UCT eResearch

ACCELERATING RESEARCH

eResearch report

2019-2020

UNIVERSITY OF CAPE TOWN

IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

CONTENTS

Unleashing knowledge in and from Afrika through UCT eResearch

As the UCT community looks toward Vision 2030 and how to future-proof our institution, there is a growing acknowledgement that partnerships are key, not only between research collaborations, but also between academic and the professional and administrative support (PASS) staff. This will ensure that we are able to navigate the increasingly complex global research environment.

Key to the Vision 2030 goal of 'research solving Afrika's* problems that contribute to global knowledge' is the need for research to be supported by effective and innovative management.

International research collaboration is critical in addressing global challenges such as climate change, human migration, urbanisation, health, poverty, and inequality. This highlights the need for openness and transparency in working together towards the greater good of society – especially now, when African researchers are positioned to contribute vital information to understand, and potentially mitigate or solve global problems.

An increase in the scope and complexity of funder mandates and local and international regulations, have created a situation in which no department or research group can be an island. The demands of these mandates and regulations require that support units work in close and co-ordinated partnership with research groups. This ensures institutions are ready to comply with data governance requirements, so that they are firmly embedded in research collaborations. These include issues of data security and open science practices of data sharing and publication.

UCT eResearch has been one step ahead in terms of predicting and preparing for the coming changes and has been working to forge a governance framework based on systems and services drawn from PASS departments.

When our researchers at UCT find themselves navigating the demands of new data protection regulations and compliance, they are able to easily access an existing network of skills, expertise, and infrastructure, all coordinated by UCT eResearch. Given this professional support, UCT stands ready to "unleash knowledge in and from Afrika to redefine and co-create a sustainable global future."

* The Vision 2030 choice to spell Afrika in its pre-colonial spelling is an invitation to reclaim Afrika's agency and use it to validate the global character of the local in the 21st century.

Professor Sue Harrison
Deputy Vice Chancellor for Research and Internationalisation

Bidding farewell to eResearch Director Dr Dale Peters

The end of 2020 marks the end of an era for UCT eResearch, as Dr Dale Peters, who has served as director since 2016, retires. In her time in this post, Dale has kept a firm eye on the horizon, anticipating changes in the research landscape and ensuring UCT is prepared through the expanding eResearch portfolio. Her impact has spanned beyond UCT as she engaged with government and other institutions to lay the foundation for national support and infrastructure around data management and governance.

Dale has been a remarkable asset to UCT and her absence will be sorely felt. We wish her all the best in the next phase of her life.

— Professor Sue Harrison

Enabling a data-driven African research advantage

eResearch is your travel guide to a scientific expedition, exploring new territories in data-intensive research practice and probing new challenges in data analysis and management, presented at every opening vista of your career advancement.

The lessons of lockdown in this remarkable year present an opportunity to reflect on what it is to be a research-intensive university in our African context. Possibilities of Vision 2030 hang in the air, tangible in the recognition of a data-driven African research advantage. eResearch is set upon a course of facilitating accelerated local scholarship that contributes to global knowledge. As witnessed during lockdown, the upsurge in the demand for online tools for communication and collaboration has undoubtedly changed the course of scientific exploration, confirming the new normal as a viable model for the delivery of global contributions to knowledge. Reaching out to over a thousand researchers during the past year, the need is clear that in fostering inter- and transdisciplinary collaboration and borderless research, we must focus on optimising our data-management infrastructure and data-protection services, to allay concerns around data sharing and more equitable global partnerships.

Increased regulation and reduced funding have eroded trust in the "system", challenging eResearch to work towards transforming its research culture and community. Scientific communication, as evidenced in these pages, is an indication of that community, leading towards a new strategic direction in enhancing the ability of researchers to translate their research into environmental and social sustainability. In acknowledging the contributions of the eResearch team in accelerating research, we celebrate the success of our researchers, especially those venturing into new territories.

Dr Dale Peters
Director UCT eResearch

Streamlining our research support systems

The launch of the electronic Research Administration (eRA) system marks the transition from project to enterprise solution. The role of eResearch in providing user support for our researchers during this critical period equips UCT to manage and enhance its research profile as a research-intensive university.

UCT eResearch offers strategic oversight of the research infrastructure and, with the eRA roll-out, identifies the need for streamlined workflows. Our research infrastructure has been transformed from fragmented systems, repositories, and stand-alone data silos into a vertically and horizontally integrated research ecosystem, thus reducing the administrative burden on our researchers. Statutory reporting on research outcomes is much improved and has strengthened our research practice in open science, showcasing research excellence in improved visibility, quality, and integrity of outputs, thereby attracting potential partnerships and international collaboration.

Christina Pather
Acting Executive Director,
Research Office

Support across the research project lifecycle

It has been gratifying to see how the staff of UCT Libraries have embraced new ways of thinking and exploring new tools to enhance research support throughout the research project lifecycle.

New offerings, such as research landscape analysis, provide an invaluable service to our postgraduate researchers to help them survey a knowledge domain and identify the gaps they may pursue or fill with their current research.

Our research data management infrastructure and support continue to be exemplary and active in supporting FAIR (findable, accessible, interoperable and reusable) publication of data.

It is in the Libraries that the open access stance of the university is really brought to fruition, as it is the home of OpenUCT and ZivaHub, the institutional and data repositories.

As we continue to innovate and respond to the changing opportunities and needs of research, UCT Libraries actively contribute to the consolidation of existing knowledge and the creation of new knowledge.

Ujala Satgoor
Executive Director,
UCT Libraries

The technological foundation for globally competitive research

Many of the research stories in this report feature the face-to-face support offered by the Information and Communication Technology Services (ICTS) team to researchers, brokered through UCT eResearch and backed by the behind-the-scenes work of our technical specialists. A few large projects are underway which will further support the advancement and acceleration of research at UCT.

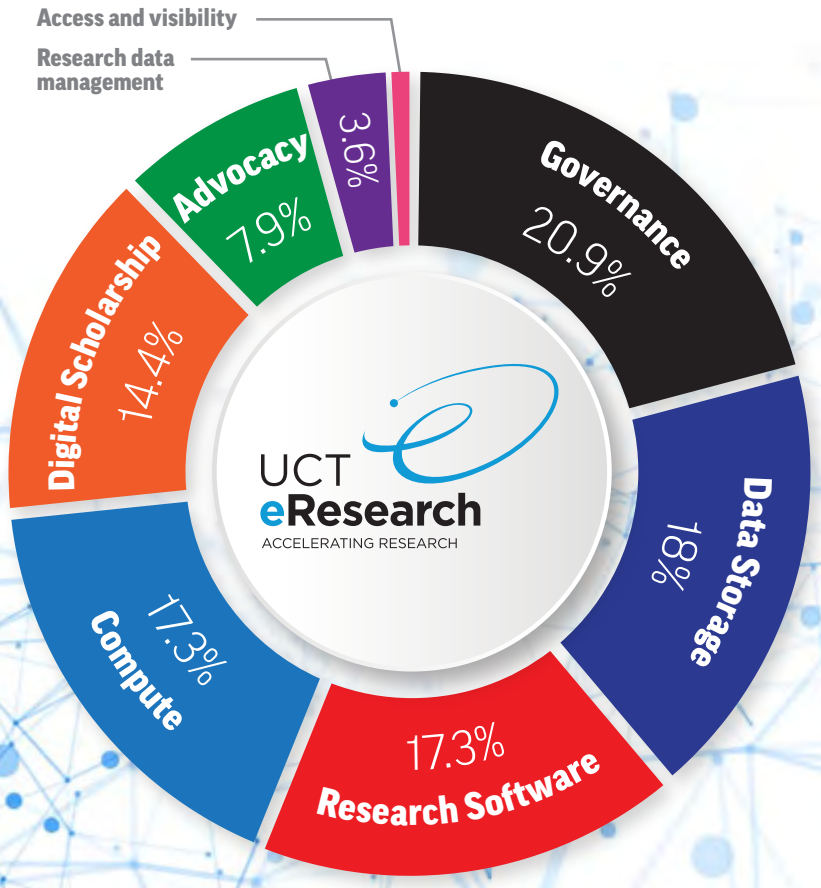
We have started our storage architecture project and are working with the data management team in UCT Libraries to provide the necessary secure storage for all institutional data, including research data. We have also begun the process of upgrading our upper campus data centre, which will give us redundant backup power and a greater ability to house and host computing facilities for research in a secure environment. In addition, our technical staff continue to provide support for the ilifu cloud computing facility for data-intensive research.

In these and other ways, ICTS provides the technological foundation underlying many of UCT's research support systems and services. This technological platform allows our researchers to remain globally competitive.

Richard van Huysteen
Executive Director,
Information and
Communication
Technology
Services
(ICTS)

eResearch support engagements

July 2019 to June 2020



1: Governance

Includes 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: Data storage

Long-term secure storage for research data.

3: Research software

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

4: Compute

Advanced scientific computing, including high-performance computing and cloud computing.

5: Digital scholarship

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

6: Advocacy

Including outreach and training, such as the Emerging Researcher Programme seminar series and tailor-made seminars for research groups and departments.

7: Research data management

This is the process which includes the organisation, storage, preservation and sharing of data collected and used in a research project.

8: Access and visibility

Includes open access publication of scholarly work to increase the discoverability and visibility of scholarly output at UCT.

Section 2

Advancing and accelerating research

In order to unleash knowledge in and from Afrika, in line with Vision 2030, we need to ensure our researchers have access to cutting-edge facilities, infrastructure and support.

Through this we can enable a data-driven Afrikan research advantage; accelerate innovation in local scholarship to contribute to global knowledge; foster inter- and transdisciplinary research; and enhance researchers' ability to translate their research into environmental and social sustainability.

Supporting fieldwork applications to understand the impact of land restitution

The South African Land Restitution Programme offers redress to South Africans and their descendants who were dispossessed of their property by the racially motivated legislation of the early 1900s. The Land Restitution Evaluation Survey (LRES) was established to evaluate the impact of the Land Restitution Programme on the economic, social and psychological outcomes experienced by the beneficiaries. This process involves fieldworkers traveling to remote areas all over the country to collect survey data. eResearch played a pivotal role in supporting the development of two novel survey applications, for in-field data collection and post-field sample management.

The Natives Land Act of 1913 was one of the darkest points in South Africa's history. This Act, and other related laws passed in the subsequent years, dispossessed over seven million black South Africans of their land in both rural and urban areas. In 1994 the Restitution of Land Rights Act (Act 22 of 1994) was passed. This Act entitles individuals and their descendants who were dispossessed of their land by racially motivated legal acts to file claims for restitution. The Land Restitution Programme seeks to undo some of the damage by offering a sizeable once-off cash or land transfer to dispossessed individuals or their descendants.

To better understand the potential impact of land compensation in lifting households out of poverty, and its effect on beneficiaries, the International Initiative for Impact Evaluation (3ie) and the Department of Rural Development and Land Reform (DRDLR), responsible for the Land Restitution Programme, commissioned a study by UCT's Southern African Labour and Development Research Unit (SALDRU). SALDRU then established the LRES, headed up by Professor Malcolm Keswell.

Supporting fieldwork virtually

To answer its research questions, LRES is conducting a longitudinal study. A nationally representative group of between 16 000 and 19 000 individuals will be surveyed twice over a period of five years. Many of the interview subjects live in remote locations, difficult for fieldworkers to access.

"With a survey at this scale we needed an effective way of tracking each interview to ensure it is done correctly the first time, as it would be too difficult and time-consuming for fieldworkers to go back to fill any gaps," says Jonathan Atkinson, senior operations manager for LRES. They also needed an effective way to keep track of the respondents over several years.

The solution to this lay in the development of two applications. The first is a case management system (CMS) which reports live fieldwork in progress. Once a fieldworker has completed an interview, they upload the questionnaire onto the survey software. It then goes to the server and is synced to a virtual machine, provided by UCT eResearch. Here it is converted into the correct format for importation into the CMS, where some basic checks are done to ensure that everything that needs to be captured is complete.

> **"Before the fieldworker drives away from the house, they can open the CMS and check that everything was uploaded and captured correctly. If the record is green, they are good to go; if the record is red, it means there is a problem. The fieldworker can then identify and fix the problem before they take leave of the respondent."**

~ *Timothy Brophy*
Data Quality Assurance & Database Admin

"So before the fieldworker drives away from the house, they can open the CMS and check that everything was uploaded and captured correctly," explains Timothy Brophy, responsible for data quality assurance and database administration at LRES.

"If the record is green, they are good to go; if the record is red, it means there is a problem. The fieldworker can then identify and fix the problem before they take leave of the respondent."

Given the success of the CMS, the LRES team decided to develop a second application: a sample management system (SMS). The SMS is designed to keep track of changes in respondents' contact details over time, an aspect which is particularly important in a longitudinal study. It also manages follow-up calls with respondents to check that everything was in order with the interviews. The SMS is where post-field data issues or questions can be flagged for follow up.

"Both the CMS and SMS are completely novel for this kind of field research," says Atkinson. "They are completely web-based, which makes them very accessible. Our researchers can be anywhere in the world and log in to look at the data, delegate tasks etc."

Developing the applications

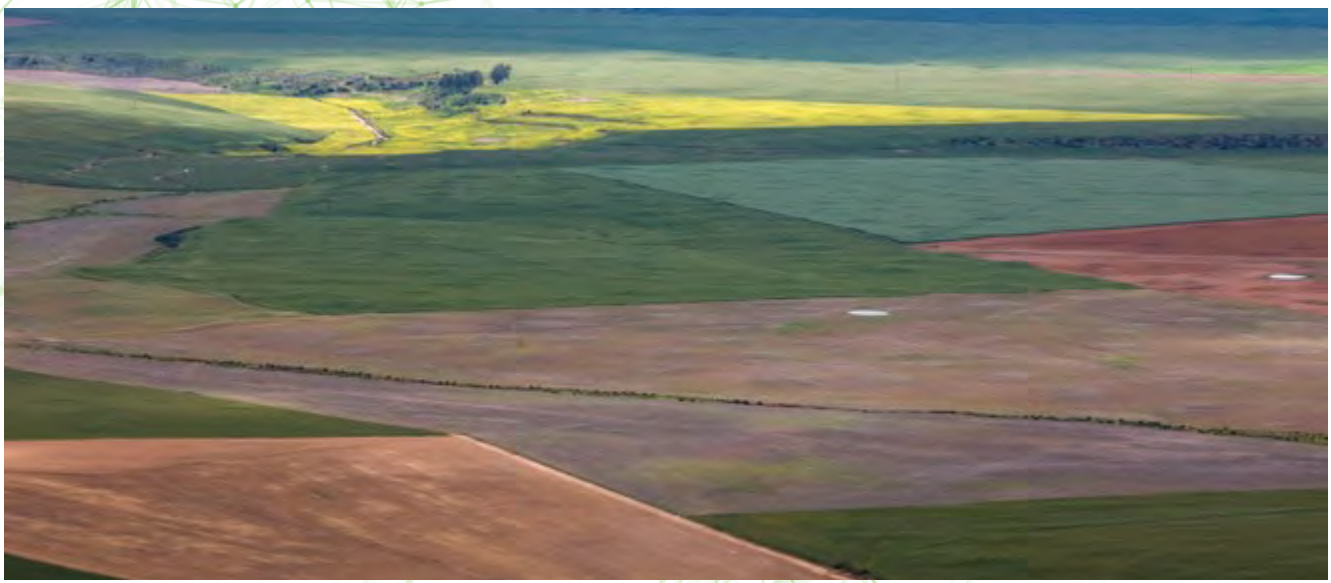
The development of the applications was done by an external development company called Busii. However, before any work was started, Atkinson reached out to UCT eResearch to support LRES through this process.

Pierre Neethling, senior manager of application services at Information and Communication Technology Services (ICTS), first set up meetings with the developers, ICTS representatives and LRES to talk through what each party could expect. Communication protocols were established and procedures agreed for ensuring that all the documentation regarding security was in place. Neethling and his team remained involved throughout the process, providing support where necessary.

> **"Where eResearch support was particularly invaluable was around managing the secure data through the application development process," says Brophy.**

Because of the highly sensitive nature of the data, the developers could not be given access to it. Yet they needed to design the application according to the data. On the advice of eResearch, the LRES team prepared manufactured (fake) data for the developers to use. Once the application had been designed, the eResearch team helped the LRES to set up a secure quality assurance server, which the developers did not have access to, where they could test the application using real data.

"This meant the LRES staff could test and break the application and go back with queries without compromising the data security or risk a buggy application that hadn't been adequately tested," says Brophy.



Above: The Native Land Act of 1913 allocated approximately 87% of South Africa's land to whites, including most of the fertile and productive lands, leaving the black population with marginal and relatively unproductive land. Image source: World Bank via Flickr (creative commons). Image on previous spread by Andrew Martin from Pixabay

Breaking out of a one-room mindset

Atkinson stresses that without the support of eResearch the LRES would not be where it is today in terms of infrastructure.

"When we began investigating hardware for this project, the standard way to go was to procure and manage our own, housed in a single server room," says Atkinson. "UCT eResearch, through their support in the provision of both servers and virtual machines, provided LRES with an organisational-level application which allows us to roll out our project at a very large scale."

In addition to the scaling up of infrastructure, the LRES team noted the importance of the expertise contributed by the eResearch team, particularly around the security of the data.

"From firewall protection to POPIA [Protection of Personal Information Act] recommendations, we relied heavily on eResearch support, because this is not our field of expertise," says Atkinson. "This kind of support allows us to innovate in ways we could not have otherwise done."

Contribution to global knowledge

"Understanding the impact of this kind of restorative justice aligns exactly with UCT's Vision 2030 goal: to conduct research solving Afrika's problems that contributes to global knowledge," says UCT eResearch Director, Dr Dale Peters. "A harsh history of colonialism and racist rule has left scars across the continent, and the Global South. Insight into whether the impact of this can be, to some extent undone, through such acts of restoration, is very relevant globally. It is great to know UCT eResearch, through a collaboration of valuable support staff departments, are supporting research in this kind of way."

"A harsh history of colonialism and racist rule has left scars across the continent, and the Global South. Insight into whether the impact of this can be, to some extent undone, through such acts of restoration, is very relevant globally."

Understanding the impact of COVID-19 and lockdown

On 26 March 2020, South Africa went into lockdown in response to the novel coronavirus pandemic. This meant all in-person data-collection activities ceased, at a time when policymakers most needed access to data to inform their decision making in an unpredictable environment. In response to this, a national consortium of 30 social science researchers from five South African universities was convened to conduct the National Income Dynamics Study – Coronavirus Rapid Mobile Survey (NIDS-CRAM). The study aims to collect, analyse and disseminate data on employment and the welfare of South Africans in light of the coronavirus pandemic and the associated national lockdowns. UCT eResearch assisted to ensure the security of this confidential data.

"Decision making is only as good as the data on which it is based," writes Dr Nic Spaull in his foreword and introduction to the first NIDS-CRAM report. "The local and international landscape is constantly morphing and changing in unpredictable ways making policy formulation and implementation as hard as it can possibly be."

This is why a consortium of leading economists and researchers across the country recognised the importance of continuing data collection to monitor the South African economy throughout the various levels of lockdown. This consortium sprang into action to put together a questionnaire for telephonic interviews to be conducted on an ongoing basis during the lockdown period.

The next big challenge was how to gather a representative pool of respondents. The Southern Africa Labour and Development Research Unit (SALDRU), through which NIDS was run, received the permission of the South African Department of Planning, Monitoring and Evaluation (which owns the intellectual property of NIDS) to use the NIDS sample group for this study.

Data collection, data quality and ensuring confidentiality

Once it was agreed to use the NIDS sample, SALDRU became more heavily involved in the project, with Associate Professor Reza Daniels overseeing the SALDRU data collection and data quality team for NIDS-CRAM; headed up by data collection and data quality heads, Timothy Brophy and Kim Ingle.

As the telephonic questionnaires were completed, it became necessary to conduct data quality checks. In addition to high-frequency data quality assurance checks, it was agreed that secondary quality assurance checks should be done by listening to randomly selected calls to check whether the data had been transcribed correctly and that the interviewers were conducting themselves professionally.

As part of the study's informed consent procedure, the respondents gave permission for the calls to be recorded, but the content of the calls remained

Below: President Cyril Ramaphosa addresses the nation on developments in risk-adjusted strategy to manage the spread of Coronavirus COVID-19 (July 2020). Image courtesy of GovernmentZA via Flickr

Above: Screenshot of some of the panellists of the webinar launch of the NIDS-CRAM Wave 1 research results and data. Photo: Haajirah Esau.

highly sensitive and confidential. Before the SALDRU team shared the NIDS-CRAM transcribed data – even with NIDS-CRAM partners – it had to be anonymised to protect the identities of the respondents.

The survey is nationally representative, which means interviews are done in all 11 official languages. To conduct the data quality checks in all those languages, the study chose to work with a trusted third-party contractor with a very strong team of translators, but they needed to protect the confidentiality of the data.

To deal with this issue, the SALDRU team reached out to eResearch.

"Even though the external team members all signed Protection of Personal Information (POPI) compliant non-disclosure agreements, there was still an issue of how to share the recordings with them in a secure manner to ensure this data remains confidential," says Ingle, who worked directly with UCT eResearch on this project.

Ashley Rustin from eResearch set up a secure folder which could be accessed by the designated external team members. They worked only within that folder, so no data was to be stored locally on the devices of the contractor, and the security of the study was maintained.

> **"UCT eResearch really understood the importance of having the correct security protocols in place," says Brophy. "This means the team members listening to the calls are doing so in a secure and encrypted fashion."**

The future of NIDS-CRAM

The first wave of the NIDS-CRAM survey made headlines across the country when the data was released on 15 July 2020 and played an important role in influencing policy-making in government, civil society and business. The anonymised data is openly available in the DataFirst Open Data Portal. The next wave of interviews has already been completed and further waves for 2020, and even early 2021, are in the pipeline.

SALDRU looks forward to continuing to work with eResearch for all future NIDS-CRAM data production efforts.

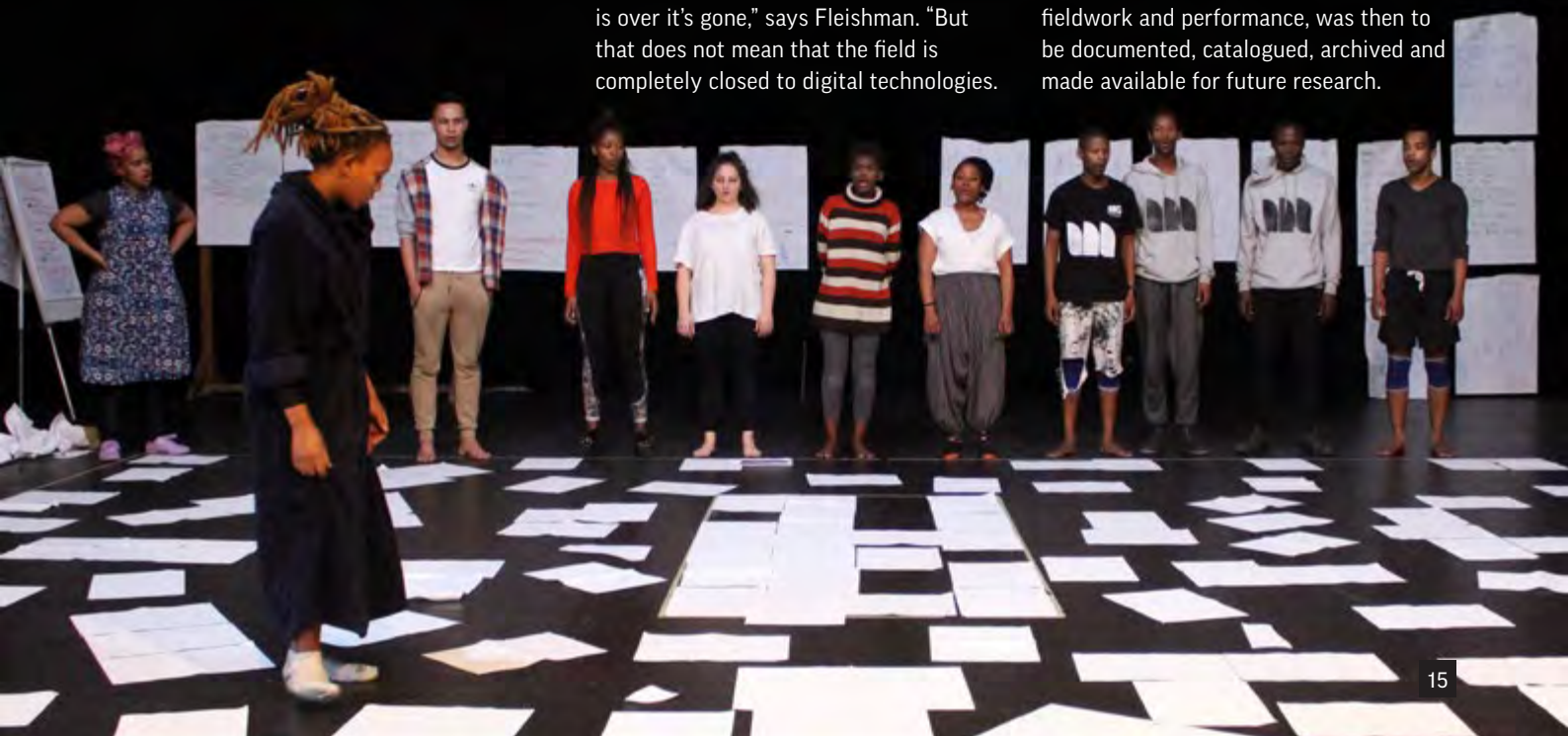
Grappling with data in the performing arts

Professor Mark Fleishman of the Centre for Theatre, Dance and Performance Studies and his team are exploring what data means in the performing arts. They are working with Digital Library Services to catalogue, archive and make available data that, among other things, documents the process of performance, pushing the boundaries of research in this field. The outcome of this may prove a useful response to the 2017 policy of the Department of Higher Education and Training on the recognition and reward of creative works and innovations.

A greater focus on research data as a scholarly output has prompted researchers in the performing arts to consider how to apply this to their work. "The argument has always been that theatre is ephemeral; when a performance is over it's gone," says Fleishman. "But that does not mean that the field is completely closed to digital technologies. It is just about figuring out how to use them in a helpful way."

In 2019 when Fleishman received a grant from the Andrew W. Mellon Foundation for a five-year funded research project on Tragedy in Africa, he made sure that the mechanisms for digital archiving and data collection were built in from the start. The project, called Reimagining Tragedy from Africa and the Global South (ReTaGS), examines the popularity of the ancient Greek tragedies of the fifth century BC in post-colonial Africa, and the reasons for this. The project is made up of a number of parts. One is a research fieldwork exercise, which involves archiving productions of Greek tragedies performed on the African continent in the 20th century; another element is to actually make theatre in response to that classical canon and see what the practice brings up.

The whole research project, including fieldwork and performance, was then to be documented, catalogued, archived and made available for future research.



"But along with this came the questions: What kind of data is generated from creative practices? How can that data be stored and made available, and to whom and under what circumstances?" says Fleishman. In ongoing engagements with the Digital Library Services department, the need for a dedicated data steward quickly became apparent.

Bringing on a data steward

Jayne Batzofin, who had, as part of her master's programme, completed a project documenting the Early Years Theatre work of Magnet Theatre, was brought onto the project. Originally, Batzofin was supposed to join the project for a period of a month. However, as the scale of the task of documenting, cataloguing and archiving the work became apparent – as well as the relevance of the work to the performing arts – she was asked to continue to develop and oversee the data management flow of the project.

"In many ways this data management element of the project grew into something separate from the particularities of the ReTaGS project," says Fleishman. "This aspect is almost like a meta-research strand dealing with how to do research in the future."

Documenting the process

Historically, in the performing arts, archiving has been about product and outcome, but for ReTaGS the group made the decision to document the process as well as the product. Batzofin was brought onto the ReTaGS project just as the rehearsals for the first theatre production of a Greek tragedy set in Africa, titled *Antigone: (not quite/quiet)*, began. Batzofin's strategy was to record the entire process as it played out.

"I came in and recorded hours of rehearsals every day," explains Batzofin.

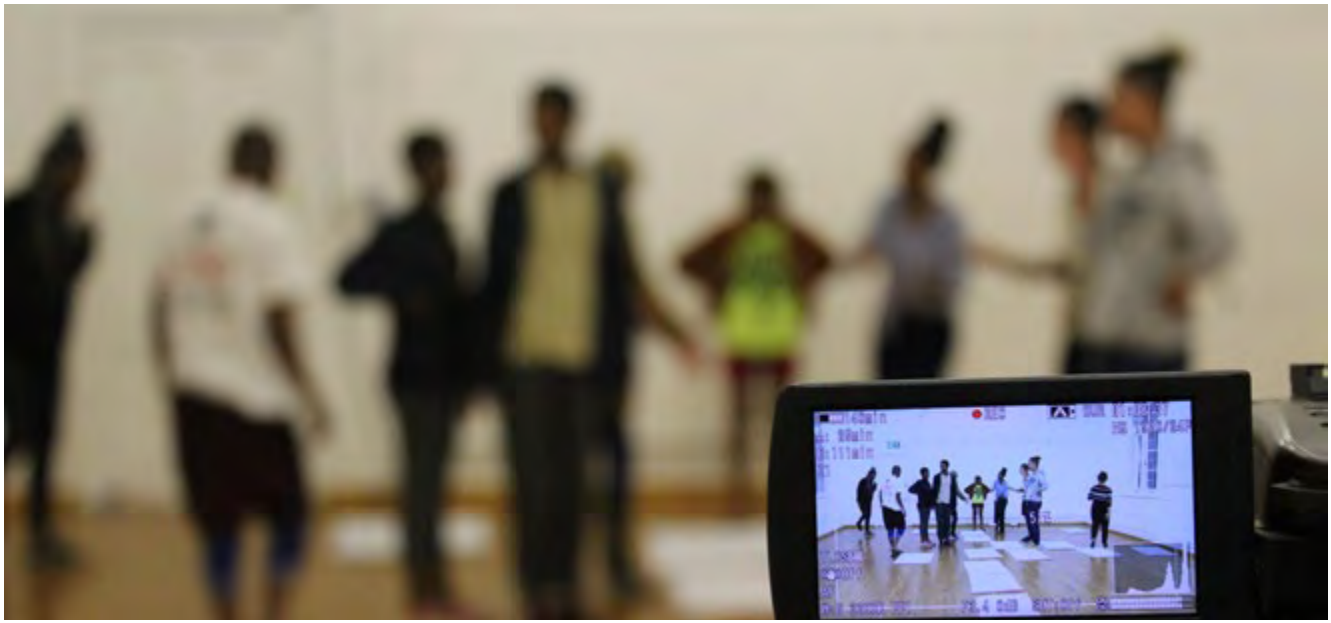
The idea behind recording this process was not to achieve an immediate utilitarian purpose, but with an understanding that this record is likely to emerge as being of value in the long run, in ways that cannot be imagined now.

Cataloguing and archiving the collection

Once Batzofin had captured – as a start – hundreds of hours of footage of rehearsals and performances, with additional plans to include recordings of field research and three more theatre productions, questions began to arise about how this collection would be archived, catalogued and curated. Early on in the process, Batzofin had reached out to Digital Library Services (DLS) for support with various elements of data management, including selecting the most

Right: Rehearsals for Antigone (Not quite/quiet) image by Jayne Batzofin. **Previous page and opposite:** All researsals were recorded by Batzofin as part of the data collection process .

Historically, in the performing arts, archiving has been about product and outcome, but for ReTaGS the group made the decision to document the process as well as the product.



suitable file formats, understanding file structures and, most importantly, working with complex, descriptive metadata.

"Metadata is the description of the data and is key to the successful cataloguing of work," explains Batzofin.

Batzofin's goal was to enable future researchers to access the collection and type in keywords, such as "lament", in order to find only videos linked to those keywords.

A big question was where the ReTaGS collection could be housed. When Batzofin reached out to DLS, they were in the process of developing a university-wide showcase website for digital collections. The site is called Ibali and is powered by Omeka S.

"The collaboration was very serendipitous," says Sanjin Muftic, digital scholarship specialist at DLS. "ReTaGS needed a site and we needed a complex collection to really demonstrate how the Omeka S platform works."

Currently, a curated version of the ReTaGS collection is available on Ibali, including the field research and the first production. As the project proceeds, this collection will continue to grow as a resource for future researchers.

Conclusion

"This has been a very interesting project in terms of how it uses digital technologies to capture the process of making, so that process can also be used for academic research," says Fleishman.

Where this project is also of great relevance is in relation to the subsidy for creative works instituted by the Department of Higher Education and Training (DHET), for which peer review of the creative output is required.

"There have been big debates about how peer review can be feasible for live performances," says Fleishman. "What this project potentially opens up is the possibility of a much more sophisticated archiving of a work, which goes beyond just the final recording of the project but also engages with its processes and theory. And that might be a mechanism that makes research review much more feasible and robust."

High-performance computing and neuroimaging

Brain imaging methods, known as neuroimaging, allow researchers to study the structure and function of our brains in a way that is non-invasive and insightful. Dr Jonathan Ipser, senior research officer in neuroimaging and co-chair of the Psychiatry Neuroimaging Group (PNG), has been working with UCT eResearch to accelerate his own research on the impact of substance abuse and HIV on the brain.

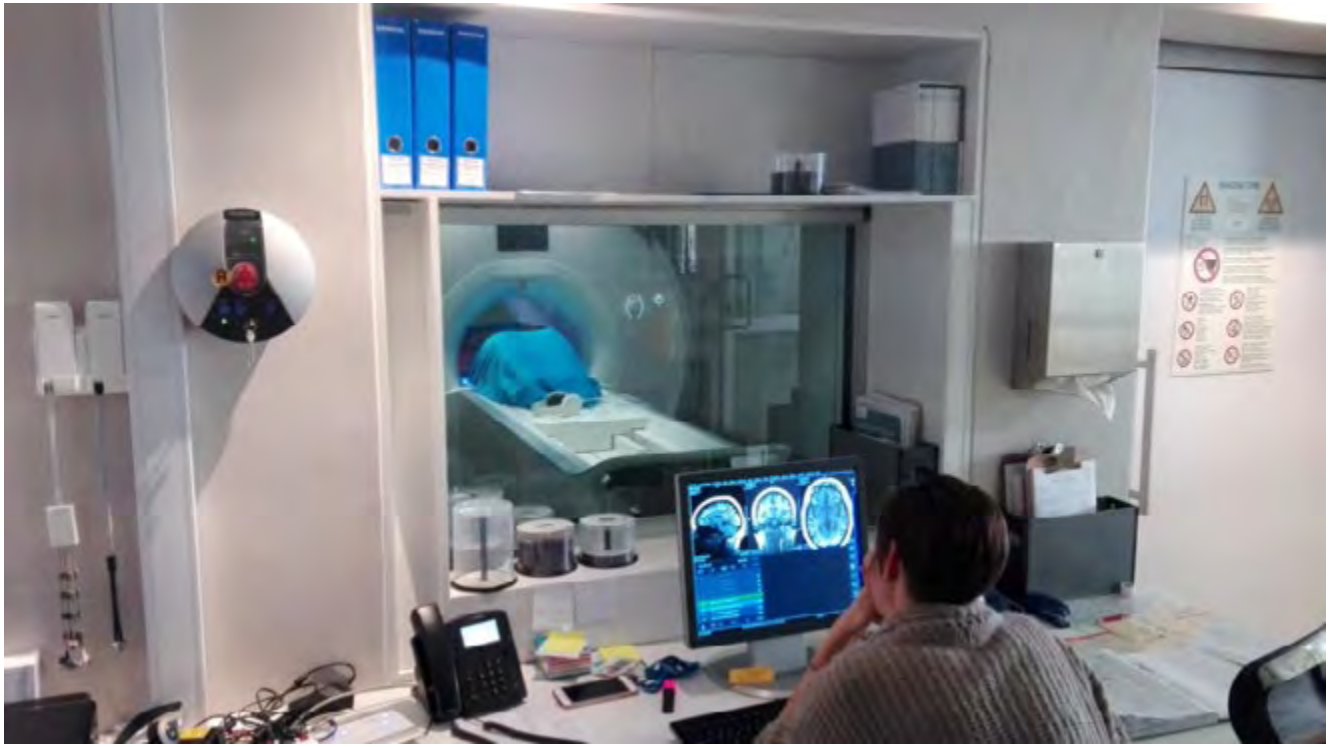
Our brains are not immune from the impact of disease and substance abuse, but the effects of these on the brain are perhaps not as well understood as the effects they have on our bodies. Ipser has been using neuroimaging techniques, specifically structural magnetic resonance imaging (sMRI) and functional magnetic resonance imaging (fMRI), to see how our brains are affected by, among other things, HIV and substance abuse, particularly the abuse of methamphetamine.

"My particular field of expertise is known as resting-state fMRI, where we look at activity in the brain when a person is not engaged in any particular task," says Ipser. "I will capture brain activity for anything from five to 15 minutes, and in that time we ask the participant to just relax while we look at connectivity within the brain," he explains. "By identifying areas that activate at the same time, we can see which brain regions are functionally connected."

His research has been focused on whether particular factors, such as substance abuse or HIV, are associated with differences in this connectivity, and if a normalisation of these abnormal patterns can be observed with treatment.

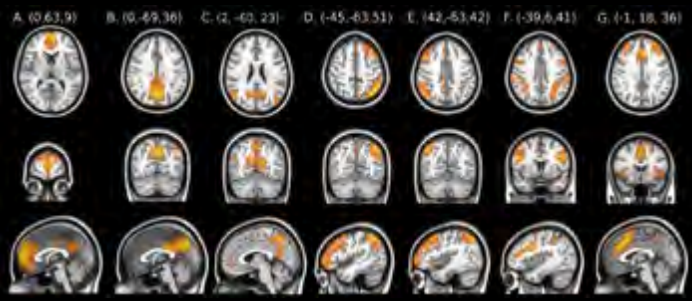
eResearch seminars

"Neuroimaging is a very data-intensive field," says Ipser. "Every brain scan creates at least a few gigs of data



More about the PNG

The PNG is the neuroimaging arm of the Brain and Behaviour Unit at UCT, a collaborative hub of psychiatric neuroscience research headed by the HoD of the Psychiatry department, Professor Dan Stein, that focuses on work particularly relevant to the South African and African contexts. The PNG's core team includes investigators with expertise in a variety of neuroimaging techniques and is led by Dr Jonathan Ipser and Dr Nynke Groenewald.



Above: Seven cognitive control resting-state connectivity images. From left to right: anterior default mode network (DMN), (B) the posterior DMN, (C) the precuneus DMN, (D) the left frontoparietal network, (E) the right frontoparietal network, (F) the dorsal attention network, and (G), the salience network. **Image on left page:** Cubic Scan room: Neuroimaging scanner. Both images provided by Jonathan Ipser

per participant. If you have dozens of participants scanned regularly over a period of several years, it quickly turns into a big data set."

This is why, when Ipser saw that UCT eResearch was offering seminars for departments across campus, he was quick to sign up the PNG.

> **"We have a number of challenges around data, including storage, processing and archiving," he says. "For years now I have also been interested in issues of research reproducibility and making data more reusable for researchers down the road."**

In response to Ipser's request UCT eResearch tailored a six-session eResearch series over three months to explore the range of solutions to these neuroimaging-specific data challenges.

Data processing with high-performance computing

Immediately after one of the seminars, Ipser reached out to the high-performance computing (HPC) team at Information and Communication Technology Services (ICTS) and asked them to install specific software for processing fMRI data on the HPC cluster. He has since been using the HPC facilities to process hundreds of data sets at a much faster rate than before.

"On my own workstation it would take about 10 hours to process one patient scan, whereas on the HPC facilities I can process at least 14 scans in a 24-hour period. This has been particularly useful in processing large datasets for global collaborations with which we actively participate, such as the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium."

What Ipser also appreciated was the support with using new tools such as the Singularity and Docker software containers.

"These allow you to install the particular software package you are interested in, as well as the entire software environment, in one go," he says. "This means all the packages that your programme is dependent on are installed at once, and you are always using the most up-to-date software."

He adds that the HPC team are very responsive, and that working with them has afforded opportunities to experiment with different software options in a supported environment.

Working with HPC is just the beginning of the relationship between UCT eResearch and the PNG. Ipser says his group is likely to request further support throughout the research project lifecycle. From data management plans to storage and archiving, the group appreciates the value of eResearch's services to their work.

Making drone data FAIR

The use of drones to capture data is increasingly common across a range of disciplines, particularly the physical and environmental sciences. While this data is valuable for reuse, its provenance is particularly complex.

In collaboration with UCT eResearch and the University of Notre Dame's Center for Research Computing, Dr Jane Wyngaard of UCT's Department of Electrical Engineering is working to build an open-source data toolkit called LANDRS (Linked-data API for Networked Drones) for capturing and publishing drone-captured data.

In conjunction with other tools being developed – such as the Australian Scalable Drone Cloud – LANDRS tools will allow researchers to process and publish their drone data more easily and in a way that fully captures and exposes the data's provenance, thus helping to make it findable, accessible, interoperable and reusable (FAIR).

Increasingly, researchers, including oceanographers, climatologists, and ecologists, are collecting valuable data 120 metres above the planet, to improve both our monitoring and understanding of the climate and environmental processes. Drone-data collection is adding a valuable new scope to so-called 'small science', explains Wyngaard.

"Individual researchers, who have collected enough funding to buy a drone, generally lack data management support. They are capturing valuable data all over the planet but that data often goes no further than that single researcher and their immediate work."

"The problem," she says, "is that the provenance pipeline for drone data is particularly complicated for scientifically-robust reuse of the data, even within the same discipline."

Wyngaard's goal is to provide researchers who do not have access to extensive data curation support, with the tools needed to publish their drone-captured data with sufficient metadata for its reuse, even across disciplines. An added complication to this challenge is the need to align discipline-specific ontologies. Different disciplines use different words to describe different and even the same parameters or phenomena. In order to fuse data from different domains, an understanding is needed as to how these terms relate to one another.

The use of formal ontologies, in combination with emerging semantic web technologies, can make this process automatable. The toolkit takes advantage of this possibility providing automated data annotation that archives and publishes metadata that is readable by both machine and human. This in turn can be used to make data more FAIR at the cost of minimal manual intervention, despite the complexity of the drone data pipeline.

Australia's Scalable Drone Cloud

Wyngaard also enjoys a long-standing relationship with a group of Australian researchers, who, under the umbrella of the Australian Research Data Commons, are working to build a national ecosystem for drone-data management called Australia's Scalable Drone Cloud. A visit to Australia in early 2020 by UCT's Vice-Chancellor, Professor Mamokgethi Phakeng, and Deputy Vice-Chancellor for Research and Internationalisation, Professor Sue Harrison, resulted in plans for a more formal research collaboration between the two groups.

Connecting with UCT eResearch

The increase in rate and volume of data being captured by researchers in general, in combination with advances in data technologies, societal need, and a drive from funders to publish data, has sparked a global trend of eResearch facilities pairing with science experts to improve data publishing in general. Most recently, this has included eResearch facilities seeking to also support science drone users. Wyngaard has thus been collaborating with UCT eResearch Director, Dr Dale Peters, to support drone research at UCT. The goal of this collaboration is to support UCT researchers using drones by finding ways of deploying the LANDRS tool stacks developed by Wyngaard and the Australian Scalable Drone Cloud.

The end goal, explains Wyngaard, is to create an end-to-end support infrastructure for small-scale drone research, by combining the data publication pipeline she is building, with the drone-data cloud-processing pipeline the Australian group is building with eResearch support.

Driving the fourth industrial revolution with *the internet of things* in smart agriculture

The internet of things (IoT) is one of the key drivers of the fourth industrial revolution, connecting smart devices into a closed network which uses artificial intelligence to optimise its own efficiency. Bontle Mere, a final year student in the Department of Electrical Engineering, explored the application of IoT in agriculture for her final year research project in Engineering in Mechatronics. To do this she used a virtual machine in the ilifu cloud computing facility, a regional data-intensive research facility supported by UCT eResearch.

Technological breakthroughs around the manufacturing of data-collecting sensors is driving new approaches to improving agricultural efficiencies via data mining. These breakthroughs include making them low power consumers, small in size, and cheap to make.

IoT refers to the increasingly common practice of connecting normally non-networked devices to the internet thus enabling industries to capture terabytes of data about their processes. In many cases, by applying both traditional data processing techniques and increasingly machine-learning-based processes to this data, significant efficiency gains can be achieved via optimised resource use, early warning and detection of potential problems, and the detection of trends or patterns that were otherwise invisible. Both commercial and small-scale agriculture are rapidly adopting such techniques in so-called smart agriculture, given the significant potential gains in crop and livestock care including watering, soil nutrient control, disease and pest detection, and harvest timing.

IoT to drive smart agriculture

Mere, under the supervision of Dr Jane Wyngaard in the Department of Electrical Engineering, focused her final year research project on smart agriculture.

"It is fascinating to me how there is still so much potential to use technology to advance how we farm," says Mere. "I wanted to explore that; particularly how small-scale farmers can efficiently use different technologies to make their work easier."

There already exists a myriad of smart devices that capture and process data for various farming applications but in order to create an IoT for agriculture, these devices all need to work together.

"The trouble is", says Mere, "they don't. They are all designed by different companies or even by farmers themselves, using different standards of communication and with different electrical and mechanical requirements. These differences mean the devices are not interoperable and often cannot connect to one another."

Wyngaard thus tasked Mere with building a reference specification document to define a standard for designing and deploying agricultural devices, which will allow for these devices to work together to share data.

"This reference specification creates a guideline for farmers and companies to build devices that can work together, reducing development time and cost and hopefully increasing innovation and collaboration among farmers across the world," says Mere.

This document is available under an open-source licence and a demonstration of its implementation was created using

open-source tools. By making both the standard and demonstrator open-source, they are available to everyone at no cost and are fully accessible for someone to take and modify or add further enhancements for the benefit of all.

In demonstrating her specification, Mere used soil moisture and air temperature sensors, and sent their data to a virtual machine she was using as a server, housed in the ilifu cloud.

"I had such a great experience using the ilifu facility," says Mere. "The computational power was incredible, and the team at ilifu were always responsive and willing to help. Having that kind of support, particularly for a young researcher, went a long way to easing frustration and I am eternally grateful for the help and resources ilifu provided me."

Bontle Mere graduated from the Department of Electrical Engineering in 2019; for her final year project she explored the application of IoT in agriculture, using a virtual machine in the ilifu cloud computing facility. Her access was organised by UCT eResearch.

ilifu

Big data is changing the face of research. To ensure that South African researchers can be global pioneers in data-intensive research fields, particularly astronomy and bioinformatics, a consortium of universities and research organisations established a data-intensive research cloud called ilifu, which means cloud in isiXhosa. ilifu is a regional node in the national data infrastructure, partly funded by the Department of Science and Technology, to support the National Integrated Cyberinfrastructure System of South Africa.

A moment in time: repeat photography to understand environmental change

"Historical landscape photography is the closest thing we have to a time machine," says Professor Timm Hoffman, director of the Plant Conservation Unit within the Department of Biological Sciences at UCT. Hoffman has been collecting historical photographs of landscapes since the late 1980s and uses them to document environmental change. As his collection grew, he ran into a storage problem. UCT eResearch provided him with centralised cloud-based data storage to house the collection.

Much of Hoffman's research focuses on environmental change, and historical photographs are the best way to track that change over many years. "Once I have ascertained where the historical photograph was taken, I try and stand in the same place to retake the photograph," says Hoffman.

In this way he can analyse a changing landscape over the period of time linking what he sees in the landscape to climate models, to see if the landscape responds to the changing climate in the way the models predict. Over the years, and with the support of colleagues in the UCT Libraries and other departments, Hoffman's collection has grown into a citizen science project called RePhotoSA. Members of the public can contribute to this project by visiting the site of existing historical photographs and taking their own images or by uploading their own historical images.

A collection in the cloud

"We try to make the photos archival quality when we scan them", says Hoffman. "But this means that a single file can be as big as 50 megabytes. When you have 20 or 30 thousand photographs, you very quickly run out of hardware space."

"The ICTS team have been excellent in providing support for us, creating a specific drive in which we could work with the collection in the cloud," explains Hoffman. "The space has been very stable and reliable and has been an absolute lifesaver."

> **"Timm was one of the first researchers to realise the benefit of centralised data storage," says eResearch Director, Dr Dale Peters. "The value of this service is that the storage is secure and reliable, backed up in multiple locations, and more cost-effective than a commercial cloud solution as it is centrally hosted by the institution."**

Today RePhotoSA has about 6 000 historical photographs which are accessible to the public. These however are only the photographs that are suitable for repeat photography where the exact location is known. Hoffman's collection is close to 30 000 images in total, dating from 1876 to the year 2000.

Below Left: Historical photograph of Silvermine Nature Reserve (Western Cape) before clearing of pine trees, taken by J Cowen in 1999.
Below Right: A repeat photograph taken by J Watermeyer in 2016.
Images © rePhotoSA, UCT & SANBI; CC BY-NA-SA 3.0.

"It takes only a second to take a photograph," says Hoffman, "but in that second you capture a resource so rich in information about the landscape. A collection of images can be used to build a narrative of what has driven change in that landscape over a period of time."

Winner of the WWF Living Planet Award

Professor Timm Hoffman,
Director of the UCT Plant Conservation Unit

Hoffman is the recipient of the annual 2020 Living Planet Award from the World Wide Fund for Nature (WWF), a prestigious award given annually to exceptional South Africans who, through their contribution to conservation, inspire people to live in harmony with nature.

"Timm is not only one of South Africa's foremost arid-zone ecologists, but is also one of the humblest and most compassionate people you will ever meet", said Dr Morné du Plessis, CEO of WWF South Africa. "Through this award we acknowledge an individual whose work exemplifies how conservation truly can benefit both people and nature."

The award honours the full range of Hoffman's work including, among others, his work around repeat photography, which has highlighted massive changes occurring in South Africa's ecosystems over an extended period, as well as his work in the village of Paulshoek in Namaqualand, where he has contributed positively to the community.

Left: Historical photograph of Wagendrift Dam (KwaZulu-Natal) before it was filled, taken by D Edwards in 1961. **Below:** A repeat photograph taken by C Hundermark and H Petersen in 2019. Images © rePhotoSA, UCT & SANBI; CC BY-NA-SA 3.0.

Empowering and supporting African adolescents

The UKRI GCRF Accelerating Achievement for Africa's Adolescents (Accelerate) Hub is a global research collaborative which houses over a dozen studies. Using a variety of methodologies from narrative storytelling, participatory research and quasi-experimental analyses of observational data, to trials and economic analyses, the Accelerate Hub aims to generate evidence that empowers and supports African adolescents. This multitude of mixed-methods studies, carried out in collaboration with partners around the world, raises challenges of data management, data governance and data integrity. UCT eResearch has partnered with researchers at the Accelerate Hub to proactively identify and address these diverse challenges.

The United Nations Sustainable Development Goals (SDGs) have refocused development goals from single objectives to a wide range of integrated targets. To address these goals, governments and policymakers need timely and rigorous evidence. The Accelerate Hub, which is a research partnership between the Universities of Cape Town and Oxford University is funded under the United Kingdom Research and Innovation (UKRI) Global Challenges Research Fund (GCRF). It was developed to generate research in order to help policymakers identify which cost-effective, scalable solutions can be provided to adolescents, to help them reach targets across a number of the sustainable development goals.

To achieve this goal, the Accelerate Hub fosters partnerships between governments, international agencies, non-governmental organisations, donors, academics and the adolescents themselves. It also means doing research differently, working across disciplines and beyond single aims. This comes with a range of data challenges, from point of collection to point of publication and dissemination.

"The Accelerate Hub is home to over a dozen individual studies," explains Dr Elona Toska, the UCT principal investigator (PI). "Each of these comes with its own unique research goals, questions and methodologies which translate into specialised data processes and associated challenges. There are also overarching issues of data governance and legal compliance that apply across the hub. Fortunately, UCT eResearch, working across departments at UCT, has been able to provide support at both the micro and the macro level."

Bringing data ownership home to Africa

The Accelerate Hub has developed from a long-standing collaboration between UCT and Oxford universities on adolescent health and development research. Originally, the data was stored at Oxford as it was the main recipient for funding for this portfolio of research.

"But as the research progressed, we had to think more critically about data and its regulatory environments," says Toska. "The project PI's had always envisioned a power shift in the work from a northern institution like Oxford to an equitable partnership with a southern institution such as UCT, part of this involving the storage of the data at UCT."

This is where the collaboration with UCT eResearch really began, with questions around hardware and how to set up the server according to the requirements of the project.

Data governance

The Accelerate Hub is funded by several United Kingdom (UK) and European Union (EU) grants, including the UKRI GCRF. This means research, carried out by the Hub, needs to comply with both the EU's General Data Protection Regulation (GDPR) and the legal framework of the country in which the research is being conducted.

For example, within the Accelerate Hub, UCT eResearch provided a great deal of support to one particular study called HEY BABY (Helping Empower Youth Brought up in Adversity with their Babies and Young children). The research was undertaken in the Eastern Cape province of South Africa and needed to comply with the Protection of Personal Information Act (POPIA) of South Africa which came into effect on 1 July 2020.

The UCT eResearch team helped us navigate this regulatory minefield," says Toska

There will be more data-governance challenges down the road where the Accelerate Hub team will rely on the UCT eResearch partnership. One example is in meeting the requirements of a data-protection impact assessment to help review their data, and structure the data-management system on the storage server.

> **"We need to answer questions like who has access to the raw (non-anonymised) data and how can we ensure its integrity, but also facilitate access to that data so it can be cleaned, analysed and used to improve the lives of adolescents," says Toska.**

They also need additional support for data-sharing agreements between partners, not only academic partners but also those of NGOs and governments. Here the UCT eResearch collaboration with the Research Contracts & Innovation office (RC&I) is integral, as RC&I provides the legal insight while UCT eResearch offers expertise on what infrastructure and support is available at the institution.

Empowering the Accelerate team

The scale of data collected under the Accelerate Hub umbrella meant that, beyond the partnership with UCT eResearch, the Accelerate team needed in-house data-management expertise.

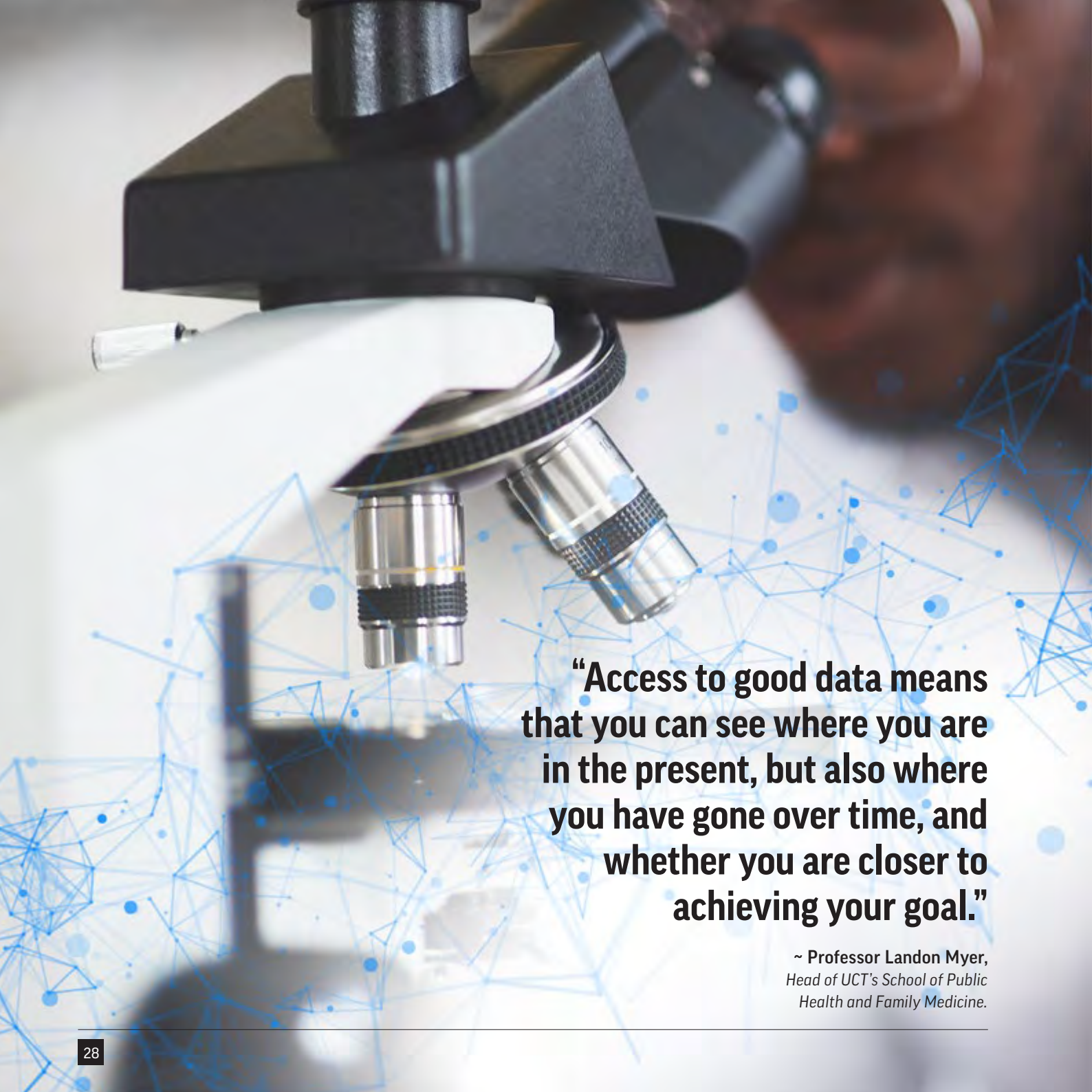
"With just the HEY BABY project collecting data from over eight different sources, and the Accelerate Hub being home to more than a dozen studies, it was apparent that we were quickly going to be in over our heads in managing the data needed to meet our research goals," says Toska.

The UCT eResearch team provided support developing the requirements for an in-house data steward post. There are also plans in place for support of the Accelerate team to help the research group better understand the different platforms which enable good research practice, including reproducible research and data management.

> **"The work with the Accelerate Hub has been an excellent example of the importance of close collaboration between research groups and the various professional and administrative support staff (PASS) departments at the university," says UCT eResearch Director, Dr Dale Peters.**

This collaboration showcases the necessary adjustments and processes that UCT eResearch, UCT Libraries, and RC&I are developing in response to the needs of UCT's research community.

"Key to this research and collaboration," says Peters, "is how researchers across disciplines and institutions are partnering with PASS staff across departments to tackle the challenges laid before us by the sustainable development goals."



“Access to good data means that you can see where you are in the present, but also where you have gone over time, and whether you are closer to achieving your goal.”

~ Professor Landon Myer,
Head of UCT’s School of Public Health and Family Medicine.

Tracking transformation in the School of Public Health and Family Medicine

The UCT Vision 2030 strategy rests on three key pillars: excellence, transformation, and sustainability. As part of the drive towards transformation, each department at UCT has a transformation committee whose mandate is to promote the actioning of practical steps towards faculty and university transformation plans. In the School of Public Health and Family Medicine, the transformation committee included people who, with the help of UCT eResearch, could bring their analytic skills to develop an application that, over time, tracks and displays the changing demographics of the school.

When it was first set up in 2015, one of the first things the School of Public Health and Family Medicine's transformation committee identified was the need, over time, to monitor the demographic profiles of the staff and students at the school.

"Access to good data means that you can see where you are in the present, but also where you have gone over time, and whether you are closer to achieving your goal," says Professor Landon Myer, head of UCT's School of Public Health and Family Medicine.

Fortunately, members of the transformation committee had previously worked with UCT eResearch, and knew they could reach out to them for support in the development of a Shiny Server web application to track the transformation demographics of the school.

"The Shiny App works like an interactive computer-based dashboard, which allows the user to view summarised data and visualisations through a 'point and click' interface, without needing to view the raw data," explains Associate Professor Maia Lesosky, head of the Division of Epidemiology & Biostatistics, who supported the app's development.

Luke Hannan, a PhD researcher supervised by Lesosky, worked with UCT eResearch on the coding and development of the application.

"Once the dashboard was developed, we needed somewhere to host it online in a way in which people could access the dashboard but ensure the school maintains control of both the dashboard and the data," says Hannan. For this, the transformation committee reached out to eResearch for support.

"Ashley Rustin of the eResearch team in ICTS was very helpful in providing and securing the back-end server that hosts the dashboard," says Hannan.

This meant that members of recruitment committees and other relevant groups could easily access the app and see the demographic data visualisations to make informed decisions based on complex information.

> **"These visualisations and analytics do not provide answers to questions," says Myer, "and we cannot pretend that the application is transformative on its own. Just like any analytic tool it does not change things, but what it does is provide us with the information we need to drive the necessary change."**

Section 3

Supporting the research enterprise

For researchers to truly be able to compete globally and 'unleash knowledge in and from Afrika', in line with Vision 2030, they need access to integrated, responsive and well-governed research infrastructure as well as the support of effective and innovative research management. Part of UCT eResearch's role is to support this behind-the-scenes work to ensure the three-pillared goals of excellence, sustainability and transformation can be achieved in the university's research endeavour.

Reflecting on lockdown and COVID-19

On the 23rd March 2020 President Cyril Ramaphosa announced South Africa was to follow suit of several countries around the world and enter a hard lockdown in response to the global COVID-19 pandemic. The introduction of global lockdowns threw us all in the deep end and researchers and staff at UCT were faced with a steep learning curve as work moved completely online. Fortunately, UCT eResearch and other departments across the university, already had systems in place for such an eventuality.

"When UCT mandated that all staff who could work from home should do so, ahead of President Ramaphosa's address, we were, thanks to the foresight of many staff and researchers, already in a good position to implement remote working," says eResearch analyst, Renate Meyer.

"Systems like Microsoft (MS) Teams for collaboration, SuccessFactors and LinkedIn Learning for online learning, and a range of relevant research software were already in place, ready for user uptake."

UCT eResearch had already worked with a number of research groups to develop use cases for how MS Teams could specifically be used for research. Making use of the regular Research Announcement, a quick notice went out to campus with some of the lessons learned.

"It has also been impressive to see how quickly researchers adapted and became proficient in systems and software that allowed for remote working," says Meyer. "In the first month of lockdown, we saw an 850% increase in the use of MS Teams within UCT.

Remote research

The other area where eResearch was able to offer a great deal of support was in requests around research software and collaboration, particularly around how best to conduct online surveys.

"Prior to lockdown, we had already seen interest from researchers in moving field research online; this included examining the related issues of secure data transfer, data storage and data security," says Meyer.

Lessons for the future

After COVID-19, the world of work is likely to be changed forever. Being forced to make working remotely possible in such a short turnaround time has brought with it many lessons as to where we need to strengthen our structures and support.

"The digital divide and stark inequalities in our society, even among colleagues, is something we need to address as a society and as an institution," says Meyer.

"Further, concerns around online security needed to be addressed as we moved online for both research and teaching & learning. We consistently examine these points of risk where we are most vulnerable and work toward proactive solutions."

> **"What we have seen from researchers and staff alike at UCT, was a commitment to keeping the wheels turning, both in research, teaching, and learning. This was nothing short of inspiring."**

Building a FAIR data culture at UCT

Digital Library Services (DLS) has been working to build the UCT Data Stewards and Champions community in order to drive the development and maintenance of a vibrant and sustainable data culture in research at UCT.

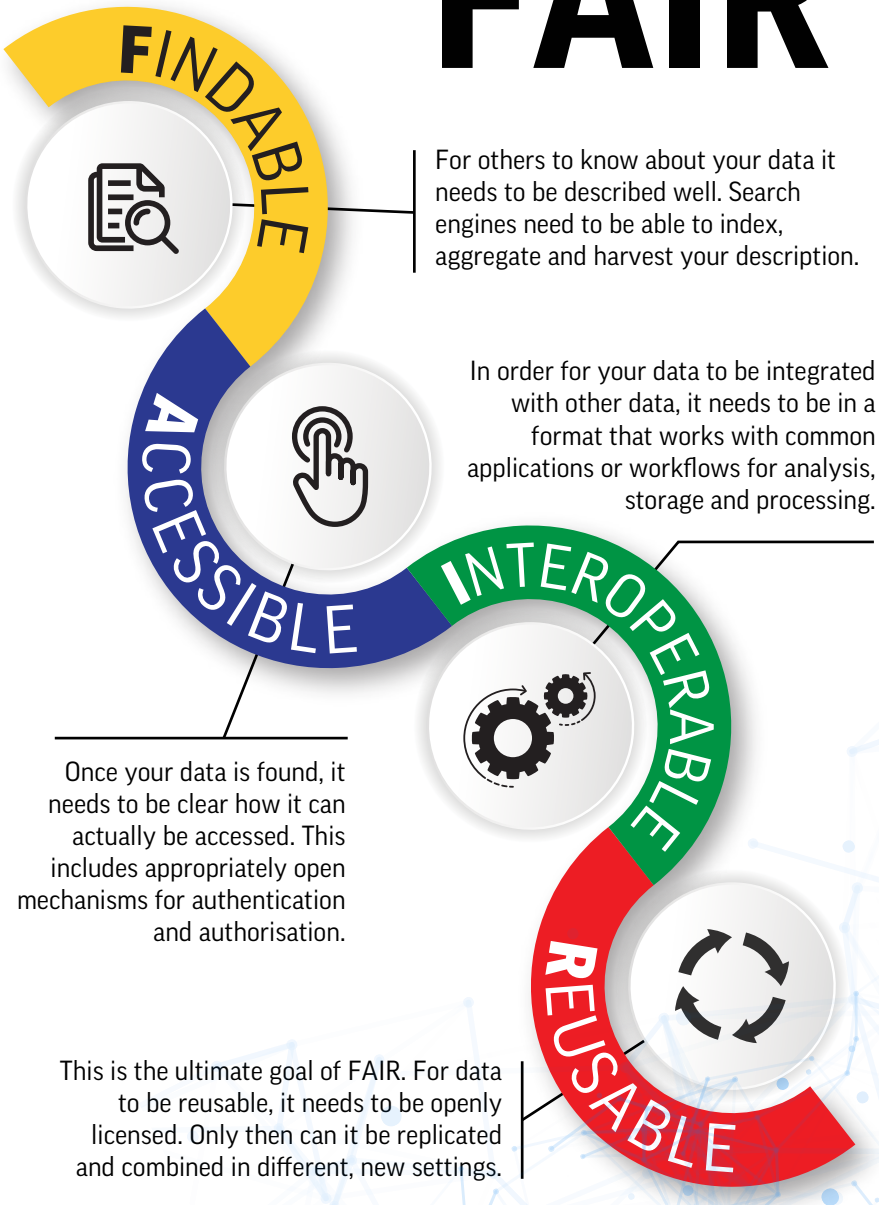
The data stewards and champions serve as a vital network between their research groups and staff at DLS, ensuring that disciplinary expertise is combined with an understanding of good data practices to drive FAIR data publication at UCT, with all its concomitant benefits.

The stewards provide domain-specific knowledge to assist the libraries and their colleagues with the processes of curating research data from their specific fields, and the champions advocate for the importance of the FAIR data principles within their departments, divisions and faculties.

"We see the immediate benefits of good data management and data publication according to the FAIR Principles on UCT's open access data repository ZivaHub," says Niklas Zimmer, manager of DLS. "Well-curated data sets enjoy far greater visibility and increased chances of citation for the researcher."

Open practices at UCT can also be seen in action on UCT's instance of the Open Science Framework, as well as UCT's Zenodo community. Furthermore, DLS supports the UCT community with FAIR compliance by managing a dedicated platform for creating data management plans called UCT DMP.

What does it mean to be FAIR?



ZIVAHUB: OPEN DATA UCT

ZivaHub, launched in November 2017, is beginning to see uptake by UCT's research community as awareness around the importance of FAIR data grows. This spread shows the uploads, downloads, total views and storage capacity of ZivaHub in the reporting period June 2018 to June 2019.

TOTAL VIEWS

125 216

Number of times ZivaHub content was viewed by internal and external users

12

TOTAL CITATIONS

556

TOTAL DEPOSITS

DOWNLOADS

29 184

Number of times items were downloaded

Supporting reproducible research with the Research Data Integration Project

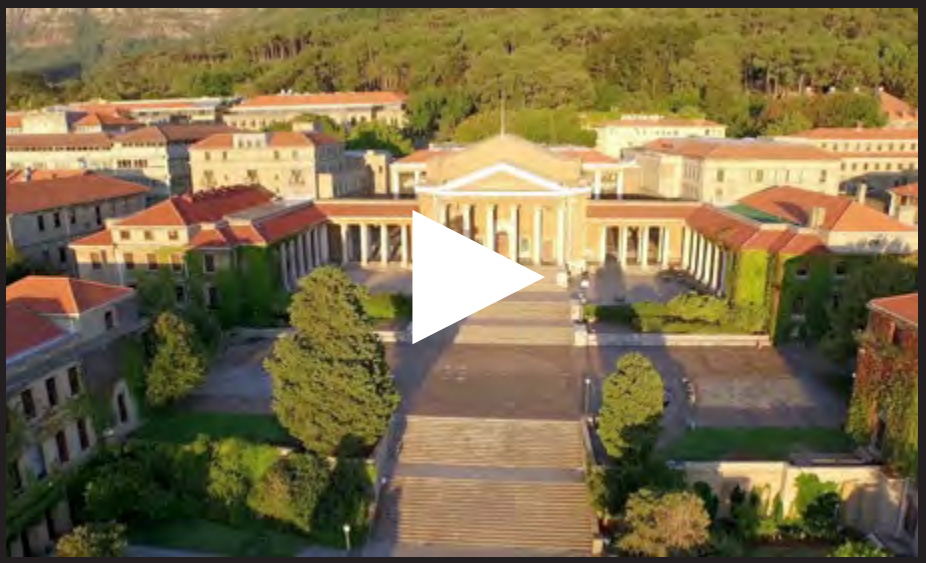
UCT's research vision 2030 includes the goal: 'unleash knowledge in and from Afrika to redefine and co-create a sustainable global future.' This kind of ambitious goal requires an integrated, responsive and well-governed infrastructure to support research.

What is RDIP?

UCT has embarked on a Research Data Integration Project (RDIP) to integrate our research support systems throughout the research lifecycle to help you, the researcher, avoid duplication and achieve the goals of reproducible research. For more about this project, watch the video below.

Reproducible research at UCT

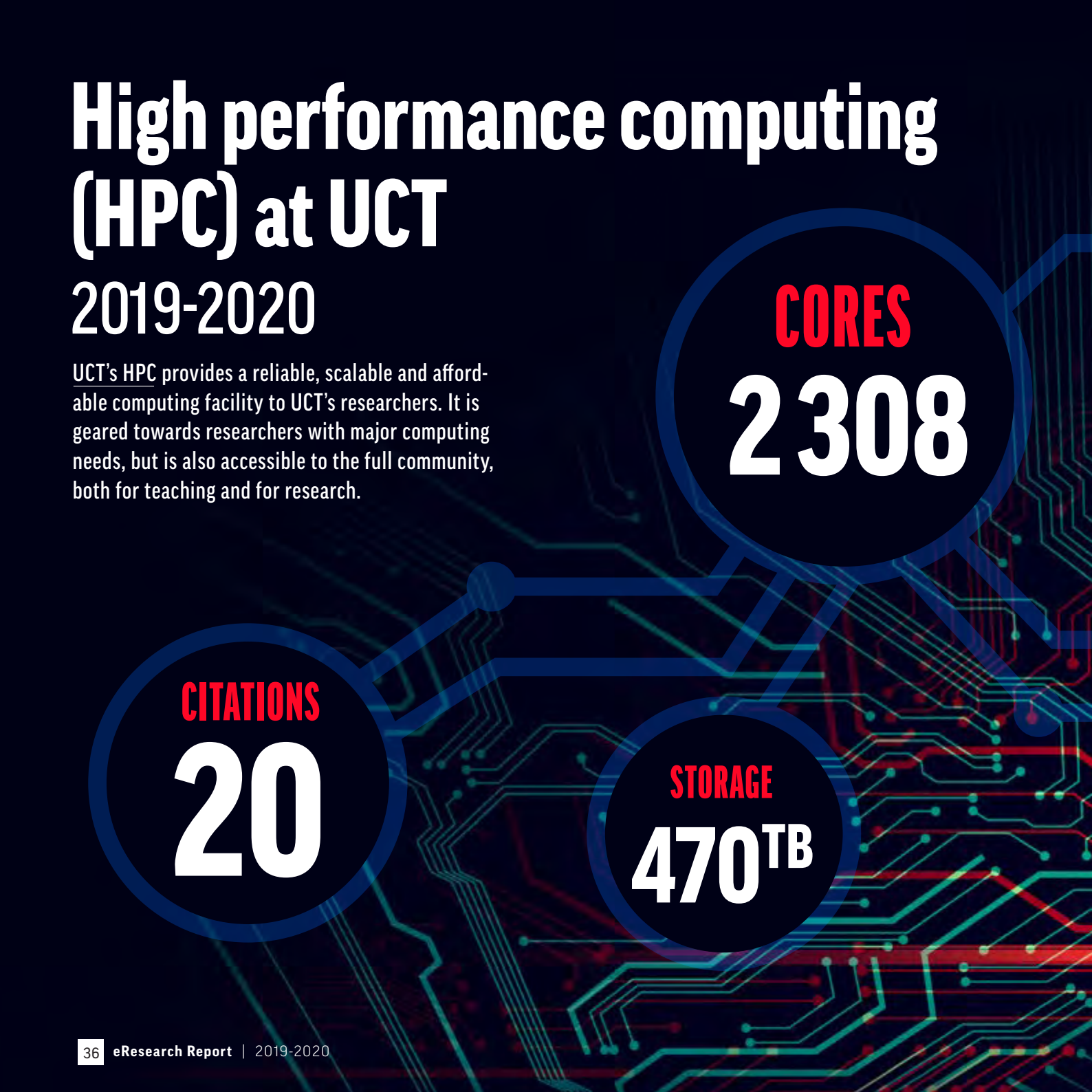
The practice of reproducible research is increasingly being mandated by funders to ensure research findings can be verified, and expanded on by others. This means researchers are often required to publish their data and software code along with research publications. Watch the video below, part of the Research Data Integration Project (RDIP) series, to find out more about how UCT enables reproducible research.



High performance computing (HPC) at UCT

2019-2020

UCT's HPC provides a reliable, scalable and affordable computing facility to UCT's researchers. It is geared towards researchers with major computing needs, but is also accessible to the full community, both for teaching and for research.



GROWTH SINCE 2015

YEAR	CITATIONS	CORES	STORAGE (TB)	JOBS	CPU HOURS
2015-2016	27	1 458	219	260 678	5 500 000
2016-2017	15	1 428	219	510 000	4 020 000
2017-2018	24	1 458	219	369 066	3 217 898
2018-2019	24	2 440	470	618 084	4 746 242
2019-2020	20	2 308	470	485 415	7 523 820

USER UPTAKE
93

7.523
MILLION
CPU HOURS

485 415
PROCESSED JOBS

UCT eResearch
ACCELERATING RESEARCH

Produced by Simon Communications
for UCT eResearch
November 2020
Writer and editor: Natalie Simon
Design: polygram.co.za