UCT eResearch
ACCELERATING RESEARCH

Emerging Researcher Seminar Series

Date	Venue	Time	Title
Thurs, 28 July 2016	**Research Office Board Room**	12:30-14:00	What is eResearch?
Thurs, 4 August 2016	**Research Office Seminar Room**	12:30-14:00	Proposal planning
Thurs, 11 August 2016	**Research Office Seminar Room**	12:30-14:00	Gathering data
Thurs, 18 August 2016	**Research Office Seminar Room**	12:30-14:00	Research collaboration
Thurs, 25 August 2016	**Research Office Seminar Room**	12:30-14:00	Analysing data
Thurs, 8 September 2016	**UCT Libraries, Hlanganani 6**	12:30-14:00	Data visualisation
Thurs, 15 September 2016	**UCT Libraries, Hlanganani 6**	12:30-14:00	Data management
Thurs, 22 September 2016	**UCT Libraries, Hlanganani 6**	12:30-14:00	Open access publishing

UCT eResearch
ACCELERATING RESEARCH

All presentations from the emerging researcher series will be available from:

http://www.eresearch.uct.ac.za/eresearch-training

Research data lifecycle

Gathering Data

Emerging Researcher Series

Thursday, 11th August 2016

UCT eResearch
ACCELERATING RESEARCH

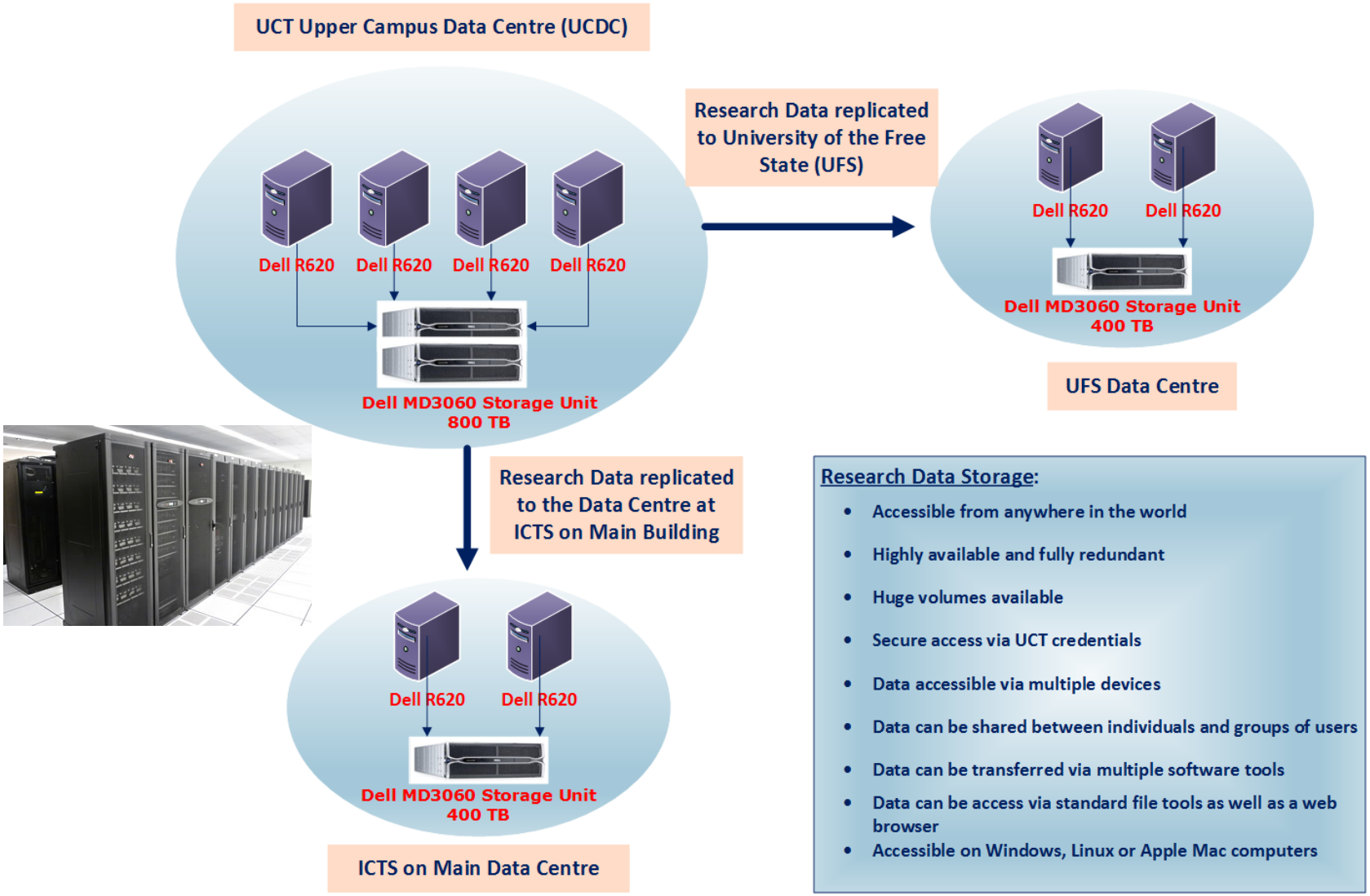
Ashley Rustin - **Senior eResearch Technical Specialist**
Niklas Zimmer – **Head: Digital Library Services**

www.eresearch.uct.ac.za

UCT eResearch
ACCELERATING RESEARCH

Data (File) Storage

UCT Research data storage

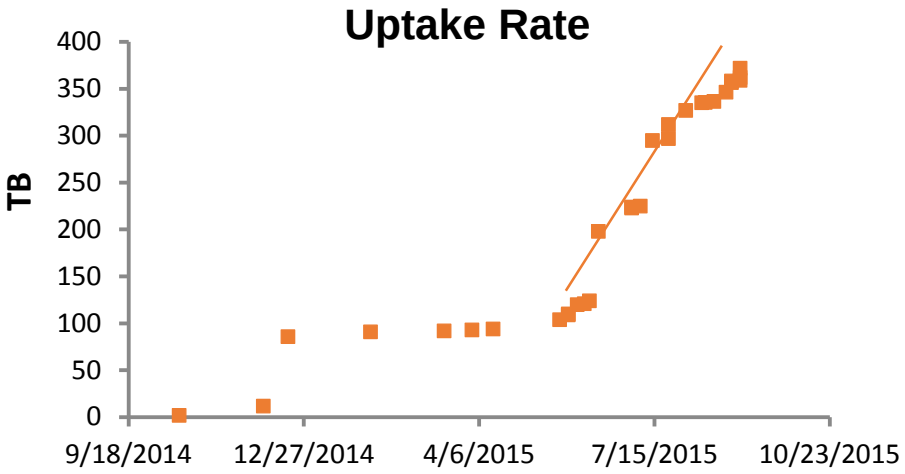


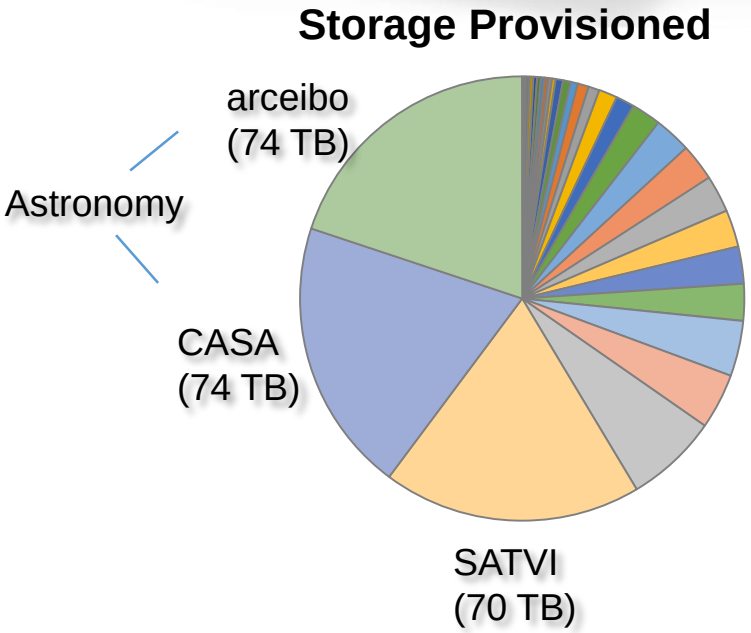
Research Data Storage:

- Accessible from anywhere in the world
- Highly available and fully redundant
- Huge volumes available
- Secure access via UCT credentials
- Data accessible via multiple devices
- Data can be shared between individuals and groups of users
- Data can be transferred via multiple software tools
- Data can be access via standard file tools as well as a web browser
- Accessible on Windows, Linux or Apple Mac computers

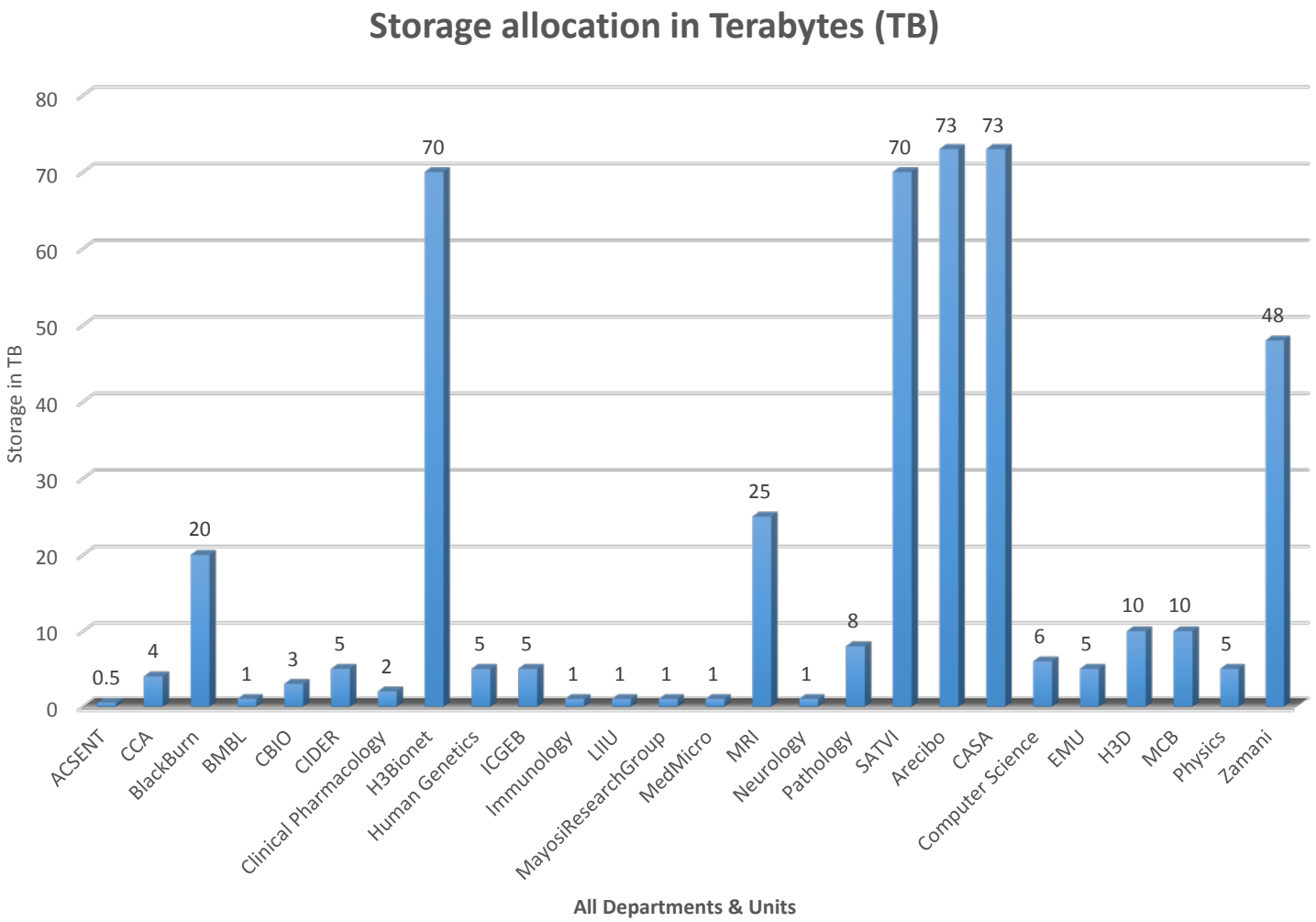
Research data storage

- Almost **400 TB** allocated to date
- +- 700 Users accessing our Research Data Storage
- Current rate **40 TB/m** provisioned
- **90 TB** fast parallel storage on HPC (fhgfs)



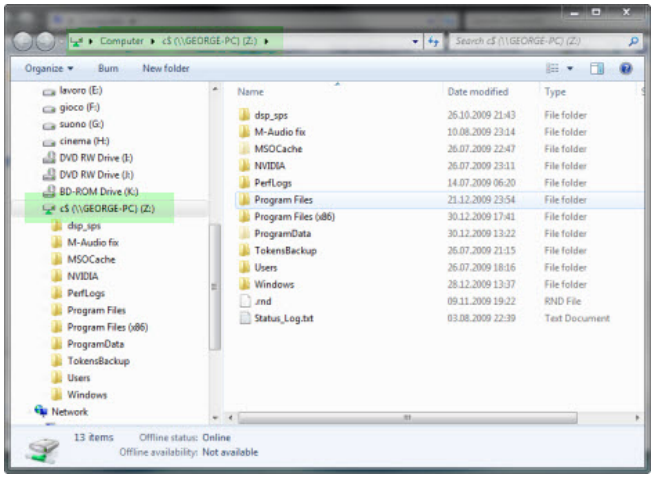


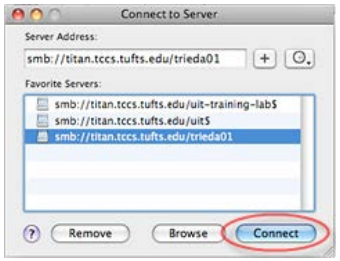
Research data storage allocation:



Accessing your Research data storage

- A **mapped network drive** on your Windows, Linux or Apple computer
- **Web browser** using NextCloud to access all your external storage, eg. Research Data, Drop Box, One Drive, Google Drive, etc.
- Storage available in increments of 1 TB
- Data accessible from anywhere in the world

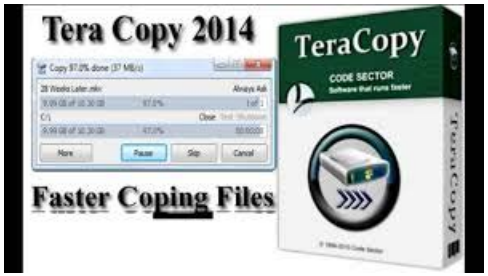




Nextcloud

Cyberduck

Tools for accessing & moving my data



Windows	Linux	Apple Mac
TeraCopy	SCP	SCP
RichCopy 4	RSync	RSync
FastCopy	UltraCopier	UltraCopier
XXCopy	FastCopy	FastCopy
FreeFileSync	FreeFileSync	FreeFileSync
Windows Explorer		
RoboCopy (Command Line)		
FileZilla		

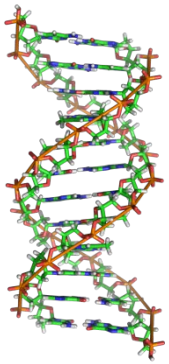
How do I request data storage?

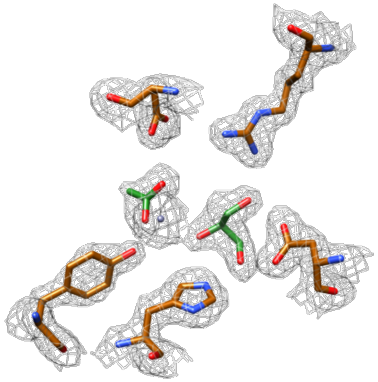
- **Email: - eresearch@uct.ac.za or visit the UCT eResearch web site at www.eresearch.uct.ac.za**
- **We can provide guidance on data volumes and growth**
- **Cost – R280 per TB per month or R3360 per year**
- **Storage available in increments of 1 TB**
- **Please contact us for more information on storage requirements exceeding one year for a discounted rate.**

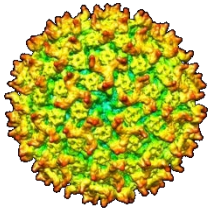
UCT eResearch
ACCELERATING RESEARCH

Automated ingest

Case Study: - EMU & SBRU







Structural Biology Research Unit

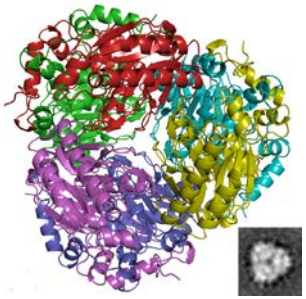
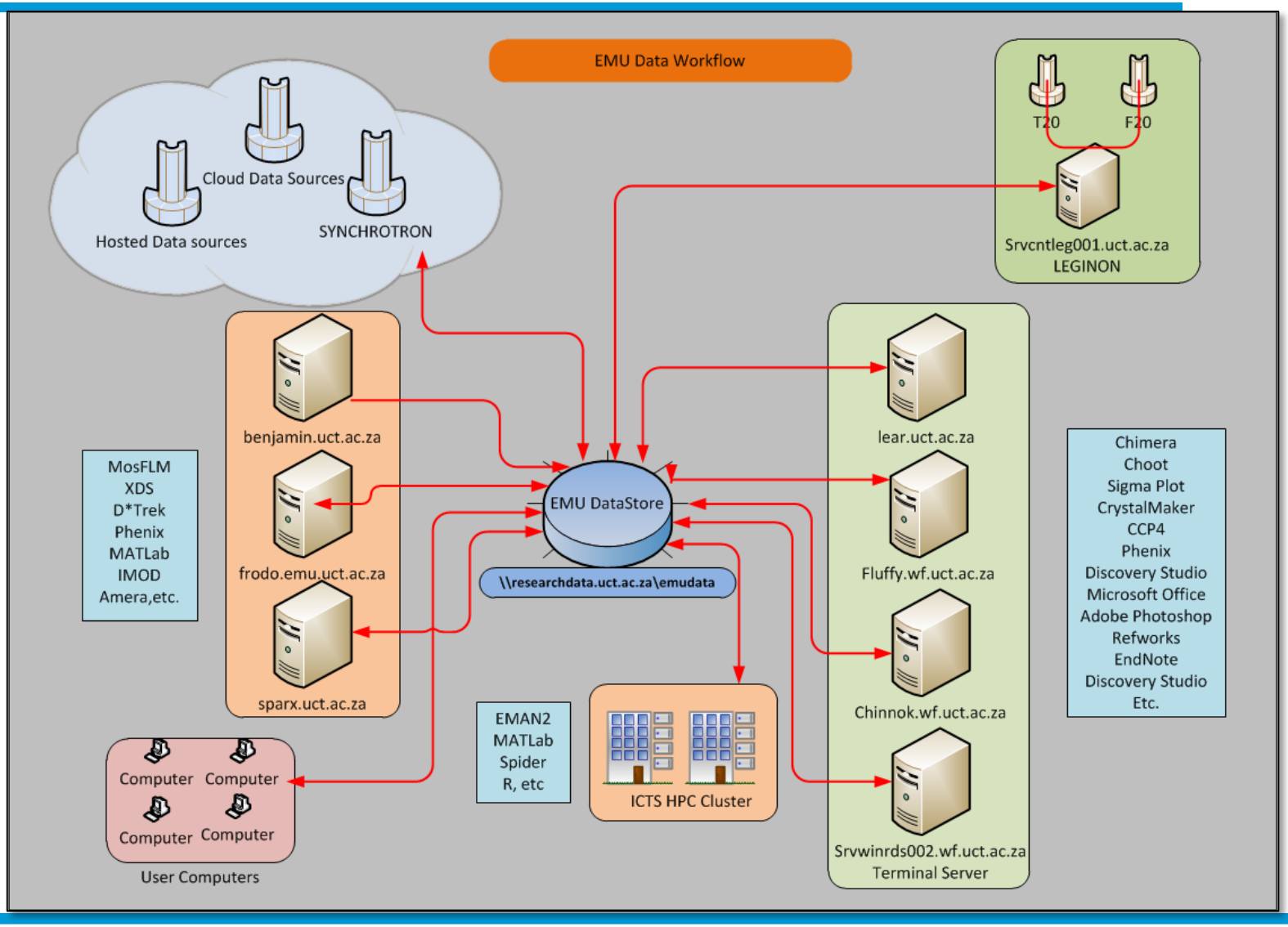


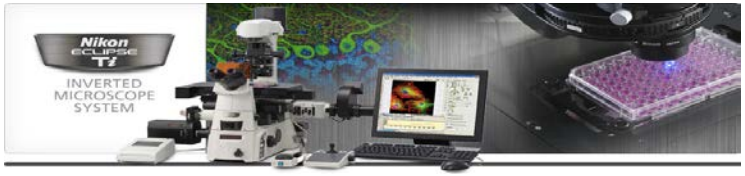
Image acquisition

Electron Microscope Unit:



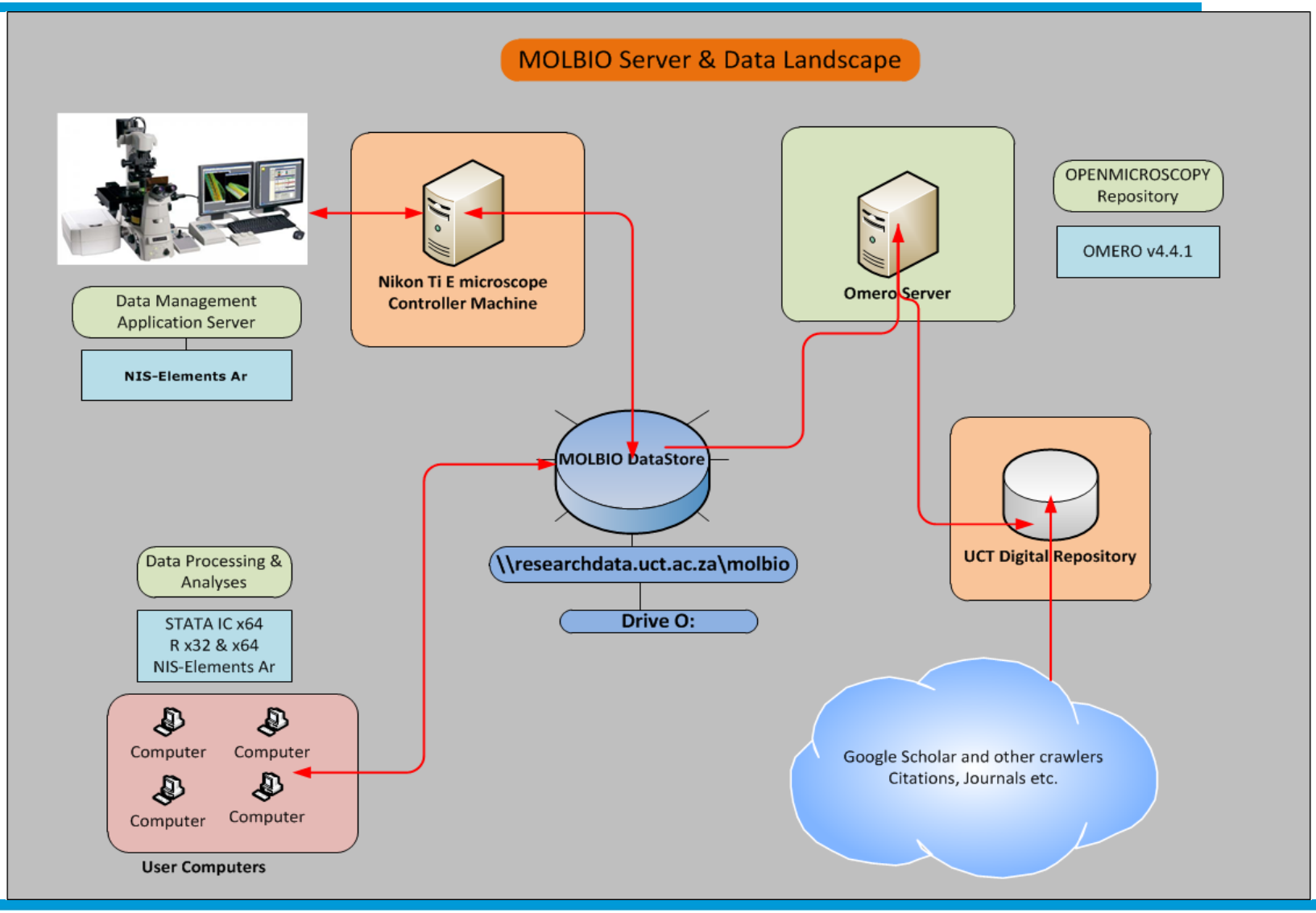
Case Study – Molecular & Cell Biology:

Prof Nicola Illing



- **Nikon Ti** Fluorescent Microscope for live cell imaging
- eResearch supported **storage**, **remote data analysis** and **collaboration**

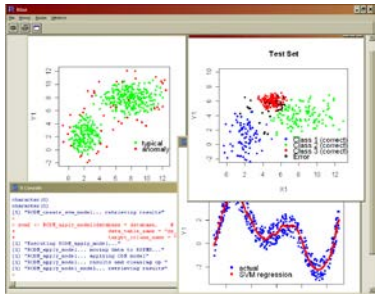
Case Study - Molecular and Cell Biology

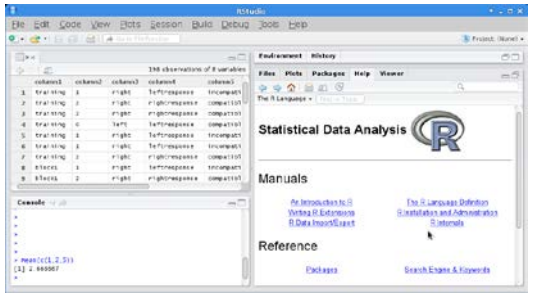


Centre for Infectious Disease Epidemiology and Research

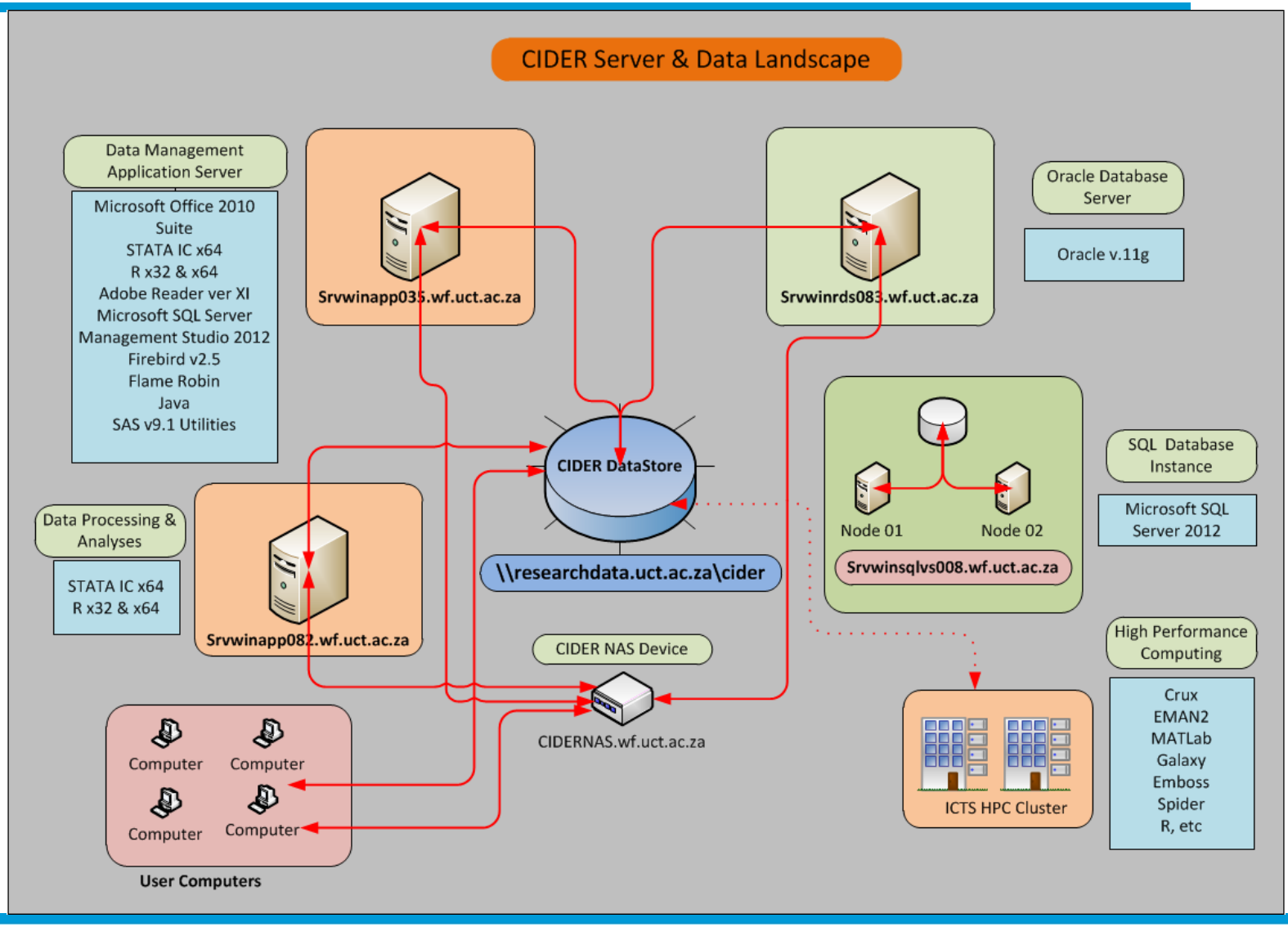
Prof Andrew Boulle

- eResearch provide assisted with setting multiple databases and processing and data analysis servers.
- Also provisioned **research data storage** and **collaboration**





Case Study - CIDER



Research data storage

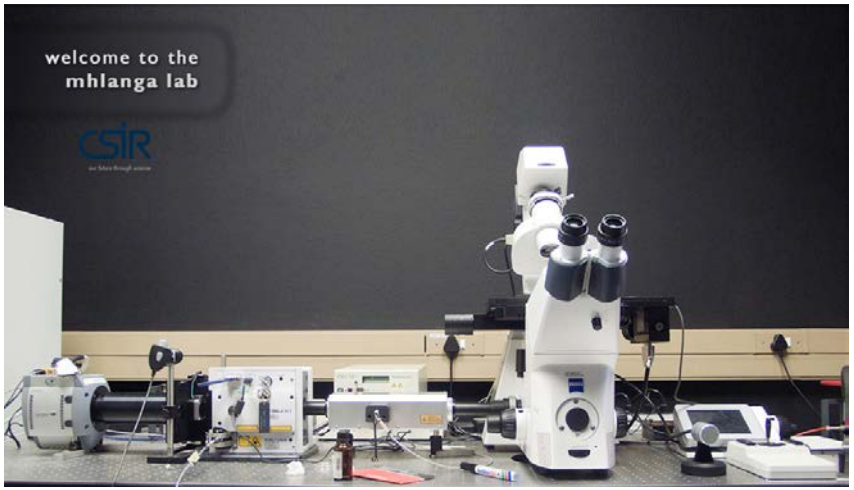
Prof Jane Yeats

- **Olympus VS120** Virtual Slide Microscope **digitizing** specimens.
- eResearch provide advice on sizing & specifying server hardware and storage. Installed and configured server software - **storage**, **remote data analysis** and **collaboration**

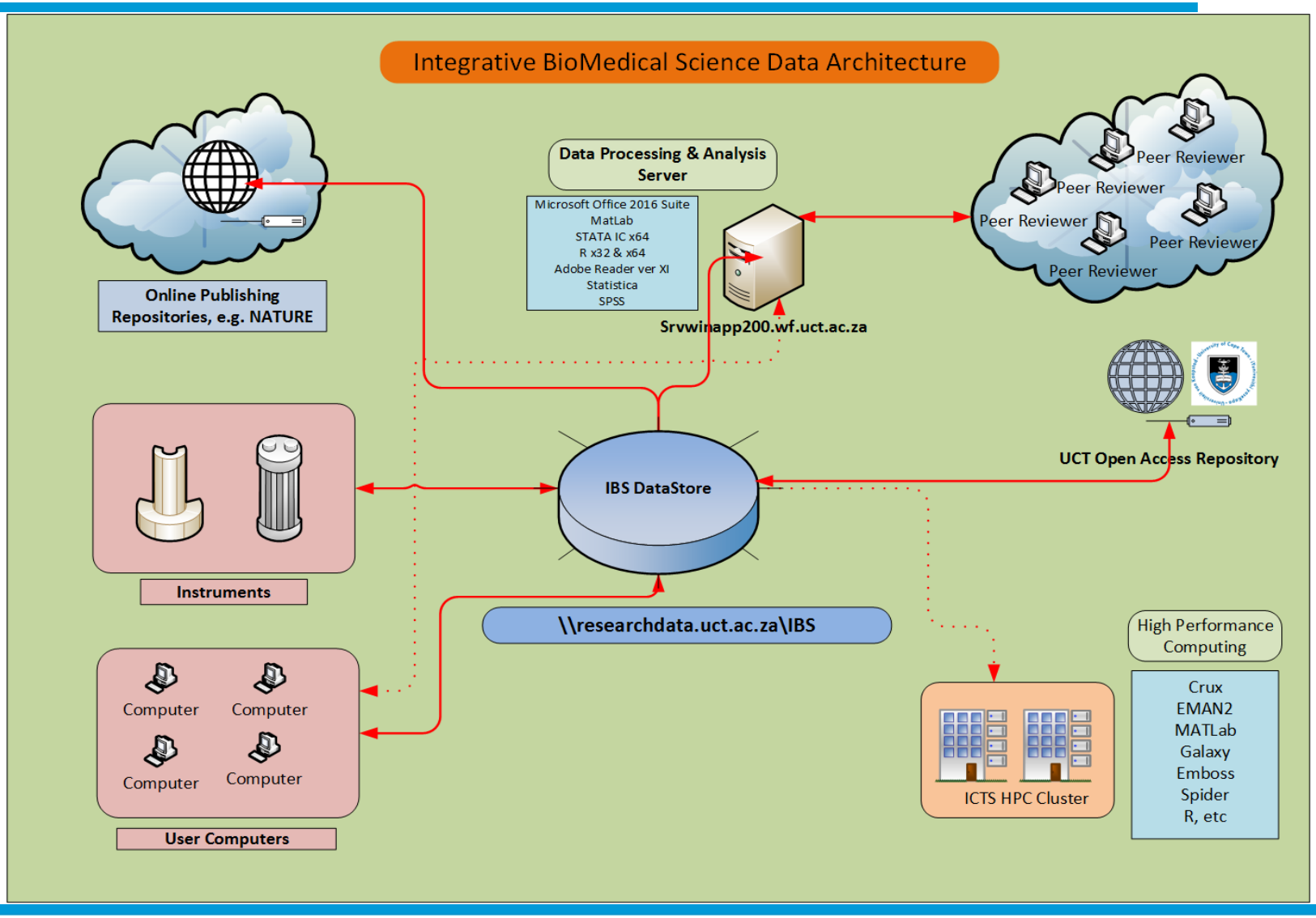
Case Study – Integrative Bio-Medical Science

Dr Musa Mhlanga

- **Dr Mhlanga is** building a super-resolution microscopy
- eResearch provide advice on sizing & specifying server hardware and storage.
- Installed and configured a server for remote access and processing, allowing international peer reviewers to **access**, **process** and **verify data** used for a paper submitted to Nature.



Case Study – Integrative Bio-Medical Science



UCT eResearch
ACCELERATING RESEARCH

Digital Scholarship

Digital scholarship

- What is digital scholarship?
- Support provided by Digital Library Services
- Good (digital) research practice around (y)our assets:
  - Metadata
  - File naming conventions / Folder structures
  - Open (non-proprietary) file formats
- Putting (y)our research funding to good use:
  - Digitising existing (primary) research materials
  - Creating ‘born-digital’ research materials
- Connecting your findings and (y)our data:
  - Open Science Framework (OSF)

Digital Scholarship

What is digital scholarship?

"*Digital scholarship is the use of digital evidence and method, digital authoring, digital publishing, digital curation and preservation, and digital use and reuse of scholarship*." (Abby Smith Rumsey, 2011)

- Digital scholarship harnesses digital tools and strategies in research, teaching, learning and publishing, such as:
  - *Data analysis & visualisation* (e.g. GIS spatial data)
  - *Text mining* (e.g. Historical studies, Law, Linguistics etc.)
  - *OA publishing*:
    - UCT Libraries **OpenUCT**
  - *Digitisation & Metadata* (e.g. Special / primary collections):
    - UCT Libraries **Digital Library Services**
    - UCT Libraries **Special Collections**
  - *Data Management Planning* (funding > DAM > reusability etc.):
    - UCT Libaries **DMPonline**

A proposed typology of digital scholarship

Adapted from: William Thomas (February 28, 2015)

	Interactive Scholarly Works	Digital Projects or Thematic Research Collections	Digital Narratives
Type of Data	Homogeneous, Primary	Heterogeneous, Primary	Integrated, Layered
Components	APIs, Scripting	Schema, Data Models	Analysis, Modules
Organization	Hypothesis	Theme or Subject	Criticism
Scope	Tightly-defined	Capacious	Problem-oriented
Interpretative Nature	Query-based	Affordances	Multi-modal
Character	Procedural Inquiry	Open Ended	User-directed, Hypertextual
Examples	ORBIS	Valley of the Shadow	The Differences Slavery Made
	Visualizing Emancipation	Whitman Archive	Gilded Age Plains City
	Railroaded	Mapping the Republic of Letters	Who Shot Liberty Valance?
	Who Killed William Robinson?	Digital Gazetteer of the Song Dynasty	Hearing the Music of the Hemispheres
	(resource): The Programming Historian	UCT Libraries Special Collections Humanitec Digital showcase	Struggle for dignity in Cape Town's informal settlements

Digital Scholarship

Support provided by Digital Library Services:

- Research Data Management:
  - Our RDM services assist you with organising, managing, and curating your research data to facilitate its preservation and access for present and future use. Together with the eResearch Centre and ICTS, we give you access to the datasets and tools you need to complete your research.

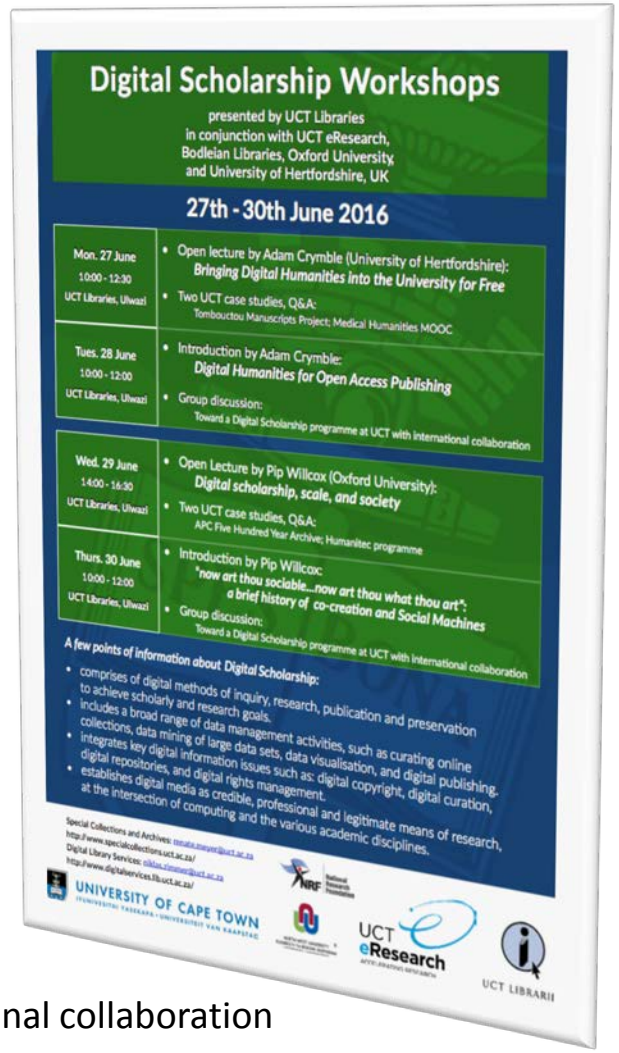
- Digitisation:
  - Our digitisation unit offers digitisation, project management, curation, and preservation services for a wide variety of audio-visual, photographic and paper documents to enable and support long-term preservation of, and access to, digital collections. We digitise objects according to international archival preservation and access standards. We also transcode, edit, and curate born-digital objects.

Digital Scholarship

Support provided by Digital Library Services:

Digital Scholarship Workshops hosted by UCT Libraries:

UCT Libraries, in conjunction with UCT e-Research Centre, Oxford University and University of Hertfordshire, hosted a Digital Scholarship workshop series from 27th-30th June 2016. We were joined by two leading international digital scholarship experts, Adam Crymble and Pip Willcox.



- Topics:
  - Bringing Digital Humanities into the University for Free
  - A brief history of co-creation and Social Machines
  - Digital Humanities for Open Access Publishing
  - Digital scholarship, scale, and society
- UCT case studies:
  - Tombouctou Manuscripts Project (HUMA)
  - Five Hundred Year Archive (APC)
  - Medical Humanities MOOC (CILT)
  - Humanitec programme (UCT Libraries)
- Group discussions:
  - Toward a Digital Scholarship programme at UCT with international collaboration

Good (digital) research practice – managing (y)our digital assets:

- UCT Libaries > **DMPonline**
- DLS > **Apply to Digitise**

- Metadata:
  - **UCT metadata entry guidelines** - Metadata Working Group

- File naming conventions / Folder structures:
  - **Best practices for file naming** - Stanford University Libraries
  - **Document Control Guidelines** - University of Edinburgh
  - **Best Practices for File-Naming** - Dept of Cultural Resources

- Open (non-proprietary) file formats
  [future proofing / digital preservation / interoperability]:
  - **Electronic Records Management Guidelines** - Minnesota Hist. Soc.
  - **Recommended Formats Statement** - Library of Congress
  - **Sustainability of Digital Formats** - Library of Congress

Good (digital) research practice – managing (y)our digital assets:

- **Metadata**:

"Metadata is the glue which links information and data across the world wide web. It is the tool that helps people to discover, manage, describe, preserve and build relationships with and between digital resources. […] It is the axis on which the wheels of the Internet turn. As users of digital resources it enables us to find what we are looking for (resource discovery metadata) or tell us what resources are (descriptive metadata). It might tell us where the resource has come from, who owns it and how it can be used (provenance and rights metadata). It might describe how the digital resource was created (technical metadata), how it is managed (administrative metadata) and how it can be kept into the future (preservation metadata). Or it might help us to relate and link this digital resource with other resources (structural metadata)."

Source: Jisc - Describing metadata

Good (digital) research practice – managing (y)our digital assets:

- **File naming conventions / Folder structures**:

"Naming conventions are rules which enable the titling of electronic and physical folders, documents and records in a consistent and logical way. This ensures that the correct records can be located, identified and retrieved from a filing system in a timely fashion, and that they are stored in an appropriate secure location. Ideally, the best time to think how to name and structure the documents and directories you create is at the start of a project. (…) Through consistency and the application of logical standards we benefit from secure storage, and the ability to locate and access information."

Source: Naming Files and Folders - University of Leicester

Good (digital) research practice – managing (y)our digital assets:

- **Open (non-proprietary) file formats**:

"When selecting file formats for archiving, the formats should ideally be:

- *Non-proprietary*
- *Unencrypted*
- *Uncompressed*
- *In common usage by the research community*
- *Adherent to an open, documented standard […]*
- *Interoperable among diverse platforms and applications*
- *Fully published and available royalty-free*
- *Fully and independently implementable by multiple software providers on multiple platforms without any intellectual property restrictions for necessary technology*
- *Developed and maintained by an open standards organization with a well-defined inclusive process for evolution of the standard."*

Source: Best practices for file formats - Stanford University Libraries

Putting (y)our research funding to good use:

Speak to us about:

Digitising existing (primary) research materials
(or: "*please don't go out and buy another scanner …*")

- DLS: Higher quality (experienced, trained staff)
- DLS: Greater cost-efficiencies (infrastructure: processes & equipment, including legacy / archival formats)

- **Creating your own 'born-digital' research materials**
- (or: "*please don't record in mp3, and call your files 'test1' …*")
  - hiring vs. buying AV recording equipment for interviews
  - Planning your outputs, i.e. managing your files from day one

Connecting your findings and your data

What is the Open Science Foundation (OSF)?

- The OSF is a *free* scholarly commons to connect the entire research cycle. It provides support with:
  - Creating and managing research projects
  - Sharing and collaborating on research projects
  - Version control, data security and privacy
  - Naming, organising and licensing files
  - Making data dictionaries
  - Creating DMPs
  - (…)
- OSF is developed by by the Centre for Open Science (COS). COS is a non-profit technology company providing free and open services to increase inclusivity and transparency of research. COS supports shifting incentives and practices to align more closely with scientific values.