomic Data Management at *icipe*

Vision: To become a vibrant, productive, innovative, and flexible bioinformatics training, research and genomic data analysis hub

Goals:

- Provide sufficient computing infrastructure and tools for genomics
- Develop standardized analysis workflows
- Build bioinformatics capacity through training
- Establish Bioinformatics research programs


Animal | Human | Plant | Environmental Health Themes
EANBiT & H3ABioNet
MBBU
Baobab LIMS
Training
Bioinformatics
Reproducibility
MSc
Students
Interns
PhD
Research Projects
Genomic Data Management
Pipelines / Workflows

Ensure:

- Compliance to SOPs
- Efficient Sample management
- Proper documentation / labels
- Compliance with ethics

Support research reproducibility

LSMS for sample storage management

Manage:

- Samples received
- Sample storage
- Sample use

GDMS takes over after sequencing

Open and Reproducible Genomic Research

Reproducible Bioinformatics Pipelines and Tools

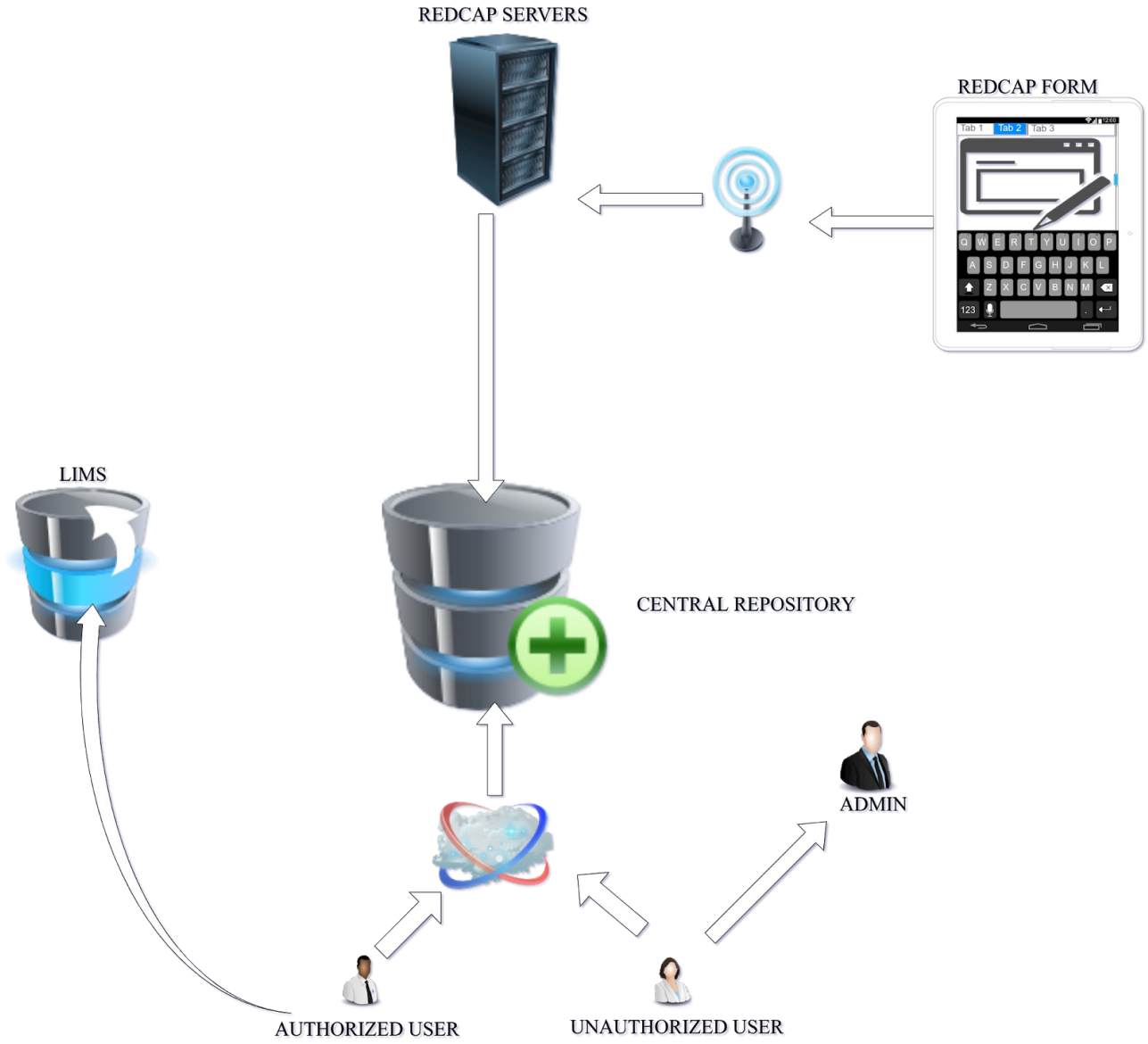
Genomic Data Management and Storage (GDMS)

- Web-based: Galaxy
- Stand-alone:
  - Snakemake
  - NextFlow
- Standard: CWL

- The data stored in the GDMS will also link to an analysis pipeline
- The pipelines also link to data for training
- Pipelines stored and managed in GitHub

- Storage
- Management
- Sharing / reuse
- Reproducibility
- FAIRness

LSMS: Laboratory Sample management System



http://lims.icipe.org:8080/index

LSMS: Impact

Visibility and Availability

- See all samples for a project immediately they're registered.
- Know what samples are available for a project at any time.

Transparency

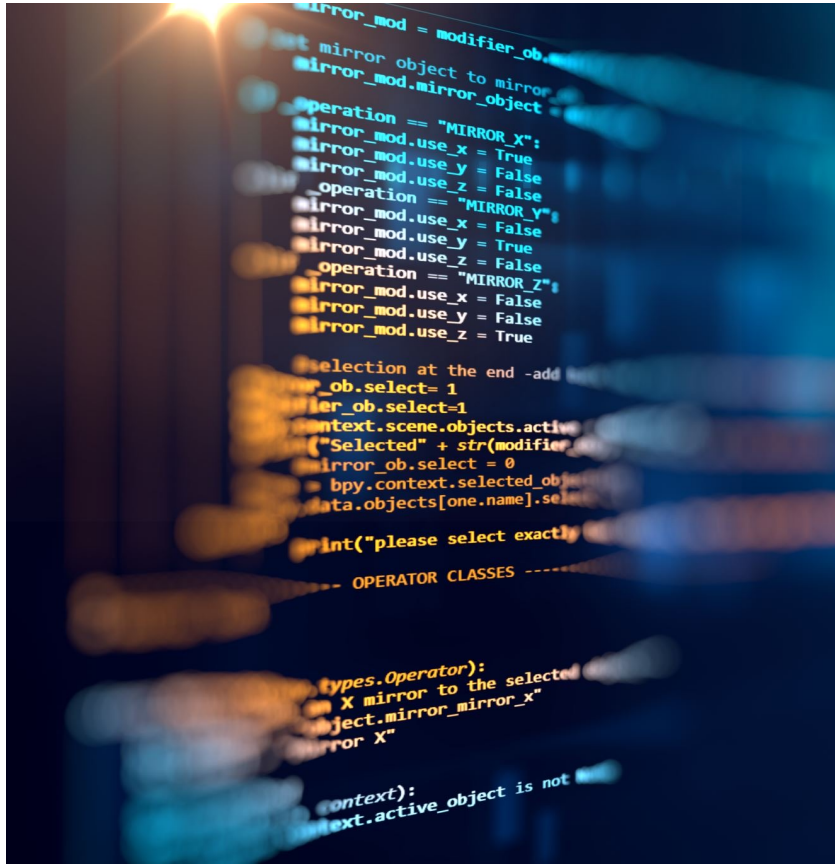
- See all your samples that were collected from a particular study block at a specific time.
- Proper flow of sample information from collection, request, approval/denial, collection and use.

Ease of Use

- User dashboards to manage data and aid in user functions.

Bioinformatics Training

- **H3ABioNet**
  - Introduction to Bioinformatics Training (>100)
  - Intermediate Bioinformatics Training (15)
  - NGS Training (15)
- **EANBiT**
  - MSc in Bioinformatics
  - Residential Training (
- **In-House Training**
  - **Bioinformatics Incubation and Mentorship Program**
  - Carpentries: Genomics, HPC use, and R
  - Machine learning for genomics

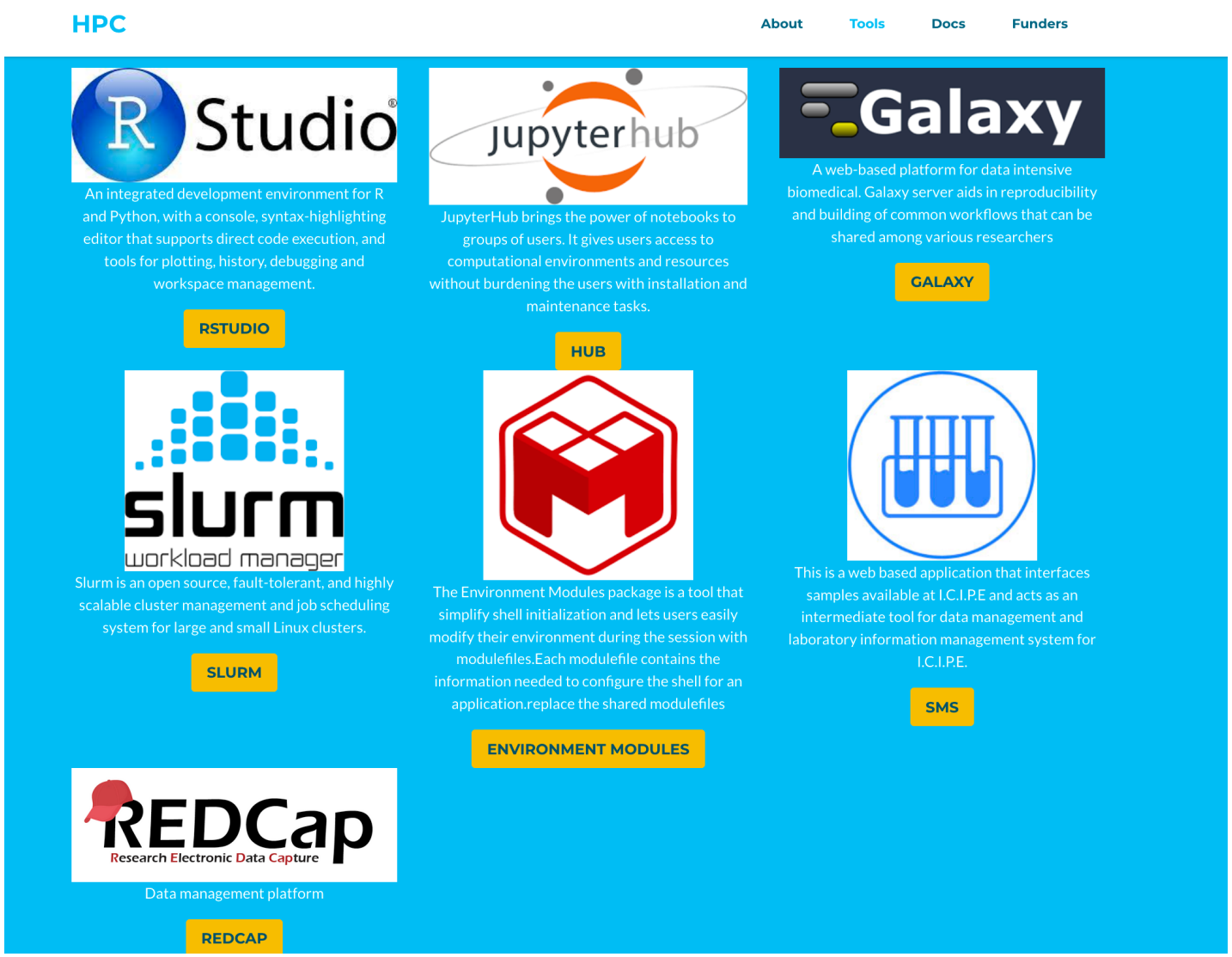


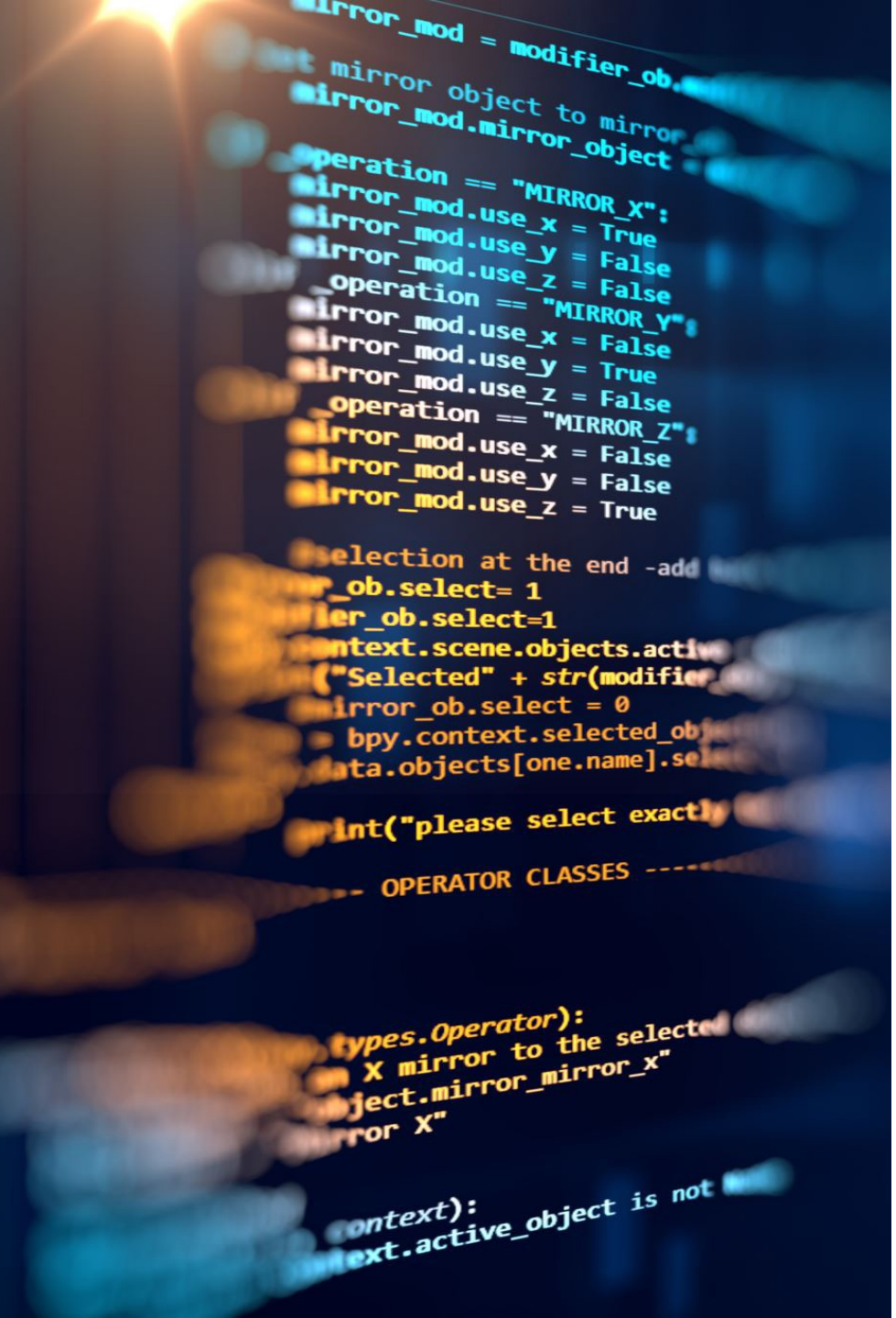
HPC: Computing capacity

Machine	cores	Storage	RAM	node
hpc	8	500GB	8GB	Log in
hpc01	64	12TB	256GB	Worker node(Intensive)
hpc02	64	8TB SSD	64GB	Worker node(fast jobs)

http://hpc01.icipe.org/hpc/

HPC Landing Page





Jupyter Hub

- Brings the power of notebooks to groups of users.
- Students, researchers, and data scientists can get their work done in their own workspaces on shared resources which can be managed efficiently
- Mainly used with Python for data analysis and machine learning.
- It is a shared resource and can be accessed through a secure web access at Jupyterhub

Galaxy

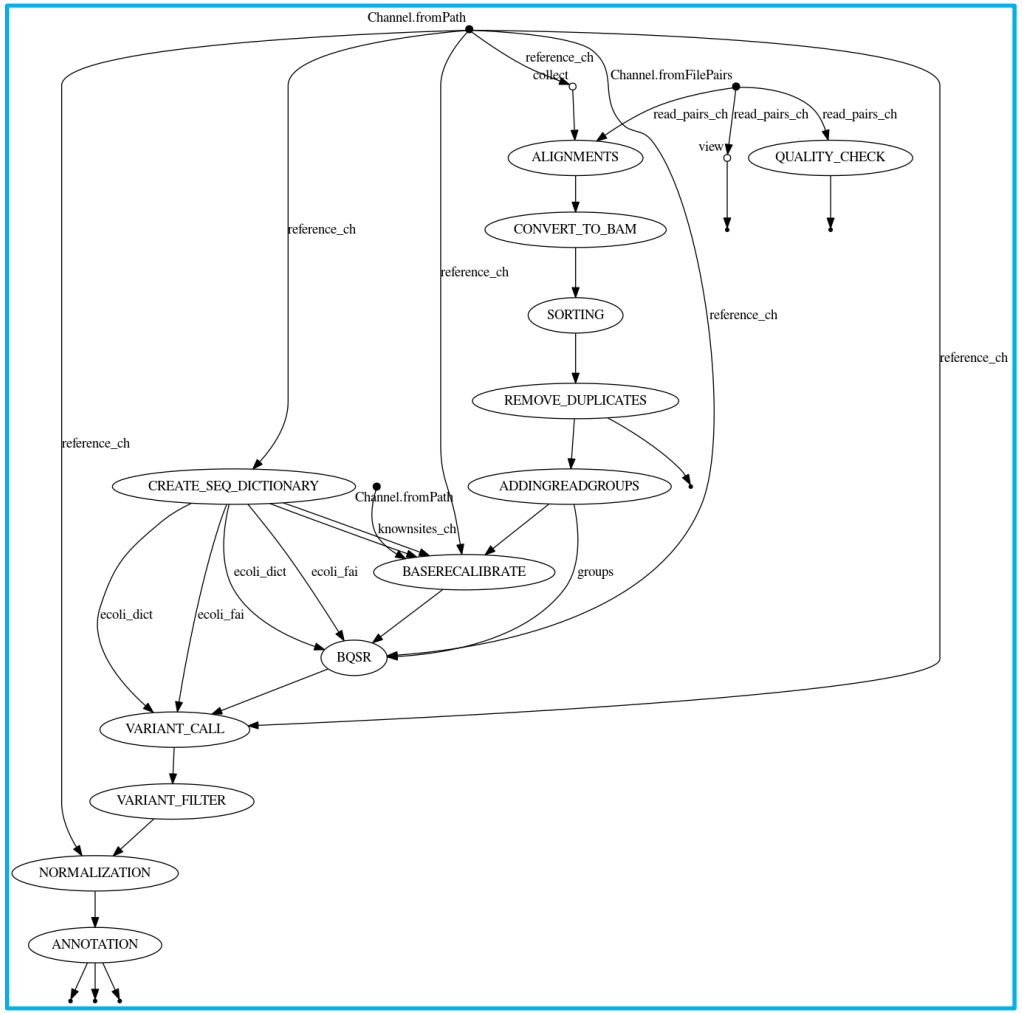
A web-based platform for Genomics analysis

Aids in reproducibility and building of common workflows that can be shared among various researchers

Galaxy server can be accessed through a secure web access at Galaxy Server

REDCap

- Secure web application for building and managing online surveys and databases.
- Facilitates compliance with 21 CFR Part 11, FISMA, HIPAA, and GDPR)
- Supports:
  - Secure data collections in the field
  - Longitudinal Studies
  - Managing applications
  - Surveys



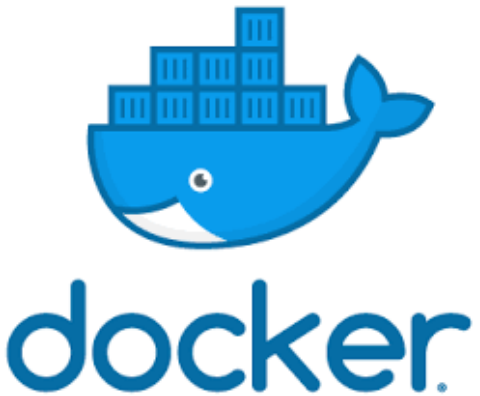
Genomic Data Analysis Capacity

We have developed standardized pipelines for:

- RNA-seq
- Variant Calling
- Metagenomics: 16S, ITS
- Whole genome assembly

How to Collaborate

- Support review of bioinformatics proposals
- Genomic Data Analysis
- Offer custom bioinformatics training



Thank you

International Centre of Insect Physiology and Ecology

P.O. Box 30772-00100, Nairobi, Kenya
Tel: +254 (20) 8632000
E-mail: icipe@icipe.org
Website: www.icipe.org

Support *icipe*: www.icipe.org/support-icipe

facebook.com/icipe.insects/icipe

twitter.com/icipe

linkedin.com/company/icipe