
UCT
eResearch
ACCELERATING RESEARCH
eResearch report
2021-2022
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

UCT eResearch
ACCELERATING RESEARCH

Published by UCT eResearch
eresearch@uct.ac.za
www.eresearch.uct.ac.za

Production: Simon Communications
Writer: Natalie Simon
Design: Gaelen Pinnock

November 2022

CONTENTS

A thank you and a welcome

Introduction by Professor Sue Harrison, DVC Research and Internationalisation

UCT eResearch, as an entity, has existed at our institution for nearly ten years, during which it has provided support to countless researchers in the use of e-resources including, but not limited, to compute, software, data management and storage, digital scholarship and ensuring access and visibility of their research. It has ensured UCT's readiness to respond to changes in the research landscape as regards open science, among other things. Most of this work was done under the leadership of Dr Dale Peters, who, despite formally retiring in 2020, has continued to play an important role in UCT eResearch, including in this reporting year.

The role of eResearch director is a difficult one to fill, requiring insight into both the practice and the support and management of research across a range of specialist services to ensure the latter meet the needs of the former. As a rapidly developing, but new, field, expertise in this area is in short supply.

I am thus delighted to introduce Professor Mattia Vaccari, who has taken on this critical role at UCT as of September 2022. Many of our researchers working in the data-intensive space, particularly the astronomers and bioinformaticians, will already know Mattia, as prior to this role he served as the director of the Ilifu cloud computing facility. Ilifu, a big-data infrastructure for data-intensive research, has been featured in many previous editions of this report for its value as a regional research infrastructure that UCT eResearch helped to establish.

Alongside that, Mattia was the founding director of the University of the Western Cape's eResearch unit, a role he served in for three years. He is also an active researcher, an astrophysicist whose research focus is the formation and evolution of galaxies over the lifespan of the universe.

Mattia has demonstrated his ability to occupy this important space between research and research-supporting PASS (professional, administrative and support staff) departments to make sure services match needs. As the researchers and PASS staff reading these pages are aware, this is not an easy space to occupy. We wish Mattia much success in this role and look forward to working with him in a thriving eResearch environment while seeing where he will take the rapidly developing field of eResearch at UCT in the years to come.

I forward to seeing where he will take eResearch in the years to come.

Professor Sue Harrison
Deputy Vice Chancellor for Research and Internationalisation

Supporting the digital transformation of research

UCT eResearch was created as a partnership between ICTS (Information and Communication Technology Services), UCT Libraries and the Research Office, to help researchers across UCT redefine how their work can best be done in the era of big data and artificial intelligence.

This brief report details a number of ways in which the UCT community has been successfully working on the digital transformation of research, research data and research services, making them more sustainable and FAIR (findable, accessible, interoperable and reusable). While a lot remains to be done, and while the needs to be met by eResearch are ever-shifting, these stories demonstrate what can be achieved when digital services are effectively rolled out and employed.

One of our main challenges going forward will be to ensure the democratisation of the digital skills and tools required in the world today – not only in order for every researcher to carry out important research but also to enable every global citizen to make informed choices. Our data acquisition, processing, interpretation and visualisation services make eResearch a trusted partner for UCT researchers and their global footprint of collaborators.

As we look back and take stock of the major disruption of our research activities brought about by the COVID-19 pandemic, the crucial role of UCT eResearch becomes clear. Promoting and supporting the use of advanced information technologies to enable better, faster and higher-impact research is vital to the success of the research enterprise in an increasingly data-centric society.

While UCT accelerates its transformation to best meet the goals of its Vision 2030, UCT eResearch is looking forward to working with the university community to support impactful interdisciplinary research better and thus "Unleash Human Potential to Create a Fair and Just Society".

Professor Mattia Vaccari
eResearch director

What is eResearch?

UCT eResearch is a partnership between three research-supporting departments, the Research Office, Information and Communication Technology Services (ICTS) and UCT Libraries, with a focus on providing researchers with the individualised and customised support they need. The value of eResearch lies in its strategic position at the centre of the support services to ensure that researchers have seamless access to what the university has to offer.

The partnership

The diagram on the left represents the full scope of the eResearch partnership across the institution. However, the arrangement is fluid, changing in response to shifts in the research landscape, new policies and mandates and research needs.

Our eResearch support stretches across a range of offices and categories. In many instances several departments need to work together to provide a service; in some cases a single entity within a department provides a comprehensive service. But for researchers who do not know where to go, eResearch can be their first port of call.

A taste of what we do:

1

Governance

Support for a range of issues around legislation, compliance and data protection, particularly in the negotiation of research grants and contracts for international collaboration.

2

Compute

Advanced scientific computing, including high-performance computing (HPC) and cloud computing.

3

Research software

Support and guidance around the use of software for research, regulatory compliance and project management of research.

4

Advocacy

Outreach and training, such as the Emerging Researcher Programme seminar series and tailor-made seminars for research groups and departments and staying up-to-date with advances in technology and governance nationally and globally.

5

Research data storage

Long-term secure storage for research data.

6

Digital scholarship

Support for the use of digital methods of inquiry, research, publication and preservation to achieve open scholarship and research goals.

7

Research data management

Services related to the organisation, storage, preservation and sharing of data collected and used in a research project.

8

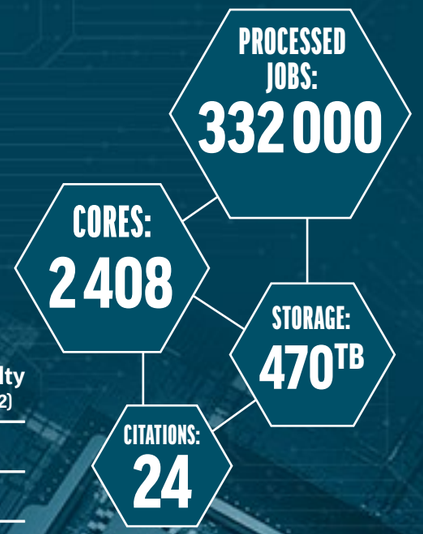
Access and visibility

Open-access publication of scholarly work to increase discoverability.

Section 2

High-performance computing at UCT

UCT's HPC facility provides a reliable, scalable and affordable computing service for the university's research community. It is geared towards researchers with major computing needs, but is also accessible to the full community, both for teaching and for research. These figures provide a snapshot of the HPC infrastructure and service in 2021–2022.



Faculty	New researchers	Active researchers	CPU hours Per faculty (1 July 2021 – 30 June 2022)
CHED	0	0	0
Commerce	9	6	33 924
EBE	31	48	3 703 973
Health Sciences	32	93	2 160 047
Humanities	0	1	54
Law	0	0	0
Science	53	84	2 183 313
TOTAL	**125**	**233**	**8 081 311**

Modelling drug dosages for African patients

The way that drugs interact in a human body is complex and dependent on a number of variables. Accurate and effective dosage relies on insight into how a particular drug works in a particular patient population. With the help of eResearch's HPC facilities, the UCT Pharmacometrics Group is working to understand how drugs should be dosed in our local populations..

"Most of the drugs we use in Africa have been developed elsewhere," says Associate Professor Paolo Denti, head of the Pharmacometrics Group at UCT. "This means that most of the information and clinical experience on their use is from developed countries, where clinical studies are generally conducted and where the drugs first become available after approval. The trouble is that, more often than not, when these drugs are rolled out in developing countries, the clinical context and the patient population are quite different and additional information is needed to use these drugs optimally. There is little to no financial incentive for research on this, especially for older drugs, whose patents may have expired."

Once a drug comes to market, there are generally still open questions as to what dose should be given to children or to pregnant women and how the drug can be used taking into account the common comorbidities of a region.

"In Africa and other developing countries, the demographics of people living with HIV are very different from, say, Western Europe and the US. There are more children or pregnant women living with HIV and, especially in Africa, tuberculosis is a common co-infection," says Denti. "It is very important to understand the interactions between HIV and TB drugs, as the TB treatment lasts several months, and a person living with HIV cannot simply stop taking their HIV drugs for that period. Unless the doses are suitably adjusted, the treatment may not work, thus exposing patients to toxic side effects or the risk of developing resistant HIV or TB strains."

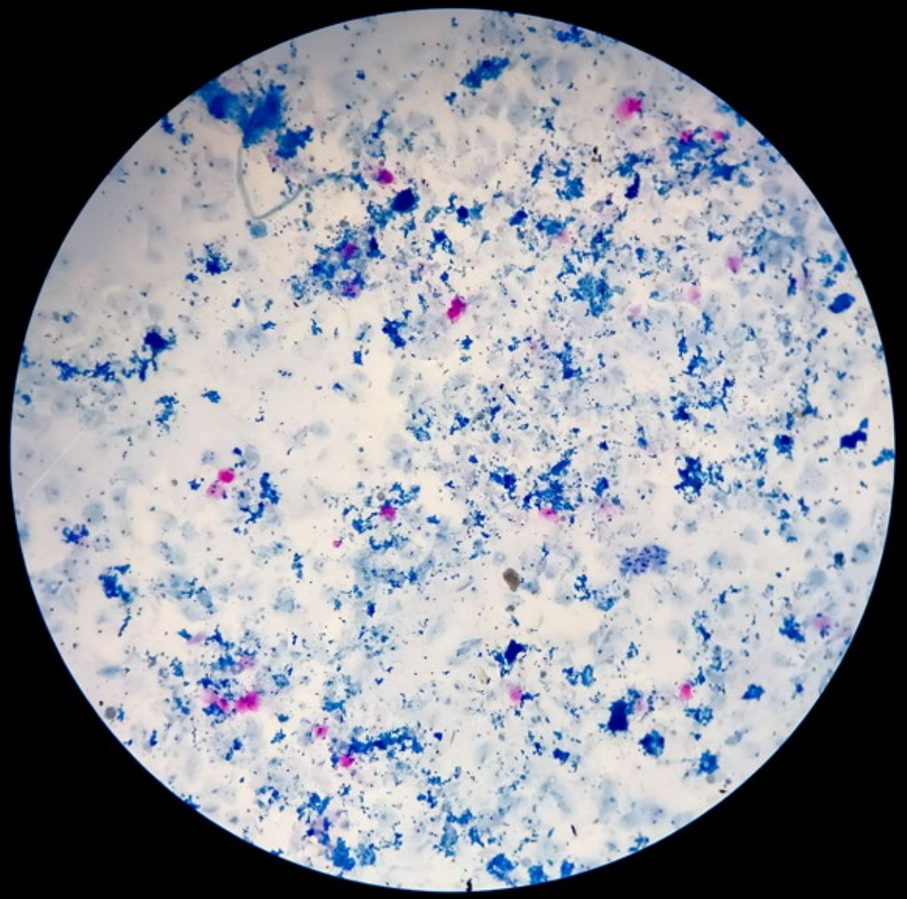
To make things more complicated, different people process drugs at different rates, and genetic differences are also relevant.

"Genetically, the African population is one of the most diverse in the world, yet it is also less well studied," says Dr Kamunkhwala Gausi, postdoctoral researcher in the group. "These genetic variations need to be taken into account when we calculate dosage, and it makes our work here on African populations all the more critical.".

Pharmacometrics and HPC

With so many factors in play during human-drug interactions, the data can be very noisy. Computational modelling using HPC is therefore critical to identify the signal through the noise.

> *"HPC allows us to take all our data and model it at one time,"* **says Gausi.** *"This is much more powerful than breaking it down by each variable and allows us to come up with a much more conclusive result."*



Denti notes that prior to using the HPC cluster at UCT, the group members were trying to run the models on their own laptops, with their own servers.

"Every second week something was breaking, security software had to be kept updated, and it was a nightmare to maintain," he says. "Migrating to the HPC cluster has been a game changer, freeing up so much time."

The HPC facilities also make capacity-building in these modelling skills easier for the group, as students can access the facilities from anywhere, through the remote server, and need no more than an entry-level laptop to run the models.

> *"These kinds of facilities allow us to build a specialised niche of pharmacometrics skills here in South Africa and become globally competitive in this space,"* **says Denti.**

> *"Unlike other research fields, which require prohibitively expensive equipment and often see poorer countries trailing behind developed ones, pharmacometrics only requires mathematical and statistical skills and computational power. The HPC facility has allowed us to unlock the potential of African pharmacometrics."*

Left: A tuberculosis sputum sample. Patients are asked to expectorate daily for a week and the bacterial load is then assessed to test efficacy of the drug dosage.

Modelling treatment for drug-resistant Tuberculosis

"Multidrug-resistant (MDR) TB is a major public health threat," says Denti. "If you don't properly treat and eradicate the TB bacterium in a patient with MDR TB, you run the risk of those bugs evolving into extreme drug-resistant (XDR) TB, which requires drugs that are more toxic and less effective, and then a patient is lucky if they survive it."

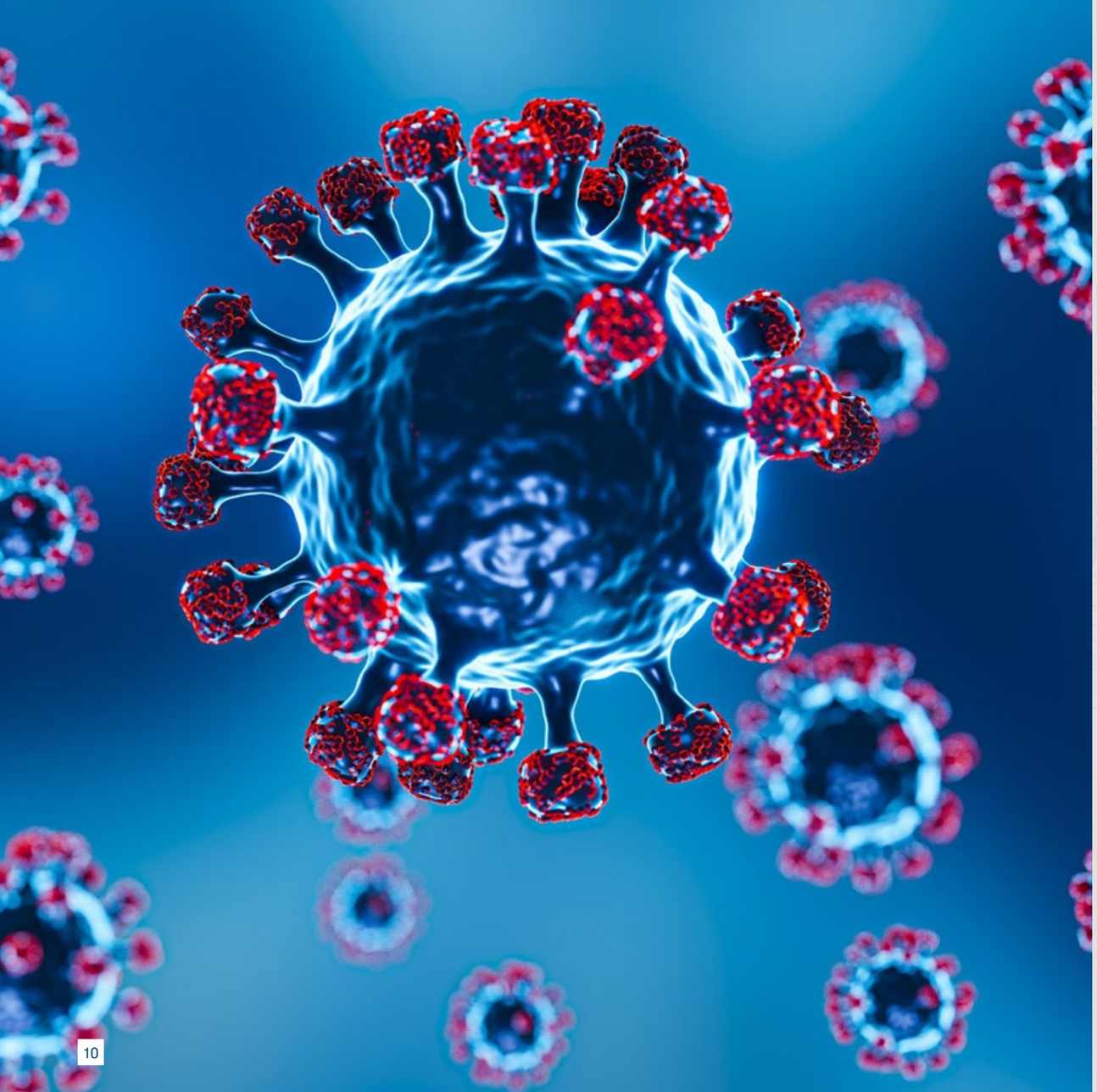
Gausi recently studied the effects of one common TB treatment, isoniazid, to see if higher doses of the drug could be used to effectively treat MDR TB. The study looked at data from a drug-sensitive TB control group and patients with an MDR TB strain that is less susceptible to isoniazid. Across both groups there were individuals with specific mutations, which indicates an accelerated clearing of the drug.

The data included daily tests from the patient groups for a week, to see if the bacterial load gradually decreased in response to the drug.

"The data was very noisy," says Gausi. "Using traditional statistics it would have been almost impossible to identify a signal among the noise. But, luckily, the modelling revealed that the bacteria were decreasing over the time period for all the patients, although MDR TB required higher drug concentrations and there were significant differences between fast and slow metabolisers of isoniazid."

The study allowed the group to offer guidelines on appropriate doses of isoniazid in MDR TB and for patients with the mutation under study.

"By using pharmacometric modelling, we were able to propose a personalised dosing strategy, accounting for the fact that different patients clear isoniazid at different speeds," says Denti. "This is expected to provide safer and more effective treatment in each patient."



Using HPC to better understand the impact of SARS-CoV-2 on the human body

A team of bioinformaticians used UCT's HPC facilities to study proteins in different organs of individuals with COVID-19. The findings have potential applications for better diagnostics and prognostics and for drug discovery and development.

SARS-CoV-2, the virus responsible for the COVID-19 pandemic, causes respiratory or gastrointestinal symptoms. However, a growing body of evidence shows that this virus is not confined to the human lungs but also affects other organs, including the heart, brain and kidneys.

Proteomics is the large-scale identification and quantification of proteins in biological samples. While proteomics has been used to understand SARS-CoV-2 pathogenesis, the data available is limited.

A team led by Dr Javan Okendo, a bioinformatician who completed his PhD at UCT and is currently a postdoctoral research fellow at the National Institutes of Health in the United States, used the eResearch HPC cluster to investigate the proteomic profile of different body samples. The samples included a gargle solution, urine and fluid from the lungs of individuals with different phases of severity of COVID-19. The team found that the body sites of individuals with the virus are characterised by different sets of proteins, which also differed according to the severity of the disease.

"These findings have potential applications for diagnostics and prognostics, personalised medicine and drug discovery and development," says Okendo.

> *"We could not have done this work without UCT's HPC facility, because it provided us with the scalable resources needed to allow analysis of the large amount of diverse and complex data at a reasonable speed. We are also grateful to the HPC staff, who made available the tools we needed in our research as modules, which expedited our data analysis."*

Left: Dr Javan Okendo, postdoctoral researcher at the National Institutes of Health in the US.

Far left: SARS CoV-2, the virus responsible for the COVID-19 pandemic. Scientists are still working to understand the impact of the virus on the human body.

Section 3

Geographic Information Systems at DLS

Geographic Information System (GIS) services is part of the Digital Library Service (DLS) offering for all UCT staff and postgraduate students. As the two case studies below demonstrate, the GIS team actively works with researchers to assist them in analysing geospatial data.

ArcGIS

LICENSES PROVIDED:

ARCGIS DESKTOP: **85**

ARCGIS PRO: **67**

PROVIDED ACCESS TO ONLINE TRAINING:

WEB COURSES: **398**

TRAINING SEMINARS: **11**

TRAINING VIDEOS: **47**

MOOCS: **27**

VIRTUAL CONSULTATIONS:

142

Too close for comfort?

Understanding the impact of conflict events on perceptions of government and democracy

UCT researcher Thomas Isbell is using GIS to pair Afrobarometer survey data with data from the Armed Conflict Location & Event Data (ACLED) project on events of conflict and insecurity in Africa. He aims to assess how exposure to these events affects citizens' attitudes to government, democracy and military rule.

Afrobarometer, a pan-African, non-partisan network conducting public attitude surveys in Africa, provides gold-standard survey data which covers about 75% of the African population. Surveys are conducted face to face, in the language of the interviewee's choice and in private, so as to allow participants to answer as openly and truthfully as possible. This survey data is publicly available, managed by the UCT Afrobarometer team and widely used for global research.

Dr Thomas Isbell, a postdoctoral research fellow at UCT's Institute for Democracy, Citizenship and Public Policy in Africa and a research assistant at Afrobarometer, is using Afrobarometer data, along with data from ACLED, to assess the effect of exposure to violent conflict on citizens' perceptions of their president, their military and their government's handling of the conflict. For this research he looked at Mali, Niger, Burkina Faso and Nigeria, all countries experiencing violent conflict.

To do this work, Isbell used the geospatial data provided by both Afrobarometer and ACLED on specific locations of events of conflict and insecurity. The GIS team at DLS helped Isbell to match the two sets of data to create what is referred to as a "near-map" – a map that shows where the different locations from the two different data sets fall relative to one another. This allowed him to easily match the Afrobarometer survey respondents' data with the conflict and instability event data, to measure how perceptions among citizens differ according to proximity to the events.

"The technically interesting part is that I computed different measures of exposure," says Isbell. "What does it mean to live in an area where there was one really big event as opposed to ongoing low-level conflict? And how do the survey respondents' perceptions change as one moves further away from the actual violence?"

While he is still unravelling the data, initial findings show that the further the survey respondents are from the actual armed conflict, and the lower their exposure is, the greater their support for the government and the military. But there is no impact on support for democracy, a finding which Isbell describes as surprising.

> *"Other than the findings, this research is valuable in that it leverages survey data with expert data,"* **explains Isbell.**

> *"Marrying local survey data with expert data is a really interesting way of cross-validating findings and getting a deeper insight into country-specific situations. GIS makes this kind of work very easily achievable."*

Deforestation in the Congo Basin

Postgraduate researcher Epie Njume is using GIS tools for his master's research in geomatics. The study looks at deforestation in the Congo Basin forest region in the last 20 years and its impacts on the environment in terms of air and water quality in the Cross-Sanaga-Bioko coastal forests.

Forest preservation, and particularly the conservation of the tropical forests of Africa, South America, and Asia, is critical to curbing climate change. Njume hails from Cameroon, which forms part of the area known as the Congo Basin, home to one of these planet-saving forests.

His master's research, which is focused on promoting sustainable forest management in the Cross-Sanaga-Bioko coastal forests, involves several layers. First, an assessment was carried out of the state of the forest from 2000 to 2021, which will provide the yearly rate of forest loss. Next, he will create a land-use-land-cover (LULC) classification map of the different LULC classes of the forest region, which will then be used to create a change detection analysis map to identify how the different LULC classes have been changing over the years. Furthermore, to test for the effects of forest loss on the environment, the LULC classification map will be used to test for correlation with air quality datasets, such as nitrogen dioxide, sulphur dioxide and carbon monoxide. This will determine whether there is a relationship between forest cover loss and air quality. The last step will be the prediction of future forest cover loss if no action is taken, in line with Sustainable Development Goal 15 (life on land).

Njume is using data captured by NASA satellites, which is made freely available, and several tools, including Google Earth Engine for the land use, land cover classification and forest loss analysis; ArcGIS Pro for change detection analysis; and a geospatial modelling system, TerrSet, for forest cover change modelling and forecasting.

Njume says the support of the GIS team has been a factor in making this research possible:

"The use of GIS tools and machine-learning models has led to the integration of multiple geo-spatial data for modelling, analysing, and predicting environmental phenomena. With the use of GIS tools, remote sensing data and machine-learning models, an awareness of the need for creating a sustainable forest environment in the region will be created."

Left: Man walking on logs that are being floated to a wood depot near Kinshasa, Democratic Republic of Congo, in May 2013. Photo: Flore de Preneuf / World Bank

Section 4

Showcasing UCT's collections

Ibali, the isiXhosa word for "story", is the digital collections showcase of UCT, hosted by UCT Libraries and managed by the Digital Library Service (DLS).

TOTAL ITEMS ADDED TO IBALI:
11 452

SITES LAUNCHED:
7

PAGE VIEWS:
98 312

VISITORS:
7 420

Bringing isiXhosa intellectual traditions into the academy

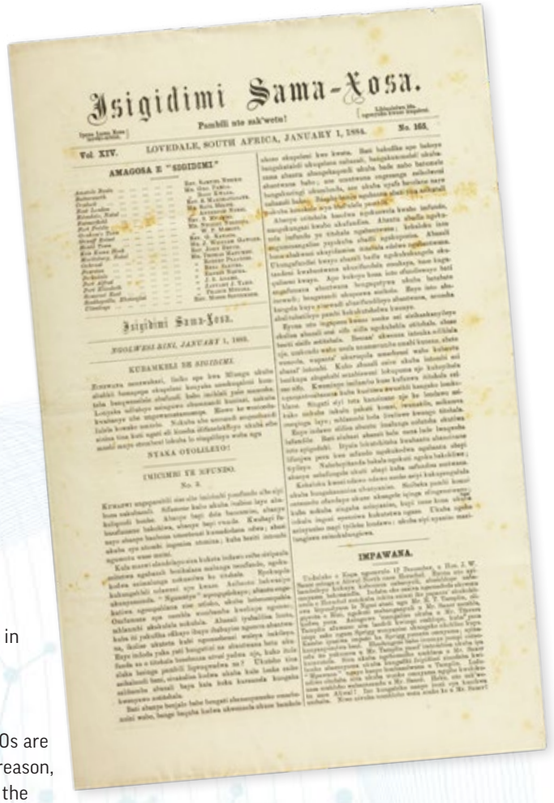
Dr Jacques de Wet of the Department of Sociology is leading a team to establish a digital archive on DLS's Ibali platform to make early isiXhosa publications authored by African intellectuals easily available for research. In this process, he has found the support provided by DLS to be invaluable.

De Wet, head of the isiXhosa Sociological Concepts and Intellectual Traditions Research Group, has been working with Amandla Ngwendu, from African Languages at UCT, and a team of isiXhosa-speaking students to digitise isiXhosa newspapers and books from the late 1800s and early 1900s. The goal of the archive is to make the writings of these African intellectuals available and "research ready" for contemporary research.

"Guided by the work of Neville Alexander on multilingualism and decolonising academia, I began to read Samuel Edward Krune Mqhayi's work in isiXhosa. I soon realised that if we are going to decolonise sociology and contribute meaningfully to the making of an African sociology, it is critically important that we engage with these and other texts written in the vernacular," says De Wet.

Part of the challenge of researching the writings of African intellectuals published at the turn of the 20th century was accessing their work.

De Wet explains: "These texts are scattered across numerous libraries in South Africa and they are in varying conditions."


Isigidimi Sama-Xosa.

Vol. XIV. LOVEDALE, SOUTH AFRICA, JANUARY 1, 1884. No. 165.

AMAGOA E "ISIGIDIMI"

Isigidimi Sama-Xosa.

NGOLWESI KINI, JANUARY 1, 1884.

KUBANIKELA NE ISIGIDIMI.

NYAKA OTOLILEYO!

UMCIMBI YE KPUNDO.

IMPAWANA.

Right: Copy of a cover of Isigidimi sama-Xosa (Eng trans: The Xhosa Messenger) was one of the earliest African-language newspapers in South Africa.

The books are easier to access, but isiXhosa newspapers from the 1800s are very fragile and, for that reason, many are not available to the public. They can only be viewed as static digital images.

Working with Ngwendu and his former student Dr Jonathan Schoots, now based at Stellenbosch University, De Wet recruited a team of isiXhosa-speaking students to build the IsiXhosa Intellectual Traditions (IsiXIT) Digital Archive, which is available on the Ibali platform. The archive includes PDF images and MS Word text files of early isiXhosa newspapers and books as well as additional information (metadata) about each publication and its contents.

> *"I need to trumpet the support of the DLS team in this project. I hate dealing with help desks and like to build a relationship with individuals committed to what we are trying to do,"* **says De Wet.** *"Sanjin Muftic of the DLS team provided exactly this support, working to understand our needs and accommodate them."*

Above: A cleansing ceremony undertaken by a Khoi-San group to mourn the fire that gutted the UCT African Studies library, home to an important indigenous archive.

Left: UCT stands on land once inhabited by the San and Khoi, and the San and Khoi Centre will foreground indigenous knowledge, rituals, languages and ways of knowing.

eResearch partners with Khoi and San Centre to build a platform for indigenous knowledge

The mandate of UCT's Khoi and San Centre is to support the knowledge production of the indigenous San and Khoi communities of South Africa. In order to achieve this, the centre has entered into a formal partnership with DLS to jointly build and maintain a platform for knowledge production.

The establishment of the Khoi and San Centre at UCT in 2020 was part of an important decolonisation initiative to foreground indigenous knowledge and support the knowledge production of the Khoi and San people, who lay claim to being the first inhabitants of the land that UCT is built on today.

> *"It has been a challenging journey,"* **says Shamila Abrahams, digital archive manager at the centre.** *"We are working hard to try and push the boundaries of the conceptualisation of knowledge – as we know it in its traditional university formula – to include the San and Khoi people's understanding of the term."*

She says it has not been an easy task to overcome the Khoi and San communities' mistrust of academic institutions, but the time and effort invested by the centre are beginning to pay off.

"We have been engaging with the San and Khoi communities to try and explain what we are doing and how we are framing it. While they were sceptical at first, they saw how we were struggling because we cannot achieve our mandate without them. Slowly our efforts began to gain traction and people started to come to us with ideas and plans."

The Ibali platform, hosted by DLS, will give the centre the medium to provide that voice to the communities. Much work has been underway in preparing the site, and it is set to be launched in 2023.

On the other end of the spectrum in the partnership, says Abrahams, is the support DLS provides. She says the memorandum of understanding (MoU) entered into with DLS means their responsibilities in the project are clear. She is confident that the technical side of the work will be taken care of so the academics are free to focus on their work, knowing that the knowledge they produce is being preserved and made accessible and discoverable to information seekers.

"The DLS team is amazing to work with," says Abrahams. "Their ethos is purely, 'How can we support what you are doing, and how can we push ourselves to achieve what you need to achieve?' Working with them has been a very refreshing experience."

Left: image taken at the cleansing ceremony to mourn the UCT fire.

Section 5

Growth of open data at UCT and leading the way with FAIR principles

It is becoming increasingly important for researchers to make their data, code, software and other research outputs accessible to all. While this practice, known as open science, is a powerful driver for transparency and accountability in scientific research and increased collaboration, some researchers have legitimate concerns about the protection of sensitive data and the risk of their openly published research being scooped by others.

Accessibility for all?

Some researchers also question how their research information can be organised effectively to make it not just openly available, but also easily found, understood, exchanged and cited. UCT eResearch provides a wide range of support and training for researchers grappling with issues around open science.

UCT, in line with best practices and mandated by many major funding bodies, supports the principles of FAIR data – research information that is as open as possible, and as closed as necessary. Even if you cannot make your data completely open access, practising good research data management helps you make your research more efficient.

Two systems support open science and FAIR data practice at UCT: ZivaHub and UCT DMP. The figures on the following spread show the uptake of these tools by our researchers from July 2021 to June 2022.

According to the FAIR principles, data should be:

Findable

This ensures that your data can be found by both humans and machines, by using a globally unique and persistent identifier

Accessible

Once someone has found your data, they need to know how they can get access to them. This could include going through an authorisation and/or authentication process – i.e. it does not have to be open access to be FAIR (ethics always trump openness).

Interoperable

The use of open formats ensures that your data can be integrated with other data and that they can be used by many applications or workflows for analysis, storage, and processing into the future, regardless of changes in software.

Reusable

This ensures that your data (and their related metadata) are openly licensed and well-described, indicating unambiguously how they may be reused without a need to contact the author(s) first.

ZivaHub
Open Data UCT

ZivaHub: Open Data UCT is an online institutional data repository that serves as a publishing and access platform for research data and scholarly outputs. It is powered by Figshare for Institutions and is available to all students and staff at UCT. Ziva is the Shona word for "knowledge".

TOTAL NEW VIEWS
627 750
Number of times ZivaHub content was viewed by internal and external users

22 091
TOTAL PUBLIC ITEMS

145
NEW DEPOSITORS

42 982
TOTAL DEPOSITS

NEW DOWNLOADS
368 671
Number of times items were downloaded

UCT DMP

UCT DMP is an online platform that assists with the preparation of a data management plan.

It offers various data management plan (DMP) templates that meet the requirements of different funding bodies and institutions, and includes detailed information on formulating your plan.

	NEW PLANS	NEW USERS	DOWN-LOADS	PLANS SHARED
JULY 2021	36	42	248	51
AUGUST 2021	35	44	239	51
SEPTEMBER 2021	53	47	304	37
OCTOBER 2021	61	44	169	42
NOVEMBER 2021	27	27	209	102
DECEMBER 2021	63	47	379	86
JANUARY 2022	317	367	1117	161
FEBRUARY 2022	255	313	1019	137
MARCH 2022	90	86	546	30
APRIL 2022	104	87	1014	25
MAY 2022	56	55	711	22
JUNE 2022	53	43	711	22

The importance of data management for sustainable research practices

Sudesh Sivarasu is a professor of biomedical engineering in the Department of Human Biology and director of the Biomedical Engineering Research Centre. He sits on the institution's Research Data Management Governance Committee. In this piece, Sivarasu speaks to the importance of data management in innovation, and how these lessons can be applied to all research.

In my field, biomedical engineering, our mantra is "safety and efficacy." Any new medical device must go through regulatory approval before it can be taken to market and rolled out to patients. And all regulatory bodies look at two things: first, safety, and then efficacy. The first thing the regulators need to understand is the risk of your device; then they look at how well it works. Thereafter, they have to do a risk-versus-benefit analysis. And they do this by carefully tracking the process of innovation through auditing the technical files, which are effectively the research data throughout the project lifecycle.

A lesson from data management requirements in innovation

Any innovator learns quickly that their data management must be meticulous, transparent and interoperable. The latter means the data can be processed and understood across various platforms, and this is necessary because regulatory bodies use different platforms. If the regulator finds flaws in the traceability of the data, the device will not be approved, and the project will need to be restarted from scratch.

> **The complete traceability of data, from origin to disposal, is the binding element of innovation; it is the key component in translating research into innovation.**

Innovators have learnt this diligence the hard way. But perhaps to address the reproducibility crisis in science, we need to apply the same stringent requirements to all research. Part of the challenge is that contemporary data management tools and requirements require a whole skill set, one that we are not necessarily equipping our students with.

Training students, training staff

In the Department of Human Biology, we have championed two things. The first is the critical importance of the data management plan in a postgraduate student's memorandum of understanding (MoU).

> **We now have our students submit their data management plan as part of their ethics application requirement. This ensures the students get training on effective data management, because if we don't teach them, who will?**

But it is also a critical risk management tool in a department where students often work with sensitive data. This ensures they know what to do with the data, and they are accountable if their data is mismanaged.

The second thing we are championing is data management training. Data management is not a one-size-fits-all solution. Some research groups use very large data sets, which require a particular set of skills; other groups are focused on clinical trials or public-facing research, which has its own set of challenges and requirements.

As a support service specialised to address the needs and requirements of data management across the institution, UCT eResearch, and its branches in various research-supporting PASS departments, remains an under-utilised service.

Assessing risk across the project lifecycle

As open-science practices around research data grow increasingly popular and mandated, a myriad of software packages has sprung up to support researchers. Mobile phones have become ubiquitous as data collection tools, with various mobile applications catering to different data-gathering needs. How often do we pause and investigate the security of the devices and applications we are using?

My research group learnt the hard way about the risks that come with open-source software in the early phases of the Russia–Ukraine war, when Russian hackers cleaned out many open-source databases, with catastrophic outcomes for many. We were affected, but fortunately we were not relying on any of these open-source databases for our medical devices. After this experience, we now develop our software in-house, minimising this risk.

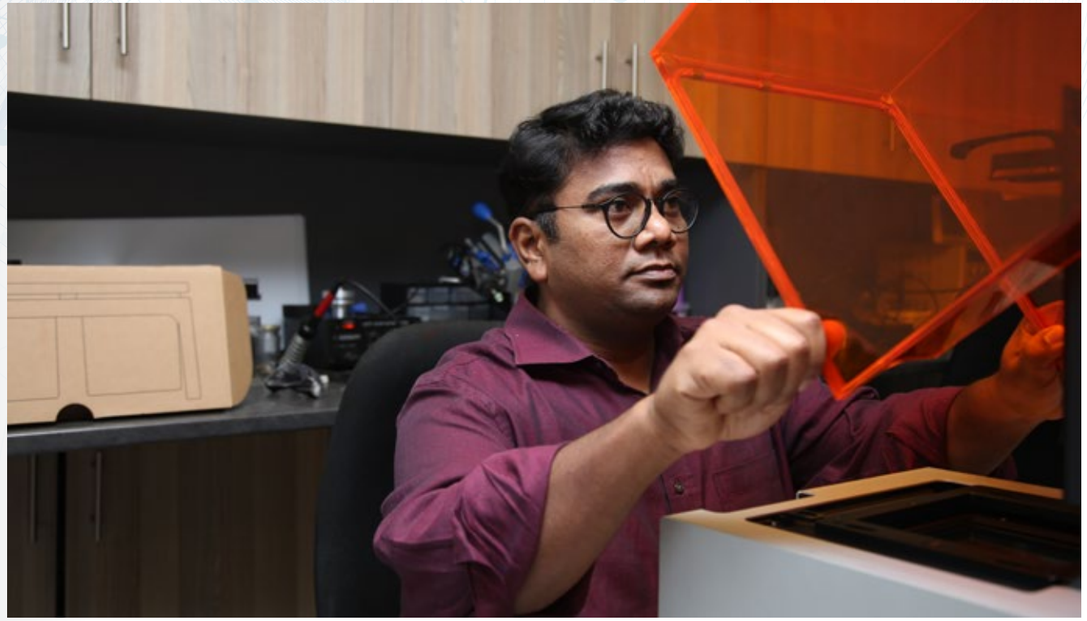
> **I encourage all researchers, but particularly those dealing with sensitive data, to investigate the security of your tools. As you set out, ask yourself: If something goes wrong and this data gets hacked, who is responsible?**

Project lifecycle close-out: Time for a policy around archiving and disposal

UCT has come a long way in our data governance policies and support tools. But there is one last gap that needs to be filled, and that is the archiving and disposal of data. We have no clear guidelines yet on how long to store data from a study or how to dispose of it when the time comes to do so. We also need to look at the succession planning of research data. When a researcher leaves or retires, do we simply lose their data with them?

This is critically important, but beyond the scope of an individual or even a department. It requires an institutional policy which includes proper archiving and data disposal guidelines.

We cannot look at research data in isolation; we need to look at the whole project lifecycle and beyond for the proper sustainability of our research.



Below: Professor Sudesh Sivarasu, professor of biomedical engineering in the Department of Human Biology.


UCT
eResearch
ACCELERATING RESEARCH