



The University of Melbourne

Research Data Services with Mediaflux®

The University of Melbourne, a world-renowned research institution, has been utilising Mediaflux for research data management for the past 16 years. Over this time, the university has greatly benefited from the comprehensive capabilities of Mediaflux, addressing the challenges of storing, managing, and sharing research data across its extensive research computing infrastructure. As the volume of research data has grown, the need for an efficient and secure data management system has become even more crucial. The university's continued reliance on Mediaflux affirms the system's effectiveness in optimising data management and ensuring that scientists can efficiently conduct cutting-edge research. In this case study, we explore how the University of Melbourne has leveraged the various functionalities of Mediaflux to enhance its research computing infrastructure and support the needs of its researchers.

Challenges

- Multiple research teams, each with its own data storage systems
- Limited data sharing among siloed researchers
- Difficult to manage user access to research projects securely
- No data redundancy, increasing the risk of data loss and making data recovery challenging
- Sensitive research data was vulnerable to unauthorised access, corruption, or loss

Results

- Significantly boosted the University's research outcomes and overall research efficiency
- Improved collaboration
- Enhanced data security, streamlined data management, effective data sharing
- Reliable data redundancy and backup
- Reliable data access, and high scalability

Challenges

The University of Melbourne faced several challenges in managing its research data. The university had a decentralised data storage scheme, with multiple research teams each having its data storage systems, leading to difficulties in organising, accessing, and managing research data. This decentralisation also hindered collaboration, as data sharing between research teams was limited, impeding the flow of information and the potential for interdisciplinary research. They had inadequate access controls, making it challenging to securely manage user access to various research projects. Moreover, the university lacked robust data security measures, leaving sensitive research data vulnerable to unauthorised access, corruption, or loss. The absence of an effective data redundancy system increased the risk of data loss and complicated data recovery. These challenges impeded research efforts, prompting the university to seek a comprehensive solution for research data management.

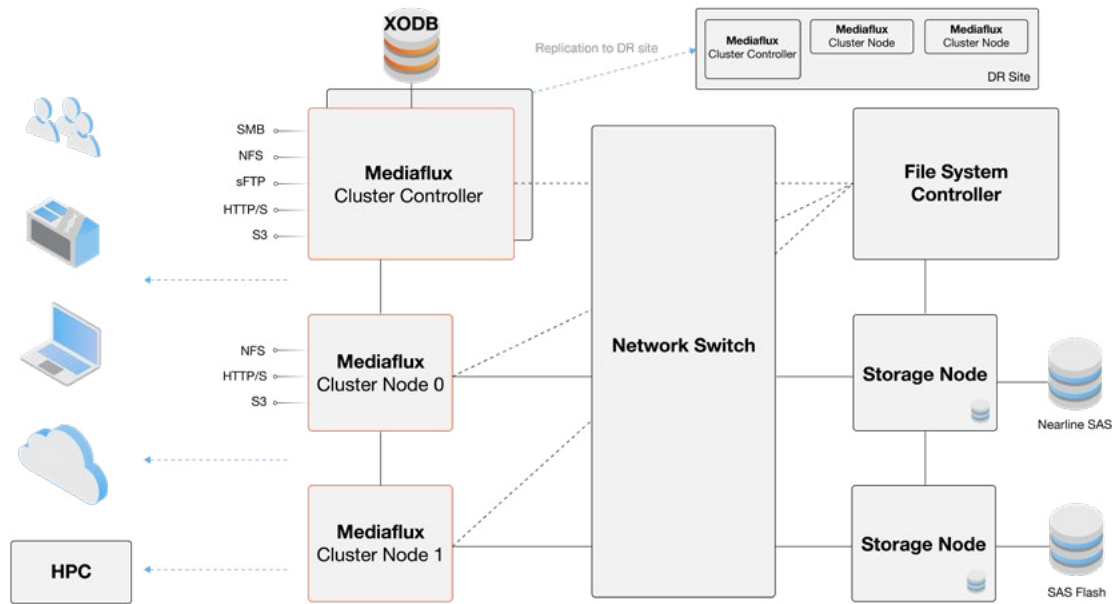
The University's Research Data Storage System

As shown in the next page, the university uses 4PB of IBM GPFS parallel, filesystem for its HPC workloads. Mediaflux front-ends the rest of the university's 10PB research storage infrastructure, creating a shared global namespace across highly scalable CEPH storage, its remote DR site at the Noble Park data centre, and AWS/S3.

Mediaflux's diverse range of features and its adaptability to the university's unique needs have made it a crucial tool for data management within the institution.

The cornerstone of the University's service offering with Mediaflux is the project namespace, effectively acting as a cloud-based file system. This feature allows researchers and their teams to securely store and share data collections for their projects. It offers great flexibility, as users can structure their namespace in any way, they find convenient and associate discoverable metadata with both the structure and the uploaded files. This makes the data more easily searchable and accessible, streamlining the research process.





THE UNIVERSITY OF MELBOURNE RESEARCH DATA INFRASTRUCTURE

The Solution: Mediaflux

Controlled Access

Authentication in Mediaflux is robust and user-friendly. University staff and students can directly login with their institutional credentials, whereas other users can use a local account created for them.

Roles are essential to authorise access to resources. To enhance user management and data access control, The University of Melbourne team implemented role definitions matched to ACLs. This means they can control what collections users have access to and the extent of their permissions within a single pool of virtualised storage. These access controls and security measures have enabled the University of Melbourne to design their own project management layer within Mediaflux. The flexibility of Mediaflux allowed the university's research team to create standard roles per project (admin, read/create/destroy, read/create, read), which can be assigned to project team members.

Mediaflux accommodates multiple access protocols, including HTTPS (browser-based access and various Java clients), SMB (network file share), sFTP (FileZilla, CyberDuck, rclone), and NFS and these are supported by UoM. Such flexibility is crucial in accommodating different user preferences and ensuring seamless data access across a variety of platforms.

Sharing Data with the Mediaflux Data Mover

Data sharing in Mediaflux is easy and secure. Mediaflux's Data Mover capability and support for instrument uploads facilitate efficient data movement, including uploads and downloads of large

data sets. This is especially important for researchers working with high-volume data generated from scientific instruments and equipment. Data can be shared with external users who don't have accounts via shareable links. These links can also be used for uploading data by external users, facilitating collaboration with researchers and institutions outside the University.

Enhanced Security

Mediaflux emphasises data security, supporting encrypted transfers through HTTPS and sFTP protocols. Selected metadata can also be encrypted, ensuring robust protection against unauthorised access. The University leverages this capability for their DaRIS services metadata framework, significantly bolstering their data security.

Versioning and Disaster Recovery

Data redundancy and backup in Mediaflux provide a crucial safety net for the University's research data. Assets are versioned, with old versions retrievable, offering an effective solution for accidental deletions or modifications. The Disaster Recovery (DR) server, located at the Noble Park data centre, holds copies of all asset versions, ensuring data recovery in case of primary server failures. Although it's not used as a fail-over, the DR server plays a vital role in data preservation.



Benefits

The University of Melbourne realised several benefits from implementing Mediaflux for research data management:

Enhanced Collaboration

Mediaflux's project namespace feature, along with the ability to structure the namespace and associated discoverable assets, make it easy for researchers and their teams to securely store and share data collections. This facilitates more efficient collaboration within the University and with external partners.

Robust Data Security

Mediaflux's authentication and authorisation mechanisms, combined with encryption for data transfers, ensures that only authorised users could access the data. This robust security significantly reduced the risk of data breaches and unauthorised access.

Efficient Data Management

With multiple access protocols, the Data Mover capability, and efficient data movement and instrument uploads, researchers can access and manage their data in a way that suits their needs. This streamlined data management process allows researchers to focus more on their work and less on administrative tasks.

Easy Data Sharing

Mediaflux's shareable links make it simple for researchers to share data with external users, facilitating collaborations with researchers and institutions outside the University. This ease of data sharing increases the reach and impact of the University's research.

Reliable Data Redundancy and Backup

The versioning of Mediaflux assets and the presence of the Disaster Recovery (DR) server ensure data availability and recoverability, protecting against accidental deletions, modifications, malicious actors, or primary server failures. This reliable redundancy and backup gives researchers peace of mind, knowing that their data is safe and recoverable.

Scalability

The scalable architecture of Mediaflux, combined with the scalable CEPH cluster for storage, offered a flexible and extensible environment capable of accommodating the University's growing data movement needs. This scalability ensures that the University's research data management infrastructure can keep pace with its expanding research requirements.

Improved Research Outcomes

All the above benefits collectively contribute to a more efficient research process, accelerating discoveries and increasing innovation. The improved collaboration, security, data management, sharing, redundancy, access, and scalability all lead to enhanced research outcomes for the University of Melbourne.



Extensive Array of High-Throughput Scientific Data Sources

The University of Melbourne is equipped with an array of advanced research technologies, spanning various fields and disciplines. Their platform technologies cover a wide range, from advanced microscopy to cultural informatics, and metabolomics to health economics – all inputting massive amounts of data into Mediaflux. The University's Research Computing Service supports scientists in achieving the ground-breaking insights that make the University of Melbourne a premier research institute in Victoria, and one of the best universities in the world.

Genomics and Bioinformatics

The University's expertise in genomics and bioinformatics encompasses human and non-human genomics, cancer genomics, next-generation sequencing, machine learning, biological data visualisation, software development, Galaxy development and resourcing, and bioinformatics tool development. They have the capability for spheroid and organoid phenotyping and screening, compound screening, CRISPR pooled and arrayed screening, high-content, live, and fixed imaging, and RNAi screening.

Microscopy

The University has a wealth of microscopy equipment for structural and cell biology, including cryo-electron microscopy (cryo-EM), three-dimensional electron microscopy (tomography, block face imaging, and array tomography), standard and high-resolution 2D imaging, elemental analysis, and nanofabrication. They also possess standard and cryo-scanning electron microscopy.

Mass Spectrometry

The University's mass spectrometry capabilities include dozens of Thermo Fischer, Agilent, Waters, Bruker, and Shimadzu mass spectrometers.

Medical Imaging

Their medical imaging capabilities include a Siemens Biograph Vision 600 PET/CT scanner and a Siemens Magnetom 7T MRI scanner.

Data Analytics and Machine Learning

The University has a strong focus on data analytics and machine learning for research.

