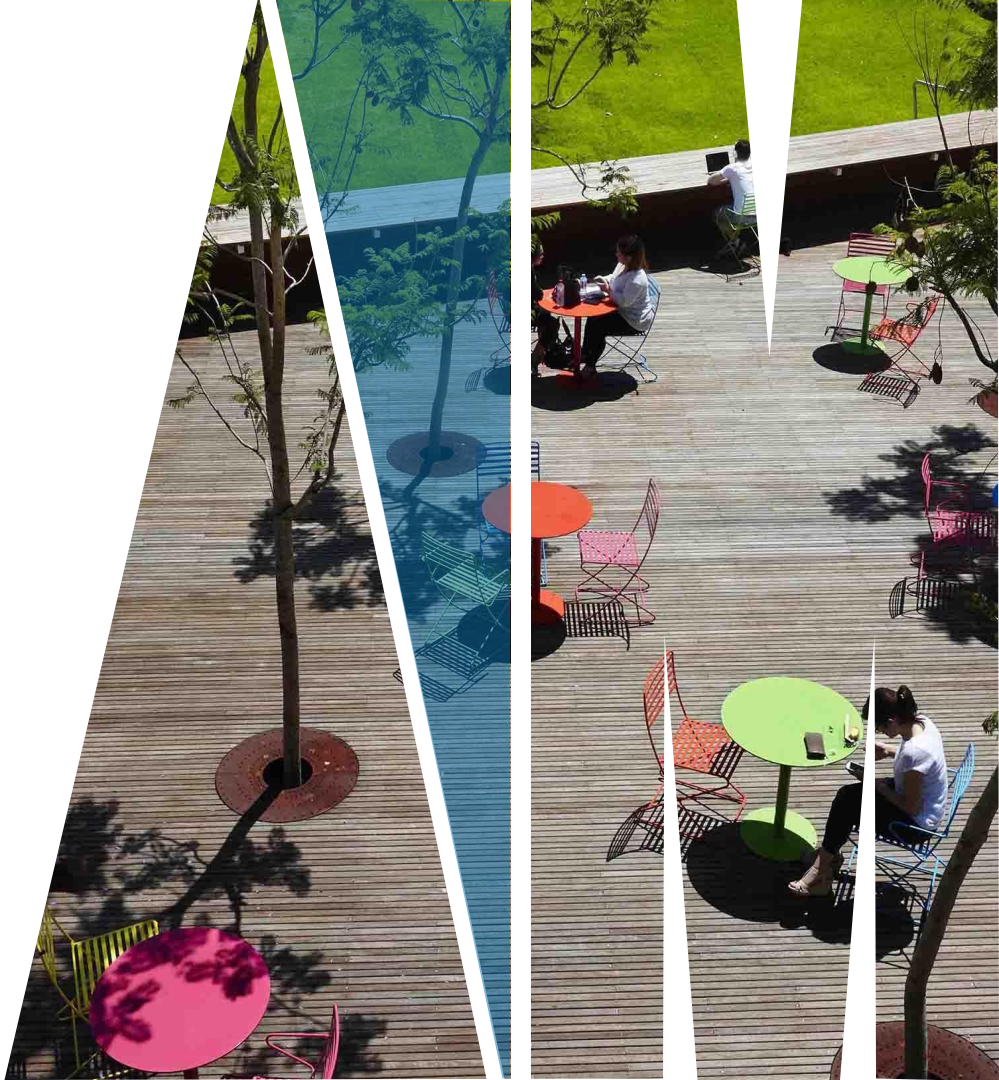
MONASH University

Monash University eResearch & MASSIVE

Monash eResearch Centre

Gin Tan



GROUP OF EIGHT AUSTRALIA

MONASH IS A GLOBAL UNIVERSITY WITH A PRESENCE ON THREE CONTINENTS AND AMBITIOUS PLANS FOR THE FUTURE

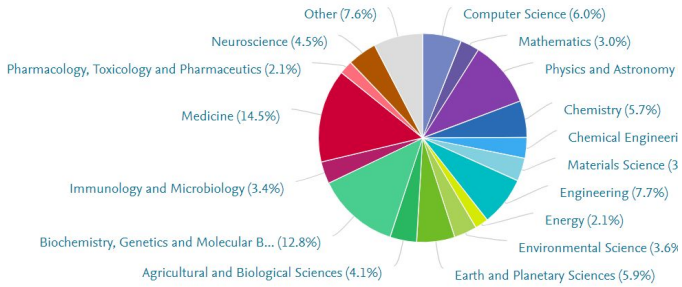
100+ partner universities

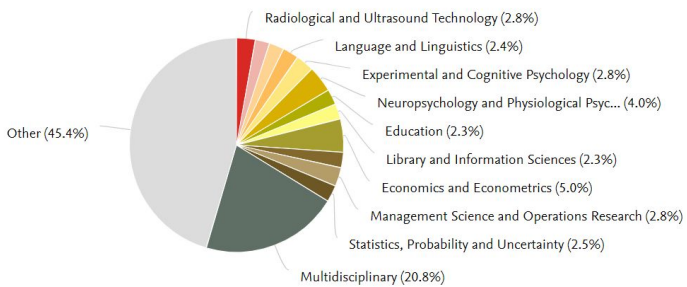
MONASH University

- Established 1958
- 347 students commenced in 1961
- Now Australia's largest university and we have 83,500+ students from more than 160 countries

Current Monash eResearch Landscape

Disciplines





Monash eResearch Projects

MASSIVE · MonARCH · RaCMon · VicNode · CVL · MyTARDIS · SeRP

Data Tools, Techniques, Infrastructure

Design and Prototyping
Visualisation
Data Science
Modelling
GPUs
VR/AR
ASICs
Software Defined Everything
Disruptive Technology
Artificial Intelligence
Deep Learning
Cloud API
IoT and Sensors
Simulation
MapReduce

Data

sensitive data
Governance
Transparency
FAIR
Self Service
Virtual Laboratories
Collaboration
Data Management
Open Science
Accessibility

Infrastructure and Capacity
Data Storage
Research Networks
Cloud Computing
Instrument Data Workflows
High Performance Computing

Contributors and Partners

NCRIS National Research Infrastructure for Australia
Microscopy Australia
BioPlatforms Australia
ANSTO
AuScope
National Imaging Facility

Gov and NCRIS
ARDC
Australian Research Data Commons
PAWSEY supercomputing centre
CSIRO
NCI Australia
aarnet Australia's Academic and Research Network
ANFF

AI and Compute
NVIDIA

Hardware
Mellanox Technologies
DELL EMC

Middleware
redhat
The Linux Foundation
cumulus
ceph

Cloud
openstack
Google Cloud Platform
amazon web services
Azure

Technical Descriptions

On Premise Technical Infrastructure

Certified System
Quality
ISO 9001
SAI GLOBAL

Research Safe Havens and Virtual Laboratories

Virtualised GPUs and HPC-interconnects on the cloud. Australian-wide access, and advanced high-performance networking. **Safe Havens** Environments support for Health Registries: ICTACCT, MPHF, TAC Propensity, TARN VSTR, VSTR Weights, VIBES, ANZSCTS, ASPREE, PCOR, VCOR, ATR. National Multi-modal Drone Data Processing VL. VL for **NetZero** and Smart-Cities Microgrid. National **Characterisation** VL

Scientific Instrument Integration and Software Platforms

Integrated: Cryo-EM (5), EMs (5), Optical Microscopes (34), Mass Spectrometers (10), MRI Scanners (2), Ultrasound Scanners (1), PET-CT (1), X-ray Computed Tomography (1), Next-Gen Sequencers (6), Macromolecular Crystallography Beamlines (2), Neutron Scattering Instruments (2).

High Performance Computing

MASSIVE M3 is composed of 6600 CPU cores, 350 GPU co-processors across a range of products suited to support parallel data processing, visualisation and machine learning, and a 3 PB parallel Lustre file system and M3 provides a combination of GPU coprocessors, including the NVIDIA K1, K80, P100, V100, and the DGX1-V. 927 Scientific Software Packages. **MonARCH** is composed of 2584 CPU cores across a range of products suited to simulation, modelling and data processing

Research Data Storage

High Performance Parallel Storage (Lustre) 5PB
Computational Cloud Storage 1.5PB
Market (Mid-Tier) capacity ~2.7PB
HSM (Slower but High Capacity) capacity 10.3PB (and can hold ~16PB with additional tapes)
Backup (of Vault and Market) ~18PB

Research Cloud (National & Private)

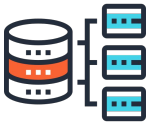
Monash Research Cloud - Located at Monash and underpins all eResearch infrastructure. Servers host the MASSIVE and MonARCH HPC cloud tenancies above, and the remainder cloud IaaS (non-HPC) is specifically composed of 7174 virtual threads, 600TB of VM storage and 20 GPUs for interactive desktops. **Open stack, Ceph technologies.** Devops and research-led co-design. Totalling >100Million CPU-hours in 2019, across self-service IaaS, safe havens, virtual desktops, facility platforms, HPC and cloud native applications.

MONASH University

Monash eResearch Centre - 2019 Performance Indicators*

5914 Users
(overlapping across Storage, HPC, Cloud)

2968 Research Projects



15.4 Petabyte (PB=1,000TB) of Research Data

131,172,560 CPU hours in 2019

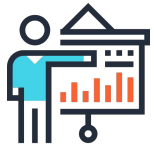
932,404 GPU hours in 2019

Current Ingest Rate: >5 PB new data per year

86 Major Instruments Connected

7824 *Secondary* users

927 different scientific software packages on HPC, adding 5 per week

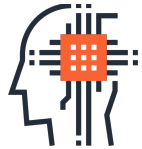


5395 Support Engagements in 2019

* Data at Oct 2019

Data Privacy

Trend: Data privacy is the most trending conversation with GDPR and other Privacy protection regulations introduced



AI & Machine Learning

Trend: Artificial Intelligence and Machine Learning is the fastest growing technique applied by researchers across different fields

MASSIVE

High Performance Computing for Data Science

Mission: MASSIVE is a data processing engine for Australian science and it empowers researchers to unlock impactful research discoveries within scientific data.

MASSIVE Focuses

- Researcher and research community focused
- Accessible to a wide range of researchers, with focus on new user communities
- Dedicated to quality and best-practice

MONASH University

MASSIVE

High Performance Computing for Data Science

Monash University-led, ANSTO, CSIRO, UoWollongong, ARC CoE in Integrative Brain Function, ARC CoE in Advanced Molecular Imaging

Technical

- 6600 Intel CPU cores over 225 servers
- 350 GPU co-processors:
  - 32 x NVIDIA K1
  - 66 x NVIDIA K80
  - 20 x NVIDIA P100
  - 72 x NVIDIA V100 (excl DGX)
  - 66 x NVIDIA P4
  - 6 x NVIDIA T4
  - 11 x NVIDIA DGX-1V, each with 8 x NVIDIA V100
- 3PB fast parallel Lustre file system
- Remote desktops for analysis and visualisation (Strudel):
  - 32 low end desktops, 30 high end desktops, 66 mid tier desktop

927 scientific software packages and libraries installed at researcher request

MONASH University

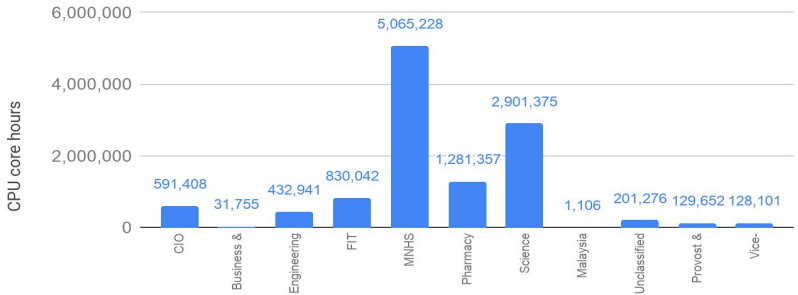
Capabilities

Support for fast parallel data processing and analytics
Support for new communities
Support for big-data producing instruments
Growing machine learning community alongside with ARDC ML
National leader in Characterisation & Data Processing - ADRC ACCS

Largest user communities:
Neuroscience, CryoEM, Machine Learning, Bioinformatics

User Community (2019 data)

302 projects, 1,187 users



Research Data Storage

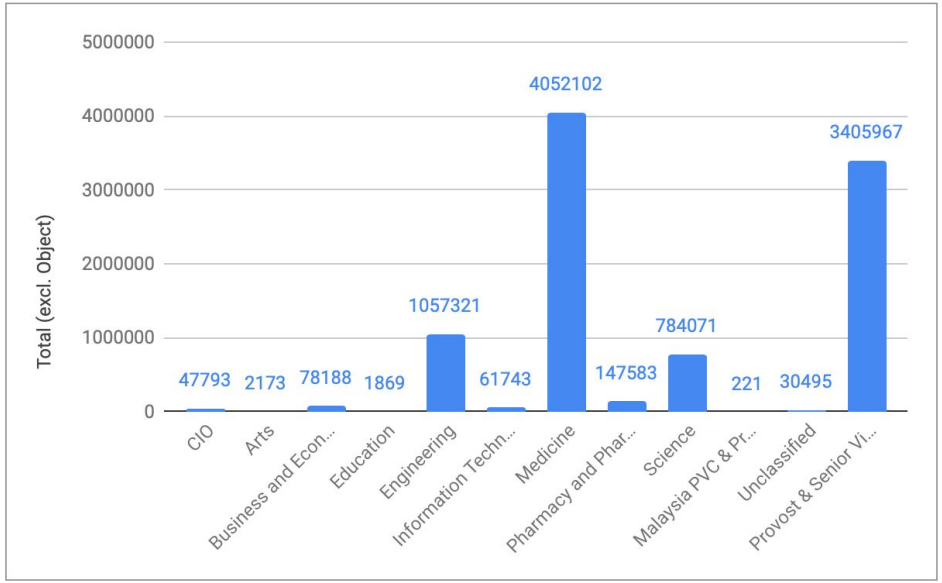
Research data lifecycle management

Breadth and scale of storage

- Tightly connected to Monash's ecosystem of **Research Cloud, MASSIVE, Monarch** and the **enterprise**.
- Storage pools of several tiers
  - **Market** (always online) for user filesystems (Windows and NFS), and research services.
  - **Vault** for big amounts of data, archiving, and infrequent use.
  - **Ceph Object with swift** for cloud-native research applications.
- Timely disaster recovery of MASSIVE and storage pools.
- Underpins data lifecycle management
  - Community resources e.g. **Store.Monash** capturing all university produced imaging data upon generation (in conjunction with **CVL** and **MASSIVE**)
  - **Figshare** making self publishing of open data easy (in conjunction with the **Library**)
  - **ASPERA** getting large data in/out Monash fast
  - **Murda** for data sentenced for retirement (in conjunction with **Records and Archives**)

User Community (2019 data)

742 major collections, 10 petabytes of presented storage
(approx 26 petabytes raw including backups/DR of HPC)



Data and workflow Orchestration

Single centralised orchestration platform

MONASH University

Acknowledgement

Gin Tan

Principle Research Systems Architect

For more information:

- monash.edu/researchinfrastructure/eresearch
- massive.org.au

MONASH University