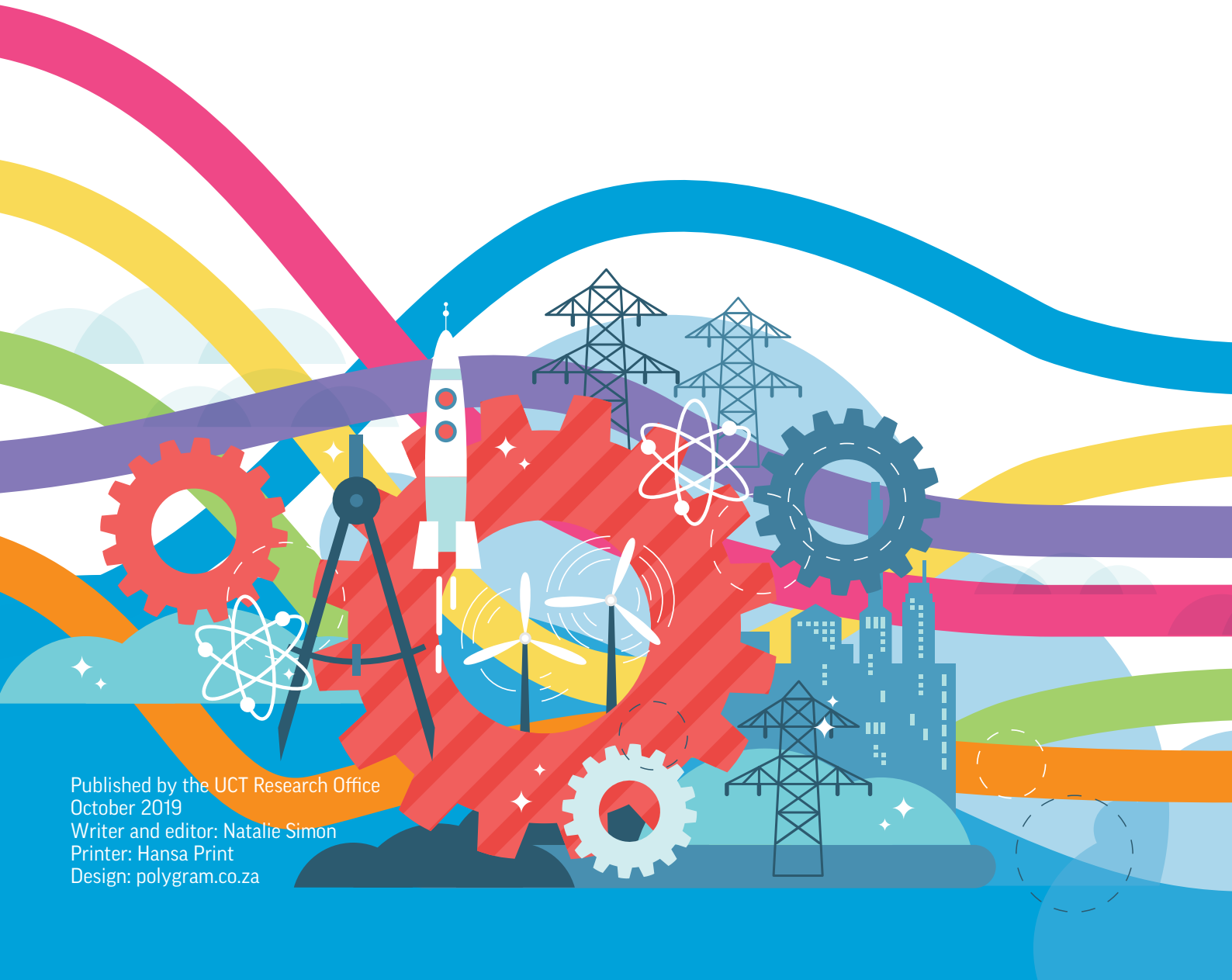

UCT
eResearch
ACCELERATING RESEARCH
2018-2019
eResearch report
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD



Published by the UCT Research Office
October 2019
Writer and editor: Natalie Simon
Printer: Hansa Print
Design: polygram.co.za

CONTENTS

eResearch support for changing research practice

Researchers the world over are feeling the ground shifting below their feet as new technology, legislation and policy drive behaviour change in our traditional research practice. Fortunately, five years ago the UCT executive identified these coming shifts in the research landscape and established UCT eResearch to respond to what was described by then DVC for Research, Professor Danie Visser as a "period of acceleration and transformation".

UCT eResearch is a distributed organisation made up of a partnership between Information and Communication Technologies (ICTS), UCT Libraries and the Research Office, and it is in the strength of this partnership that the success of UCT eResearch lies. As researchers are being asked to work in a way that is more globally collaborative, more interdisciplinary and with greater compliance demands from both funders and governments, eResearch addresses the challenges they are now facing.

Responding to these demands has required co-ordinated effort from research support units across divisions working together in new ways. For instance, the development of the Research Data Management policy and the infrastructure to support that policy has brought together wide-ranging experts faciliated by eResearch. Disparate services – in particular UCT's data repository, ZivaHub (UCT Libraries), legal and contracts (Research Contracts & Innovation), research ethics (Office of Research Integrity in the Research Office), cybersecurity (ICTS) and technical infrastructure (ICTS) – have pulled together to ensure UCT is able to respond to new requirements of publishers and funder mandates for open publication of data. The alacrity with which eResearch is responding to changing research practice underpins our strategy to advance UCT as a research-intensive university that makes a distinctive contribution to knowledge both locally and globally.

Professor Sue Harrison
Deputy Vice Chancellor for
Research and Internationalisation

UCT eResearch: the changing landscape

The research landscape is rapidly changing, with new technologies enabling new forms of collaboration, data gathering and analysis. UCT eResearch is this university's response to the changing research landscape, offering a range of services to assist with the new skills required of researchers in today's digital age.

This is a moment to acknowledge the importance and resilience of core services – such as ensuring the security of the network infrastructure and the provisioning of power and cooling to data centres – which are barely visible to researchers but provide the backbone of the eResearch value chain. So too can we celebrate the growth of the more visible services such as high-performance computing and data storage as well as the notable shift in recent months reflecting new forms of social accountability for researchers. Publicly funded research is valued as a public good.

Researchers and other stakeholders should be able to discover, reproduce and rapidly advance preceding scientific discoveries. This open science paradigm highlights the need to professionalise data-mangement practice, in response also to a growing awareness of data-protection regulations.

As a result of this shift, the interest in integrating eResearch capabilities into research practice is growing. In disciplines like chemistry and physics, eResearch is an established tool, increasingly essential to globally competitive research. In other disciplines, such as the humanities, the introduction of a new service supporting digital scholarship has seen remarkable uptake in the preparation of grant funding proposals and in the adoption of reporting mechanisms in response to recent policy developments in the Department of Higher Education and Training to recognise and subsidise quality creative research outputs.

Our response to the changing research landscape has been a coordinated effort to showcase new and established research support services in training programmes for various target groups, provided across stakeholder partners in the Research Office, UCT Libraries, ICTS, and beyond.

Dr Dale Peters
UCT eResearch director

Supporting and informing our research

Thanks to the eResearch partnership, we have been able to identify and respond ahead of time to changing legislative and policy funder requirements. This has strengthened the research support we are able to provide, particularly through the early career researchers seminar series, as well as new communication mechanisms we have put in place this year to keep our research community informed and up to date.

A pilot project on dedicated staff support for advanced proposal development that includes advice on data management has similarly borne fruit and will be expanded in future.

It is anticipated that the new electronic Research Adminstrative system (eRA) will play an increasingly important role in supporting data management and data integration across the university.

All of which serve to help us support our researchers through all aspects of their work.

Marilet Sienaert
Executive Director, Research Office

FAIR data at UCT

One of our goals is to ensure valuable research data produced by UCT is published in a manner that is FAIR: findable, accessible, interoperable and reusable.

This involves creating standardised metadata (data describing data) and assigning a unique and persistent digital object identifier (DOI). Data and metadata is stored in a trusted repository, to be available for reuse across different disciplinary communities, workflows and systems. Data can only become reusable and citable if an appropriate open license is attached.

Library staff, including the team at Digital Library Services (DLS), work tirelessly with ICTS and the Research Office, through the eResearch partnership, to provide the tools and services that our researchers need in their international research collaborations, and to make their data practice FAIR, in the service of open scholarship.

Ujala Satgoor
Executive Director,
UCT Libraries

The unsung heroes of research support

Many of the services our researchers and students take for granted, such as a reliable and secure network, compute and data storage are the result of the behind-the-scenes work of our technical specialists, particularly based with Enterprise Infrastructure Services (EIS).

In addition to this direct support, EIS and the other divisions in ICTS are involved in several projects aimed at providing improved support for the administrative and operational processes related to research, including the management and protection of research-related data

Much of this seemingly invisible infrastructure and support provides the backbone to the eResearch support services discussed in this report. We look forward to these ICTS-supported infrastructure and services expanding in the future as we provide the technological platform to allow our researchers to remain globally competitive.

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


DVC Research
Leadership
Governance
Strategy
Resourcing
UCT Libraries
Visibility & discovery
Research data management
Digital scholarship
Research Office
Advocacy
Collaboration
Science communication
Faculties
Data science
Shared facilities
Data stewardship
ICTS
Data storage
Compute
Networks
Cloud
UCT eResearch

Section 2

Insights on a changing research landscape //

Researchers are facing a rapidly changing landscape and UCT eResearch is working to ensure they are equipped to face the challenges.

Responding to the changes in technology, legislation and policy requires coordinated effort from research support units and changes in behaviour from researchers themselves. This section provides some thoughts, guidelines and insights to help researchers think through some of the issues they may find themselves facing now or in the near future.

Cybersecurity and your research data

In the digital age, information is a commodity that criminals will go to great lengths to get their hands on. Research data is no different. Roshan Harneker, senior manager for ICTS, offers a few key points to bear in mind to maintain the integrity of your research data.

Classify your data

The first step is to know what data needs to be protected and to what extent. Currently there are no institutional data classification guidelines so Harneker suggests research groups develop their own data classification systems in order to easily define public, sensitive, confidential and restricted data.

Who has access to your data?

When using a cloud-based storage service, like MSTeams or Google Drive, be sure to set permissions carefully according to your data classification system. Once you have classified and set your permissions, be sure to clearly spell out terms and conditions associated with that data access.

Where is your data?

Consider every device your data may be housed on, says Harneker. While the intention may be to keep the data securely on password-protected hard-drives, you may find yourself working on a subset of that data downloaded onto a laptop that is not password secured. Harneker advises ensuring all devices that may contain sensitive data be protected or even encrypted in the same way.

Are you complying with all relevant legal frameworks?

Familiarise yourself with the legislative requirements around data collection, use and storage, not only our own Protection of Personal Information Act (POPIA), but that of your collaborators too. The General Data Protection Regulations (GDPR) adopted by the European Union in May 2018 are applied widely in contractual agreements between collaborative partners to ensure that adequate security measures are in place to protect the management and transfer of personal information.

Is your hardware secure?

A number of researchers buy and manage their own hardware without thinking about important cybersecurity details like encryption and access control. In these cases, Harneker suggests at least consulting with ICTS over security measures in both the short and the long term.

Big data will influence us, it will empower us to make decisions – but not all decisions using big data will result in the outcomes we are looking for. We need to rely on our experience, our insight and understanding, as well as the data.

~ Professor Ulrike Rivett

Decision-making in a time of big data

Big data has changed the world forever – just ask Netflix. But, argues Professor Ulrike Rivett in her keynote address at eResearch Africa 2019, when it comes to society, public sector and government, big data is not the silver bullet we imagine.

"It is worth reflecting on how big data is used, by whom and for what," says Rivett, director of the UCT's School of Information Technology. "If you look at the top businesses in the world right now, the biggest transport company doesn't own a single vehicle, the biggest property manager doesn't own a single property, the biggest entertainment house doesn't pay actors and the biggest shop doesn't employ a single cashier. These businesses – Uber, Airbnb, Netflix and Amazon – are all digital businesses set up entirely on data systems."

Big data in the private sector

Netflix makes a wonderful case study of how a private company has used big data to disrupt the television industry and set a trend.

"How have they been so successful? Big Data, with capital letters," says Rivett. Netflix is always capturing data to monitor how you watch, when you watch and what you watch, all to allow them to respond perfectly to your exact needs.

If the use of big data to guide decision-making has been so successful in the private sector, can we do the same in the public sector? Unfortunately, the same rules simply don't apply, argues Rivett.

Collecting data: by whom, for whom?

Reflecting on her own experiences both in the health and water and sanitation sectors, Rivett stresses the incentives of the person on the ground you are asking to collect this data for you. What is this data being used for?

In one study Rivett was involved in, the funder requested that 180 data points be collected on every patient in the clinic. This, she says, was far beyond the normal scope of a nurse or doctor's work. Nor is there any benefit to the staff or the patients in a public clinic to develop massive automated systems for patient data records.

The nurse won't become a better nurse because she collected the 180 data points for somebody located on another continent," says Rivett. "This data is not going to help her care better for patients – and that is what government clinics in developing countries are all about, caring for the public – not collecting data.

Data, decision-making and behavioural change in the public sector

Rivett worked very closely with the City of Cape Town during the water crisis of 2017/18 and says it offered fascinating insights into how far data can take you, and how data is used in the context of government decision-making.

One of the things the City used as an incentive to citizens was to make water-usage across the City transparent, to encourage Cape Town residents – particularly those in the wealthier areas where the most water was being used – to save water.

"The response to the publication of the data was immediate public outcry that the City had manipulated the data. People could not believe it was them wasting the water."

In the end, attitudes and behaviour changed with the announcement of Day Zero: the day the taps would run dry in Cape Town.

"The moment Day Zero was announced everybody went into Armageddon mode," says Rivett. People began to buy bottled water for household use and the social condemnation for water-wasting behaviour became a driving factor for change. Everybody was seeing who could save the most water. And that strategy worked. There was no big data and there was not even a good model for estimating the occurrence of Day Zero.

> **"In the end what had the biggest impact, what turned Cape Town into an amazing example of catastrophe averted, did not need to be driven by big data," says Rivett. "It was driven by an understanding of human nature and behaviour."**

Data sharing: challenges around ethics and compliance

"Open science is good science," says Dr Lyn Horn, director of UCT's Office of Research Integrity. "It is transparent, replicable and verifiable." In recognition of this funders are increasingly demanding data be published openly in a reliable repository.

"But as much as data sharing is now an ethical imperative," says Horn "there are major hurdles and challenges to it. We see a clash of competing values and principles, particularly in a health research environment."

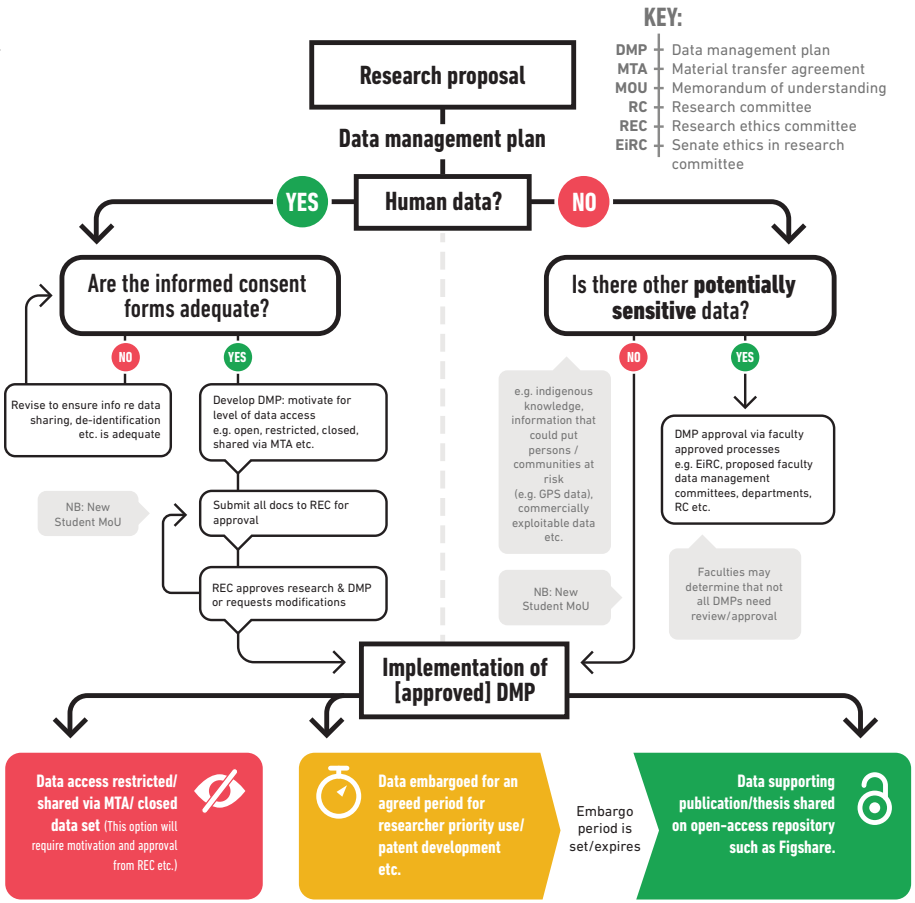
> **The research ethic of solidarity and reciprocity and the value of open science for future discovery needs to be carefully balanced with the privacy rights of an individual and that individual's right not to be exploited.**

Research participants must be offered the opportunity to consent specifically to the use of their data in research and must be made fully aware of how their data will be processed and shared, if applicable.

The local and international regulatory environment also needs to be borne in mind, advises Horn, particularly our local Protection of Personal Information Act (POPIA) and the European Union's General Data Protection Regulation (GDPR).

"For instance, POPIA states that protected health information data may not be transferred across a border unless equivalent data protection laws exist in the receiving country," says Horn. "This has major implications for research collaborations."

These competing values and principles highlight the importance of careful planning ahead of embarking on a research project, and the value of a good data management plan.



There are human beings behind that data: ensuring good data governance for ethical research

Globally, the volume of private and personal digital data has massively increased, accompanied by rapid expansion in the generation and use of digital health data.

These technological advances promise increased opportunities for data-driven and evidence-based research. However, it also increases the risk to individuals of data misuse or data breach of their sensitive personal data, especially given how easily digital data can be accessed, copied and transferred on electronic platforms if the appropriate controls are not implemented. This is particularly pertinent in low– and middle-income countries, where vulnerable populations are more likely to be at a disadvantage in negotiating digital privacy and confidentiality.

Based on their experience working in health and genomic sectors, Professor Nicki Tiffin of the Department of Integrative Biomedical Sciences and her collaborators developed a framework for digital data governance, comprised of four main pillars:

- **ethics and informed consent**
- **data access**
- **sustainability**
- **legal requirements.**

To help guide researchers through these four pillars, Tiffin developed key points for data governance.

Key points for good data governance

1 The study participant knows what I am using their health data for, and is ok with it.

2 The ethics board knows what health data I am using, and how, and is ok with it.

3 This data use complies with legislation (POPIA, Healthcare Act, in SA).

4 All data sets and drives are password protected and encrypted.

5 All participant identifiers are stored and transferred in a separate file to any clinical data.

6 All anonymised data cannot be re-identified.

7 I never use email to transfer identifiable health information, and I always send passwords separately (not by email).

8 If I were a study participant I would be happy with the way my personal data is being used.

Source: This article is a slightly condensed version of the abstract from the following paper: Tiffin N, George A, LeFevre AE. *How to use relevant data for maximal benefit with minimal risk: digital health data governance to protect vulnerable populations in low income and middle-income countries. BMJ Global Health* 2019;4:e001395.

Rise of the data steward

As the academic environment becomes increasingly digital, student and staff researchers are producing ever more data. To stay on top of the deluge, UCT is working to build and support a community of data stewards and champions who advocate and implement best practices around data.

"This community is not only comprised of dedicated data stewards embedded in their respective research communities, but also researchers who act as champions to share knowledge about best practices in working with scholarly data," says Dr Patricia Chikuni, data curator at Digital Library Services, and coordinator of UCT's data stewards and champions community.

The stewards and champions are vital: their subject expertise adds a quality assurance layer to the data curation process that somebody outside of the specialist field could not hope to achieve.

One of these data stewards is Dr Mbandi Kimbung, a research officer at the South African Tuberculosis Vaccine Initiative (SATVI). For Kimbung, his experience working on a project funded by the Bill and Melinda Gates Foundation – which required that the data be published in a reliable open-access repository – made clear the need for a dedicated steward to think through the various complexities in data curation, publication and sharing.

"It has been very gratifying to be able to support other researchers who wish to make their data sets publicly available on ZivaHub," says Kimbung. "Now that we have best practices in place, the road is paved for the publication of FAIR (findable, accessible, interoperable and reusable) data to become standard practice."

Below: Dr Patricia Chikuni, data curator at Digital Library Services, is also the coordinator of UCT's data stewards and champions community which works to advocate for use of best practice in working with scholarly data.

Breaking down the silos: a shared software for facilities management

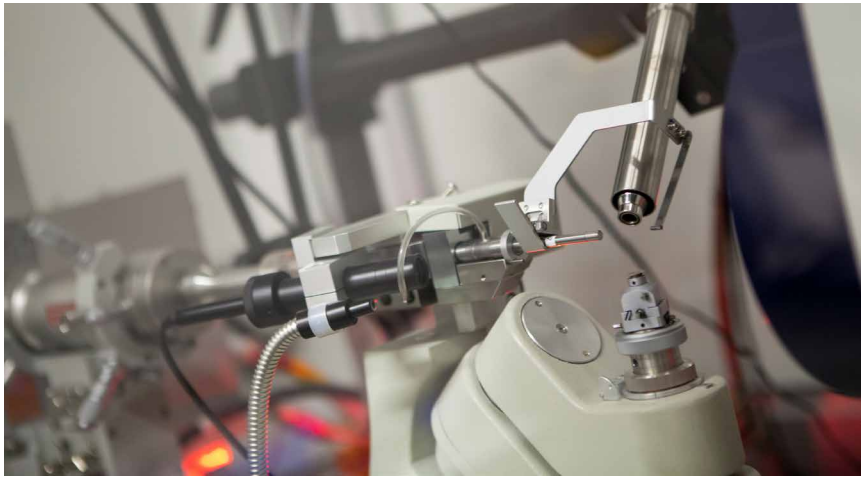
Researchers across UCT own and operate a myriad of facilities – instruments, software packages and services – all of which are expensive to buy and maintain. The solution is to ensure these facilities are adequately shared so they don't stand idle and that users outside the institution pay to cover maintenance costs. But this too is an administrative burden. This year UCT eResearch completed the facilities management project and identified Calpendo as the preferred software to be used university wide to optimise the use of these facilities.

Renate Meyer, eResearch analyst, consulted widely across faculties to understand the varying needs and responsibilities around the management and upkeep of facilities at UCT.

Finding a way to make facilities cost-effective has benefits way beyond saving the university money: "The higher education sector is looking for ways to improve graduate education, develop sustainable career paths for emerging researchers and encourage collaboration through research and innovation," says Meyer.

> **"These facilities play a large part in enabling cutting-edge research, developing and retaining leading scholars and serve as a nexus for inter-disciplinary collaboration."**

Below: An x-ray diffractometer is one of the instruments offered for use by the Electron MIcroscope Unit who use Calpendo to manage their facilities.
Image by Stephen Williams.



The challenge of giving scientific software its due

Better software, better research, reads the motto of the Software Sustainability Institute. And it is true that today, as digital technologies mean researchers can collect larger data sets, the software to interpret that data is indispensable to the research process. But what are researchers doing to ensure the quality and longevity of their code and to recognise the work of those often early-career researchers writing the code?

"Programming skills are required more and more across all research disciplines, and university curricula have the potential to integrate these skills into undergraduate and postgraduate training," says Renate Meyer, eResearch analyst.

The next step, she says, is software transparency and sustainability of the workflows developed in research groups by offering informal training programmes for developers to complement formal data science curricula.

"Platforms like GitHub are widely used and already integrated with ZivaHub, enabling open science publication of technical papers," says Meyer, "but we are also looking to validate the career path of those early-career researchers and research engineers through a UCT-based community of practice and academic acknowledgement."

Section 3

Supporting our research //

As a globally competitive, research-intensive university we need to ensure our researchers have access to cutting-edge facilities, infrastructure and support.

The digital era in which we are currently operating brings new challenges and opportunities for scientific research, offering new ways of global interaction within disciplines, as well as collaboration between disciplines. The UCT eResearch mandate is to support our researchers through these challenges so they can access the opportunities of the digital era.

This section contains just two case studies of many which show how eResearch supports our researchers through the research lifecycle.

The devil is in the detail: planning for the data lifecycle in a longitudinal, multi-institutional study

Professor Catherine Ward of UCT's Department of Psychology is working on the South African piece of an eight-country longitudinal birth-cohort study led by Cambridge University which seeks to understand violence against children. UCT eResearch worked with Ward to help her anticipate and resolve the many research data problems that are likely to crop up during the course of the study.

End all forms of violence against children – so states the United Nations Sustainable Development Goal 16.2. In order to achieve this critical goal, policy-makers need to understand the complex factors at play. This is the objective of the eight-country birth cohort study coordinated by a team at Cambridge with research collaborators across the globe. The goal of this study is to track a total of 12 000 pregnant women in eight low-income countries through their pregnancies and into the adulthood of the children in the birth cohort themselves.

> **"Trying to think through the various implications of data management in a research study of this level of complexity was daunting," says Ward. "Without the proper support there are data issues a researcher can stumble into very naively, with devastating consequences ten or even 20 years down the line."**

Ward thus turned to eResearch analyst Renate Meyer to support her through the various steps of the research lifecycle. Ward used UCT Digital Libraries Services' data management planning tool, UCT DMP, which she says was invaluable in the process.

Data collection

"These days this kind of data is collected electronically through a tablet," explains Ward. This has a number of advantages including that the participants in the study can answer sensitive questions completely confidentially.

Meyer and Ward discussed various data collection software options and the pros and cons of each. "While the decision as to what software to use will be made at Cambridge, the advice from UCT's Information and Communication Technology Services (ICTS) allows me to contribute meaningfully to the decision-making process," says Ward.

Secure (African) storage versus ease of collaboration?

Once the data has been collected and uploaded to a server, many of the real data management challenges begin.

"We are asking questions about highly sensitive issues, including substance abuse, intimate partner violence and HIV status among others," says Ward. "In addition, because we are tracking these women over decades we need to keep records of personal information - names, addresses, telephone numbers, telephone number of family members, etc."

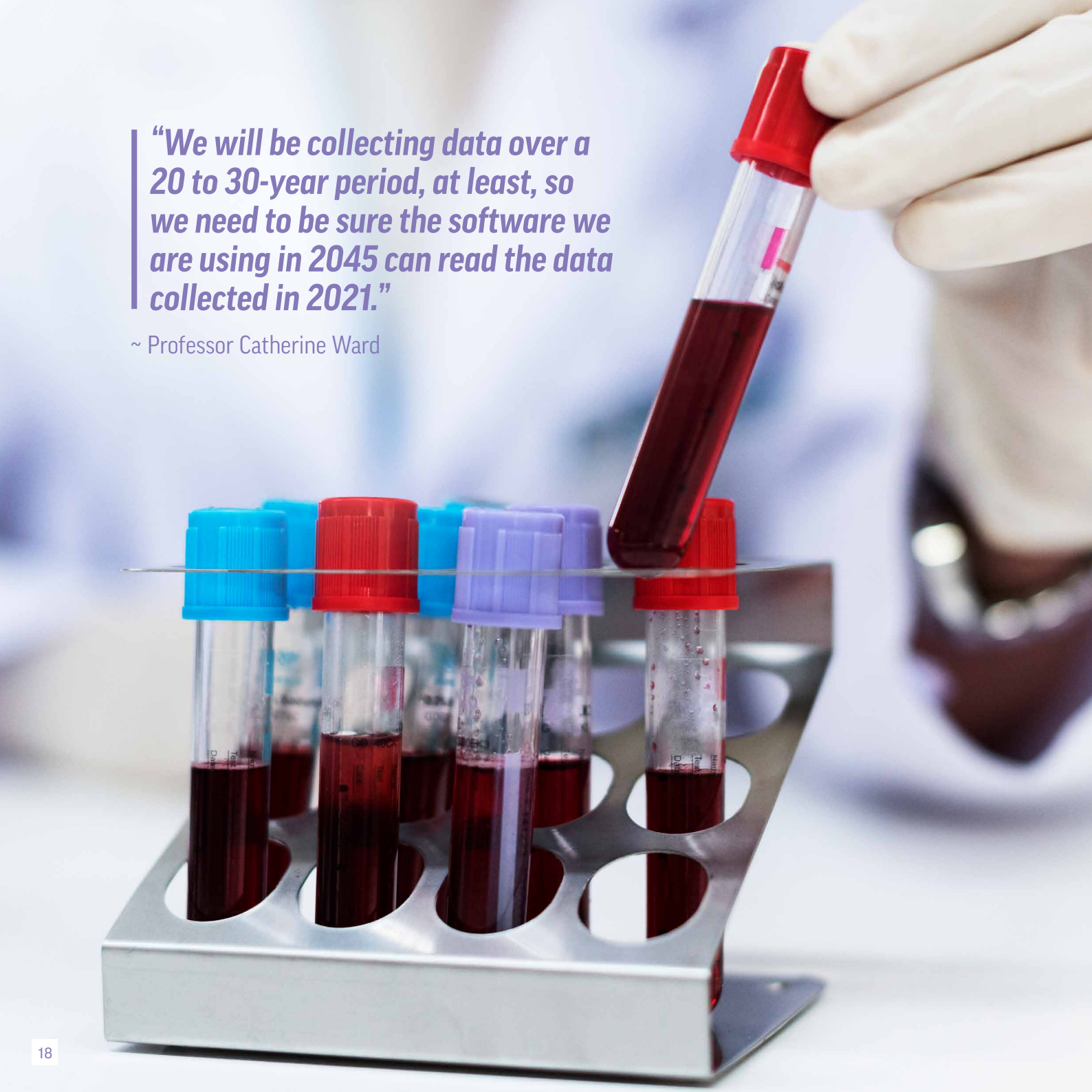
The project therefore cannot risk a security breach, but at the same time needs to be able to share anonymised data with collaborators all over the world. It is also extremely important to keep the data in Africa, under local jurisdiction.

Meyer worked with Ward to develop a plan to keep the sensitive data highly secure and offline – accessible only to the South African team – with the anonymised data for collaboration available to research partners but protected as UCT intellectual property.

Contracts and compliance

There are also a number of legal and contractual hoops a researcher needs to jump through in a multi-institutional, multi-country study such as this. And it was an awareness of the complexity of the compliance issues she faced that prompted Ward to first contact eResearch.

Meyer and eResearch Director Dr Dale Peters advised Ward to work with Research Contracts and Innovation (RC&I) on a data sharing agreement between sites to prevent battles further along the



"We will be collecting data over a 20 to 30-year period, at least, so we need to be sure the software we are using in 2045 can read the data collected in 2021."

~ Professor Catherine Ward

line as to who owns the data; who has access to it; and who is responsible for its long-term preservation.

This, according to Andrew Bailey, senior manager at RC&I, needs to be dealt with right in the beginning at the research contract or consortium agreement stage of the project.

"These contracts will need to take funder requirements into account, different university policies and also different geographical locations as the legislation in different countries will differ," says Bailey.

Long-term data management

Working with UCT eResearch Ward says she could also plan for eventualities she would never have considered herself, particularly the issue of data decay.

"We will be collecting data over a 20 to 30-year period, at least, so we need to be sure the software we are using in 2045 can read the data collected in 2021."

This means the project needs to budget for software upgrades, translating data and a dedicated data steward position to ensure the data remains useable throughout the length of the study.

Offering a competitive edge

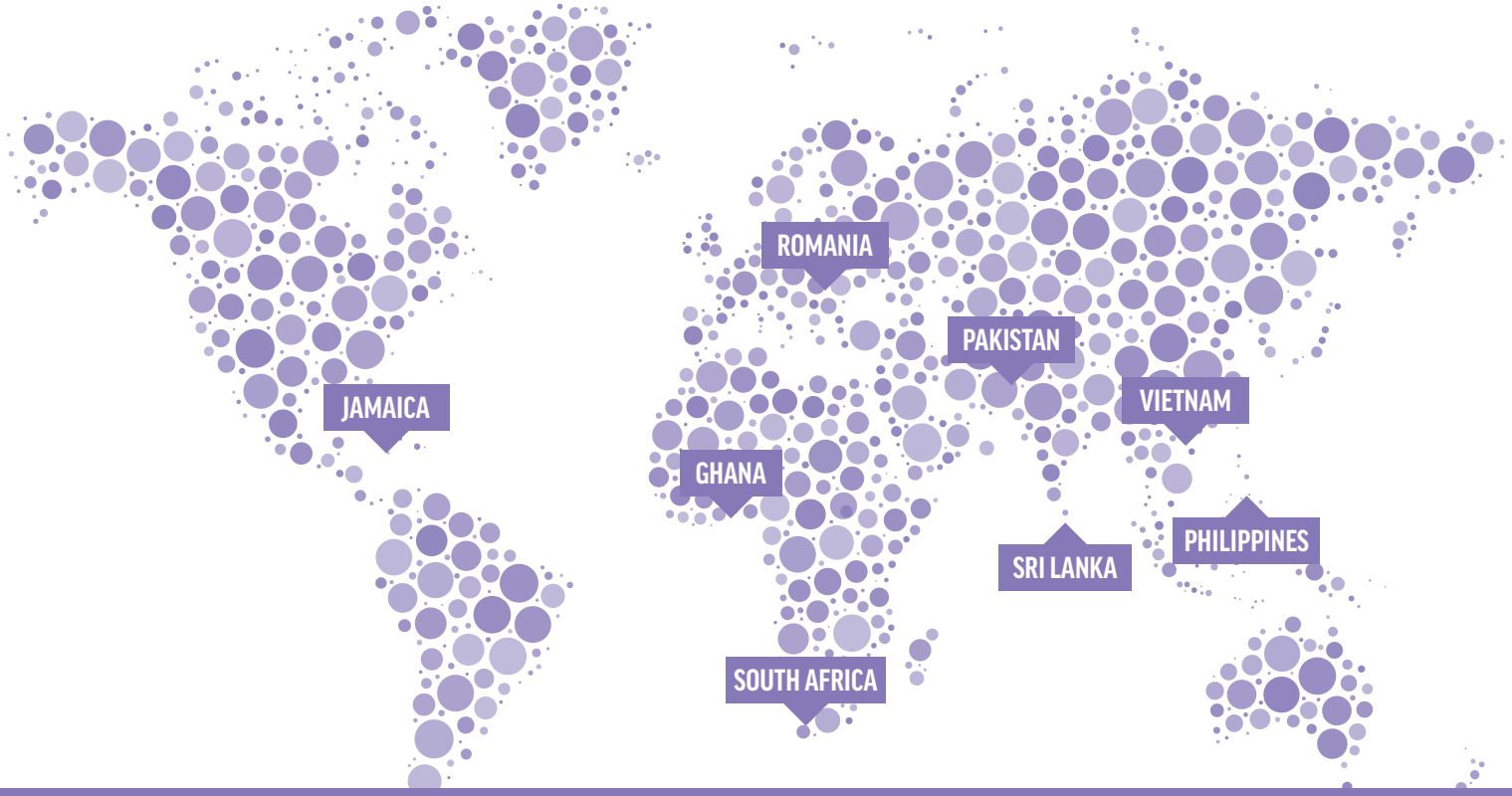
For Ward, the eResearch support she received in working through these complex data issues at proposal stage was a huge relief.

"I am a psychologist with no insight into many of the complexities and technicalities we have to grapple with around data management and compliance," she says.

Ward says she is grateful to UCT eResearch not only for the technical guidance and advice, but also for the legwork and time they put into helping her with her proposal.

"Having this kind of support," she says, "has given us a real competitive edge."

Below: The study plans to track 12 000 women across these eight lower-to-middle income countries around the world.



Tracking poisoning in South Africa

There were 15 poison information centres across the 54 countries in sub-Saharan Africa in 2019, compared with one in every major city in high-income countries. Of those 15, Cape Town's two poison information centres work together, with a little help from UCT's eResearch, to form the Poisons Information Helpline (PIH) of the Western Cape.

Based at the Red Cross War Memorial Children's Hospital and Tygerberg Hospital, these centres are the only ones to keep a database of incoming calls. The data collected from these calls helps researchers to understand poisoning in South Africa. UCT eResearch, through its ICTS arm, Enterprise Infrastructure Services (EIS), supported the PIH with a virtual machine to ensure they have access to the software they need to maintain this valuable database called the AfriTox Telelog.

"The PIH essentially provides the only national emergency poisons service in the country," says Dr Cindy Stephen, director of the PIH. "About two thirds of our calls come from health practitioners, and one third directly from the public."

She says they often assist doctors with particularly tricky poisoning cases, or those based in under-resourced rural settings who are on their own trying to manage a poisoned patient with limited resources or support.

The AfriTox Telelog is possible thanks to the volunteer work of a UCT alumnus, Dr Selig Leyser, who, back in 2014, designed a program to record telephone information; generate reports in real time; and provide immediate statistics.

The PIH still uses this program, but it relies on Filemaker software, and in order to upgrade their software, the PIH needed a Windows server on their machine.

"For assistance with this I approached ICTS for support," says Stephen. "We could not afford the expensive sever software so the ICTS technician recommended we opt for a virtual machine instead."

"A virtual machine is not dissimilar to a cloud-computing platform," says ICTS senior technical specialist Ashley Rustin. "Instead of putting all the software and data on the computer at the PIH, we at ICTS run the server in our virtual environment and then offer remote access to the staff at PIH through their local machines."

This takes care of the system administration, including keeping the software up to date, making the data more secure, and giving the PIH access to greater central storage facilities.

Stephens notes that since the Red Cross War Memorial Children's Hospital and Tygerberg Hospital merged their telephone services to create a 24-hour poisons emergency helpline in mid-2015 they have seen about a 10% growth year-on-year of calls received, as the PIH gains awareness.

> **"There is very little poisons data available from low- and middle-income countries in Africa," says Stephen, "this data set may well be the only one of its kind."**

Right: The Poisons Information Helpline (PIH) provides the only national emergency poisons service in the country. Many doctors in rural, under-resourced settings rely on the PIH to assist them in poisoning cases.


T
Giftig
Toxic
Toxique

Section 4

Publishing the data //

Technology offers new ways of not only acquiring but also sharing and storing research data, allowing for greater collaboration among researchers as well as more rigorous scrutiny.

Our goal at UCT, expressed in the Research Data Management Policy, is to ensure our valuable research data is published in a manner that is FAIR: findable, accessible, interoperable and reusable.

This section offers case studies of researchers who saw the value in the FAIR publication of their data for future reuse and greater collaboration.

Goosing around on golf courses

"It's not just golfers who enjoy spending a day on the course," says Dr Rob Little, manager of the Centre of Excellence at the Fitzpatrick Institute of African Ornithology. "Geese are also drawn to their luscious, irrigated lawns interspersed with artificial water bodies." Fortunately research by Little and colleagues suggest ways the geese and golfers can co-exist.

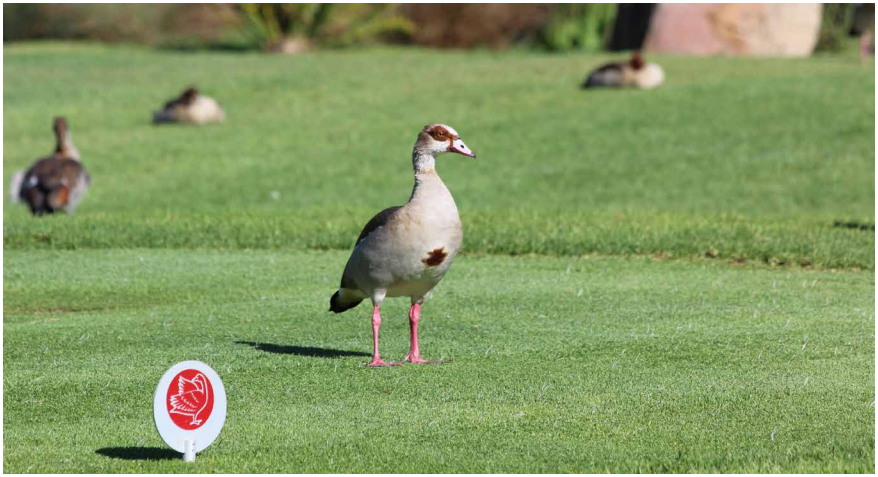
Geese on golf courses are a global problem, with suggested solutions ranging from blasting goose distress calls across the greens, deploying trained dogs to harass the offending geese or the display of garish scarecrows. None of which make for a particularly pleasant game of golf, or work effectively in the long term.

A team of staff and students at the Fitzpatrick Institute of African Ornithology studied the behaviour of Egyptian geese on Cape Town's golf courses and created a valuable data set which shows that geese seek out particular habitat features that make them feel safe.

"The best solution for both geese and golfers would be to design golf courses in a way that makes them less appealing to geese," says Little. "Then, as a secondary measure, falconry is effective at changing the perception of safety for the geese, which makes them less attracted to the golf course."

Little and his team published the data that supported these findings on UCT's institutional data repository, ZivaHub: Open Data UCT to encourage other researchers to reuse and build on this data.

Below: Egyptian Geese have invaded many golf courses in the Western Cape in large numbers and have become a nuisance. Photo by: Richard Gie.



"The more we understand the behaviours of geese and other birds that are perceived as pests," says Little, "the easier it will be to create non-harmful solutions to manage them."

Published data set offers a peek into the history of electrification in post-apartheid South Africa

South Africa's electrification project since democracy in 1994 has been a remarkable success story. Behind this success story lies a remarkable research endeavour with a number of roleplayers who worked together to collect and process the electricity use of about 15 000 households for a year-long period. The collection process took 20 years, and thanks to the work of research and data scientist, Wiebke Toussaint, at UCT's Energy Systems, Economics and Policy group, this data is now available for research use.

> **"This data set provides fascinating insight into how South Africans, primarily in rural areas and townships, use energy and how that energy use has changed in the post-apartheid era as living standards evolved," says Toussaint.**

The academic research outputs were driven by collaborators Professor Trevor Gaunt of UCT's Department of Electrical Engineering and Professor Ron Herman then at Stellenbosch University and later at UCT's Department of Electrical Engineering. The two worked with municipalities, South Africa's electricity public utility (Eskom), external consultants and the Council for Scientific and Industrial Research (CSIR) to collect this data set and used it to inform significant decisions related to the design of the country's electricity network.

When funding for continued data collection and management diminished, the closed data set lay idle for several years until Toussaint, recognising its value, worked with a range of stakeholders to publish the data in UCT's DataFirst facility, the only internationally-certified data repository (with a CoreTrustSeal endorsement) in Africa.

In mid-2019 the original data, together with two research-ready versions of the data set, were published by DataFirst for academic use only.

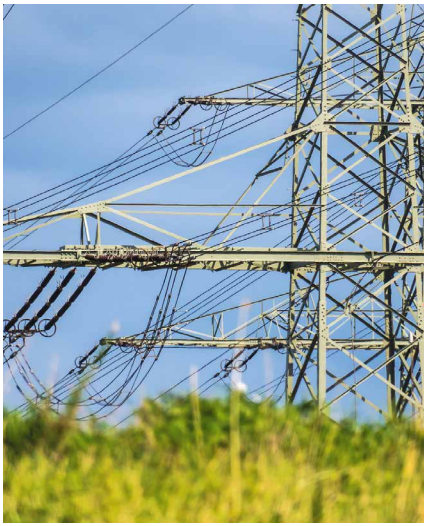
"The anonymised versions of the data can be requested online from DataFirst, while the full confidential data set can only be accessed through DataFirst's secure facility," says Toussaint.

"One of the most challenging aspects of this project was getting buy-in from the many stakeholders involved in the project over the years," she says.

> **"The CoreTrustSeal endorsement given to DataFirst made our work a lot easier as it gave the stakeholders the necessary confidence that sensitive, personal information would be protected."**

Toussaint's work was funded by the South African Energy Development Initiative (SANEDI).

Below: Since the dawn of democracy in 1994 in South Africa access to electricity went from 36% in 1994 to 90% to date.



Using ZivaHub to kick up dust

Dust feeds the Amazon and fertilises depleted oceans: feeding phytoplankton which in turn absorbs carbon emissions. We know that dust plays a vital role in the world's physical, biological and chemical cycles, but the sources of dust emission are less well understood.

Dr Johanna von Holdt, then a PhD researcher in the Department of Environmental and Geographical Sciences, was working to understand dust emissions from the Namib desert in Namibia, and published her data on ZivaHub: Open Data UCT as part of a drive to create a consolidated data set of dust emissions globally.

Von Holdt, who is now doing postdoctoral research in the Department of Chemical Engineering, spent about a month camping in the Namib desert, using a combination of satellite imagery and ground measurements from a portable wind tunnel to test dust emissions at selected sites.

"There are several researchers around the world doing similar field research, in the United States, Mongolia and Australia, among other places." says von Holdt. "But at this stage the research is still very siloed."

To try and break out of the silo, Von Holdt is suggesting the use of a standard protocol and classification scheme which, if followed by other researchers doing this kind of work, would allow for the merger of these varied data sets. In publishing her data openly on ZivaHub she is striving to set a standard for interoperability to enable research collaboration, and to create more powerful data sets.

> **"With a global data set we could come up with more accurate estimates of how much dust is actually emitted," explains Von Holdt. "And considering dust's importance in earth systems, a better understanding of dust emissions will have positive impacts on a range of models, particularly climate modelling."**

Right: Dust from the Namib desert plays an important role in the world's physical, biological and chemical cycles.

Below: Dr Johanna von Holdt measuring dust emissions in the Namib desert.

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

29 820

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

DOWNLOADS

5794

Number of times items were downloaded

NUMBER OF DEPOSITORS

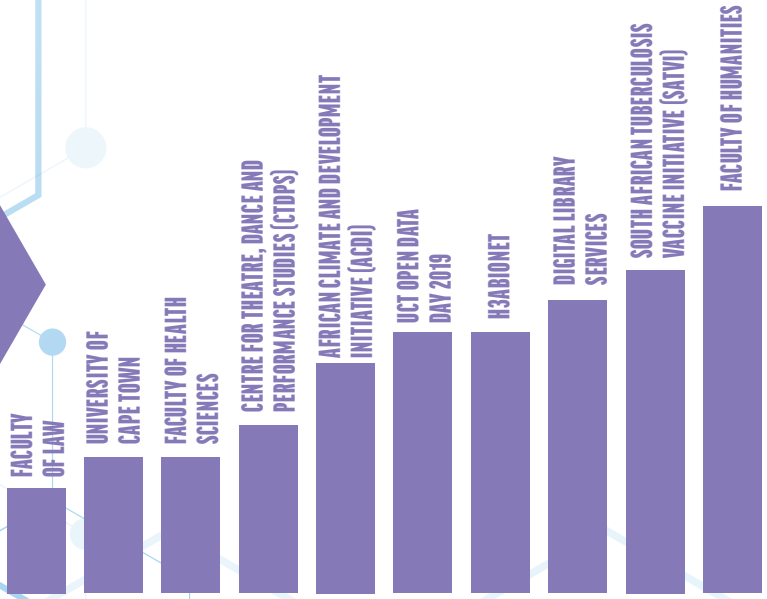
124

STORAGE

52.3 GIGS USED

TOTAL DEPOSITS

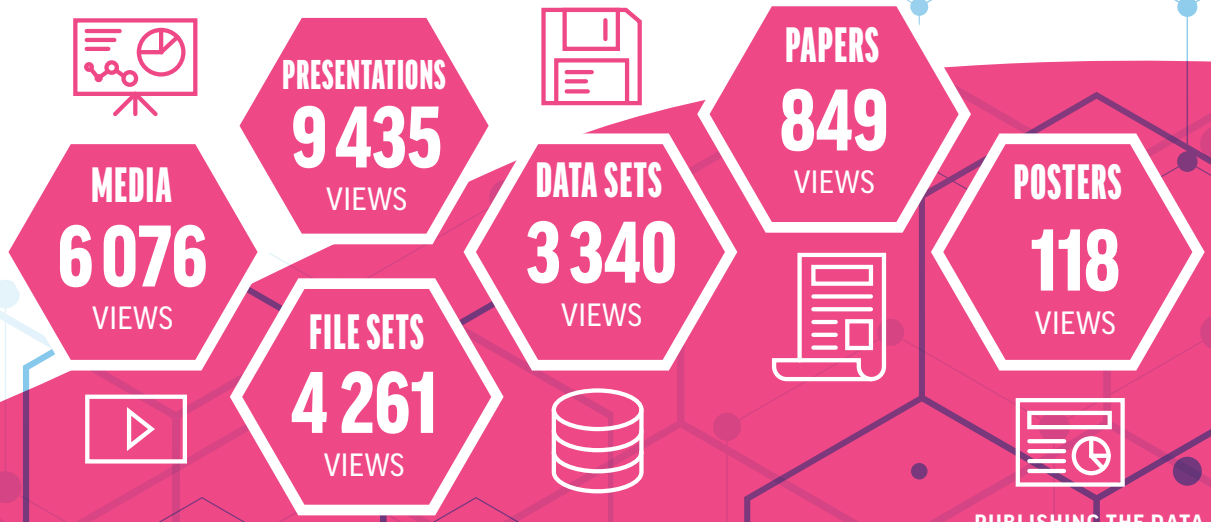
507



UPLOADS BY GROUPS

ZivaHub allows researchers and the institution at large to make research data FAIR: findable, accessible, interoperable and reusable.

TYPES OF FILES:



Section 5

Infrastructure to support our needs //

Information and communication technologies (ICT) have become the lifeblood of research. Acquiring, analysing, comprehending and visualising data on the scale at which it is being generated now requires new tools, innovative software and powerful hardware and storage.

The eResearch journey began with the establishment of the high-performance computing (HPC) facility in 2009. The ten-year anniversary of HPC facilities, as well as the recent investment in the upgrading of these facilities, is a mark of how far we have come. This section offers just two case studies showing how good ICT continues to underpin excellent research.

High-performance computing (HPC) upgrade complete to mark 10-year anniversary

UCT's new HPC cluster was installed and operational in January 2019 and continues to provide a reliable, scalable and economic computing facility for researchers. The upgraded facility provides an HPC cluster geared towards researchers with major computing needs but is also accessible to the full community, for both teaching and research.

Above: UCT's research data centre, housed in the basement of the ICTS building on Main Road in Mowbray. This is one of several data centres housed in different locations across the campus. The HPC facilities are housed in the Upper Campus data centre.

"June 2019 marked ten years since we started providing HPC services to UCT researchers," says Andrew Lewis, a senior technical specialist at ICTS. "In those 10 years we've run almost 26 million CPU hours, completed just under 3 million jobs and have been acknowledged in 139 publications that we're aware of."

The research infrastructure landscape has changed considerably since 2009, says Timothy Carr, a senior technical specialist at ICTS. "We no longer have to rely on hand-me-down servers and cobbled-together file services. Our new cluster has been running for six months and has seen a very good uptake of new researchers."

The new cluster is split into several partitions, featuring newer, faster nodes; graphic processing units; high-memory nodes; and the older, original cluster nodes now available for teaching and lower-priority jobs. It provides state-of-the-art 100GB Infiniband interconnect for both storage and inter-process communication as well as increased storage space for temporary processing.

Given the cost of electricity both for powering and cooling central processing units (CPUs) this affordable and environmentally-friendly computing resource is of great value to researchers.

The HPC team provide the following support for UCT researchers:

- Account creation on UCT Linux HPC infrastructure.
- Basic introductory training to get users started.
- Access to basic Linux skills and advanced HPC training.
- Installation of scientific codes (software applications).
- Access to fast IO scratch disk for processing of data.
- Access to archive storage.

UCT's high performance computing facility

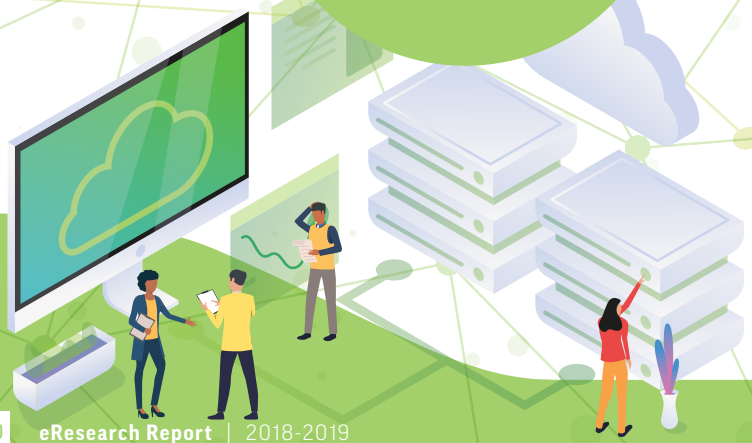
2018-2019

These numbers show updated HPC facilities – since the upgrade – and reflect the reporting period of June 2018 to June 2019. The UCT HPC cluster is geared towards researchers with major computing needs but is also accessible to the full community, for both teaching and research.

CORES
2 440

CITATIONS
24

STORAGE
470 TB



618 084
PROCESSED JOBS

USER UPTAKE
86

4.746
MILLION
CPU HOURS

GROWTH SINCE 2013

YEAR	CITATIONS	CORES	STORAGE (TB)	JOBS	CPU HOURS
2013-2014	19	1 142	169	985 216	4 163 686
2014-2015	26	1 458	219	152 690	5 997 849
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

Investing in a shared resource for scientific software sustainability in computational chemistry

Understanding the role of proteins and sugars embedded in cellular membranes, particularly on diseased cells, could be key to cracking the code in diseases like cancer or malaria. The action of multiple enzymes results in complex glycoproteins that decorate the cell. To gain insight into enzyme function, advanced computational expertise is needed to build and run molecular simulations. Dr Chris Barnett has invested in the research cloud-computing platform ilifu to create a central resource where a like-minded research community can come together and share these tools and expertise, using the open source platform Galaxy.

Barnett, a computational chemist based in the Chemistry Department and Scientific Computing Research Unit, has been studying the role of sugars in cells in specific diseases. Simulation of these enzymes is key to understanding them because different molecules will behave differently in particular contexts.

"Reusability and reproducibility are big buzzwords in science right now," says Barnett, "when you publish in journals these days they also want access to the data and the code to be able to confirm that it works."

It's very difficult to build software that other people can use, and that is sustainable in the long term. This then becomes a very complex problem for researchers.

> "Time and money need to be invested to keep scientific software in a good state, and these are resources most researchers don't have in abundance."

Software development and aptitude in the skills needed to install software or even to manage simulation data is a major barrier to entry.

Galaxy: the solution

Barnett believes the solution to these issues lies in a web-based platform called Galaxy where a user-community can form and share tools and workflows around enzyme modelling.

Galaxy is a platform for accessible, reproducible and collaborative science. Galaxy Europe hosts multiple tools including a cheminformatics subdomain which is a webserver for processing, analysing and visualizing chemical data, and performing molecular simulations.

> "It's a totally open and transparent platform," explains Barnett. "It provides workflows and histories so you can share it with other people and build up complex experiments."

The web interface is designed to prevent people from making easily avoidable mistakes like typos; this is critical as such errors can negate the value of an entire scientific study. Galaxy is 'batteries included': the provenance, metadata and choice of simulations parameters are readily accessible for review.

Ilifu: a home for Galaxy South Africa

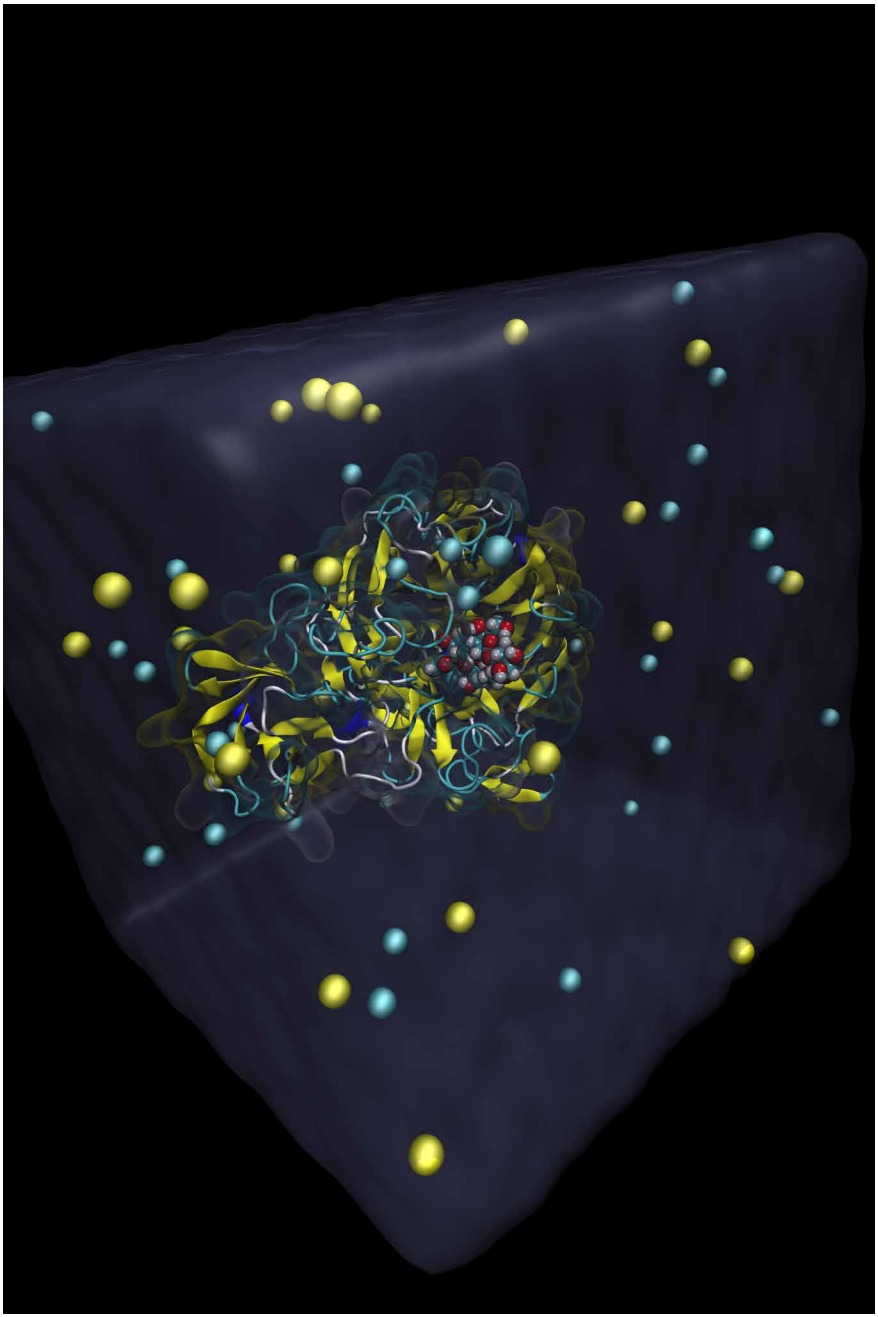
Barnett's end goal was to create a collaboration platform - similar to the cheminformatics subdomain of Galaxy Europe - for the community of glycobiology and computational chemistry researchers in South Africa, and hopefully in time, Africa. He held introductory workshops and found an enthusiastic group of researchers ranging from very interested novice users who have never run a simulation before, to advanced users who would find it valuable to repeat simulations on a platform like this. The trouble was finding a platform to house it. For Barnett, ilifu – a cloud-computing research infrastructure operated by a consortium of universities and research organisations in the Western and Northern Cape – was the obvious choice of a home.

The challenge was just how Barnett could access the infrastructure as it was built to first service the data-intensive fields of astronomy and bioinformatics. He did this by buying into the infrastructure with grants provided by both the National Research Foundation (NRF) and the UCT

Advanced Computing Committee (ACC).

For Barnett, the fact that ilifu is an infrastructure managed by researchers for researchers makes it preferable to commercial cloud computing platforms like Amazon Web Services or Google Compute Engine.

> **"It's a virtualised and flexible computing infrastructure where I don't need to worry about any of the hardware as the experts take care of that," he says. "In addition, unlike the commercial platforms, the costing is transparent. And I know the people behind the infrastructure, Andrew Lewis, Timothy Carr and Dane Kennedy from my previous work, and am confident I will get great support."**



Right: Cartoon representation of NanA sialidase (PDB:2VVZ) from Streptococcus pneumoniae solvated in water. A trisaccharide substrate (van der Waals representation) is bound in the active site of the enzyme with sodium and chloride ions present in solution.

Section 6

Looking back at our year //

Highlights from the year

It was an eventful year for eResearch and included the incorporation of new services, a new training model and a successful eResearch Africa conference in April, the fourth bi-annual conference hosted by UCT.

New services: 2018-19

The pace of development of new technological tools is impacting research practice in all domains, evidenced in the significant uptake of the new digital scholarship service, both in the support for digital humanities, and in the integration of geospatial data to address complex research questions.

Digital scholarship

Digital scholarship is simply the use of digital technologies for purposes of studying, teaching or research. Digital Library Services (DLS) offer a new service called digital scholarship to support our researchers and students with finding and using the right digital tools and processes.

"There are a lot of tools out there," says Sanjin Muftić, digital scholarship specialist at DLS, "to help you manage your files, help you work with others collaboratively, and analyse and visualise your data, among others. Developing your digital scholarship also means being aware of the processes and steps you can take to make your research lifecycle smoother, maximising the actions that happen with every keystroke and mouse click."

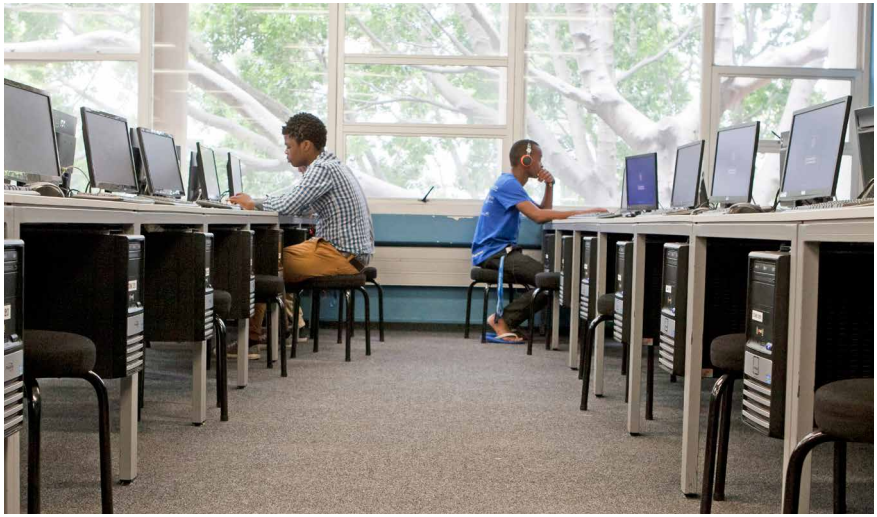
Support for digital humanities

Digital humanities is a field of scholarly practice that finds itself at the intersection of humanities research and digital technologies. This can be humanities enabled through digital media, artificial intelligence or machine learning, software studies, mapping and geographic information systems, or information design and modelling. As primary sources of information, especially within humanities, are more frequently digitised and available, online scholars invent new ways of looking at problems. Digital technologies introduce new methods, tools and frameworks to support scholars in finding novel innovative ways of enquiry, and visualisation techniques to query their data.

Geographic Information System (GIS) services

GIS at UCT Libraries serves as a support centre for all UCT staff and postgraduate students. While GIS mapping technology services have been available to the campus community for some years the service itself has now been integrated into DLS.

> **"Everything happens somewhere," says GIS officer Thomas Slingsby. "Capturing the location of a scientific phenomenon and using it to create interesting information involves analysing geospatial data, and here the GIS service is well-equipped to support trans- and interdisciplinary research."**



Right: New services offered through UCT eResearch can make the research process smoother by maximising the digital technologies. Image by Stephen Williams.

eResearch Africa 2019

Digital technologies have changed the way research is conducted globally. These technologies have enabled researchers to collect vast volumes of data which then need to be processed, analysed, managed, secured and stored for future reuse. The data scientists who wrangle that data contribute to real-world solutions for the world's wicked problems. In April 2019 UCT eResearch hosted its bi-annual eResearch Africa conference to explore the practice, principles and opportunities of data science in knowledge production and decision-making.

A meeting of minds across disciplines

This conference brought together IT practitioners, librarians, researchers and other interested individuals for two days of workshops and two days of presentations under the overarching theme Data Science for Development (DataScience4D). Richard van Huysteen, executive director of UCT's Information and Communication Technology Services (ICTS), described the conference as a "meeting of the growing community of practice that supports the way research is done in the current and future eras."

For Tawanda Chingozha, a development economics fellow at the Office of Astronomy for Development, the take-home message from the conference was the importance of this kind of interdisciplinary collaboration, both between researchers but also IT technicians, librarians and researchers.

> **"It became apparent to me during the course of the conference that whatever research data problem you may come across, somebody has either solved it or is working to solve it," he says. "So we need to be holding more conferences and meetings like these to break down the silos."**

Chingozha was awarded a scholarship to attend the conference and present his research on remote sensing data, machine learning and citizen science for development in sub-Saharan Africa. He says the conference helped him to realise he is not alone in the data-related challenges he has faced and was a valuable opportunity to share what he had learned on his research journey.

Europe's federated infrastructure a model for Africa?

The keynote address was given by Tiziana Ferrari, technical director of the EGI Foundation, the coordinating body of EGI: the federated e-infrastructure set up to provide advanced computing services for data-driven research and innovation in Europe and around the world.

Ferrari explained how the digital infrastructure that is today EGI was originally built to tackle the grand challenges of high-energy physics with

Left: Research Africa 2019 brought together IT practitioners, librarians, researchers and other interested individuals for two days of workshops and two days of presentations.

experiments in CERN. From here it has grown into a pan-European digital infrastructure supporting all research disciplines worldwide.

Rather than pour a lot of money into a single huge data centre at CERN in Geneva, European countries decided rather to invest nationally in a federated system so they would have IT experts locally to support their own researchers, explained Ferrari.

She said this could be a model Africa could use to boost its own research on the continent.

> **"Skills building is a very important topic at this conference," she noted. "It is essential for scientists to have IT experts available for them at universities and research centres as this kind of support is critical for data-intensive research."**

Above: UCT's fourth bi-annual eResearch Africa conference took place in April 2019, the focus of this year's conference was Data for Development.

Below: From the left: Keynote speaker, Tiziana Ferrari, technical director of the EGI Foundation, Dr Dale Peters, director UCT eResearch and Emma Kaye, then executive director of the Cape Digital Foundation

The next generation of data scientists

A key theme of the conference was the next generation of data scientists: education models from universities, summer schools, bootcamps and hands-on training programmes. This stream included presentations on topics as diverse as an online tool for genomics education and tackling the STEM shortage in our youth to astronomy as a platform for skills development.

Focus on development at eResearch Africa

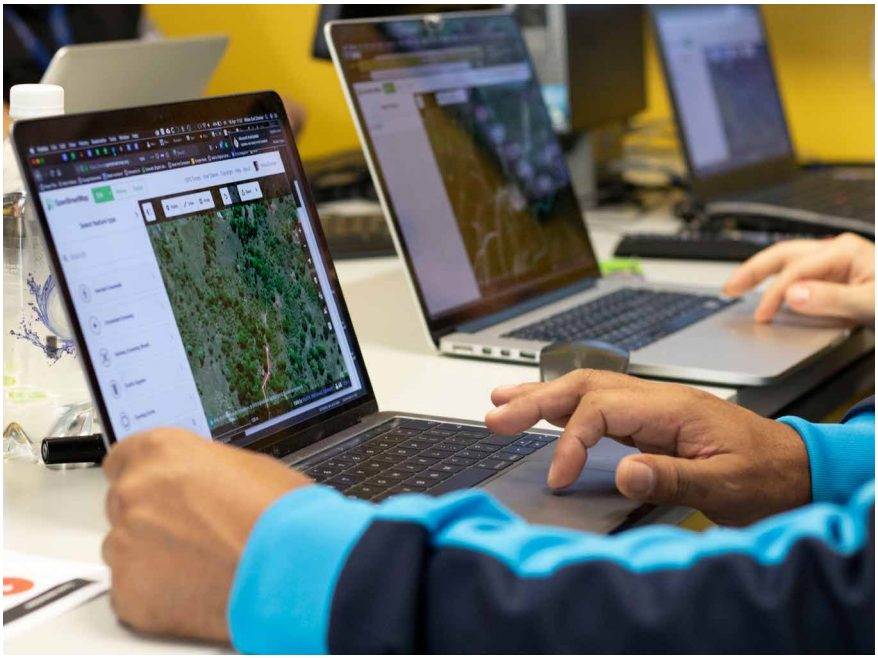
True to the theme DataScience4D the conference left its own positive impact. Instead of giving gifts to the speakers, the conference organisers donated instead to the Cape Digital Foundation, a local non-profit organisation working to create 'smart townships' in South Africa.

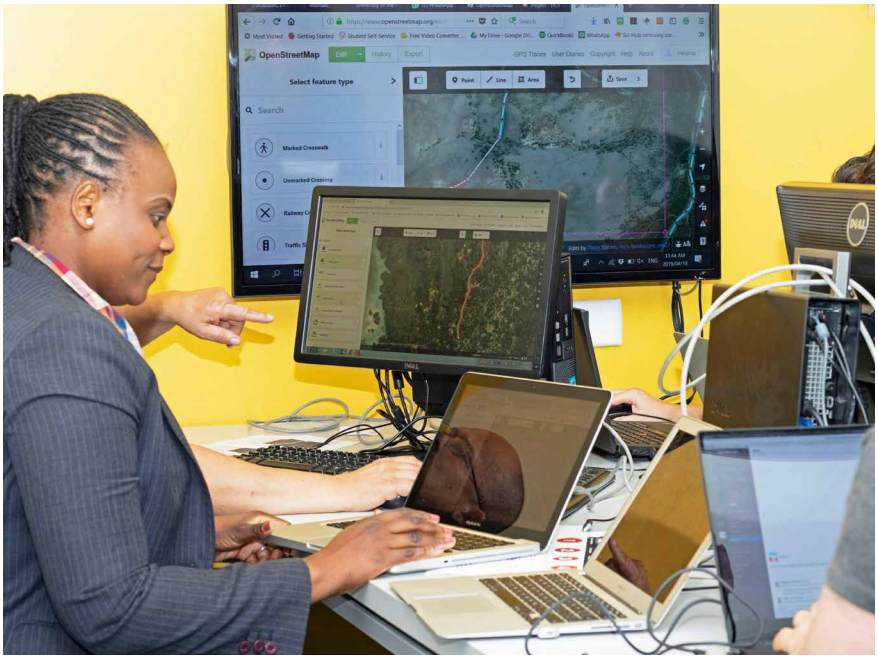
UCT eResearch, working with DLS, hosted a Mapathon for the duration of the conference. Through the Missing Maps project and Humanitarian OpenStreetMap Team, the conference delegates worked to use satellite imagery to support the disaster response operations of humanitarian relief agencies working in the cyclone-hit areas of Mozambique in the wake of Cyclone Idai.

Around 100 000 people are killed and around 200 million people are affected or displaced by natural disasters every year. Accurate maps are an integral part of emergency response planning in humanitarian disasters, as they allow ground crew to reach those affected.

At eResearch Africa 2019, delegates were invited to help remap the affected region with the use of satellite images and online GIS tools. Volunteers were able to make a meaningful difference in real time to the distribution of aid to those whose lives have been shattered by the cyclone.

> **"It was a great experience knowing my contribution, even if it was a small one, was not in vain but used to help those thousands of people in distress."**
>
> **~ Mark Hanmer**
> UCT staff member and Mapathon volunteer





eResearch seminar series: **coming to a department near you**

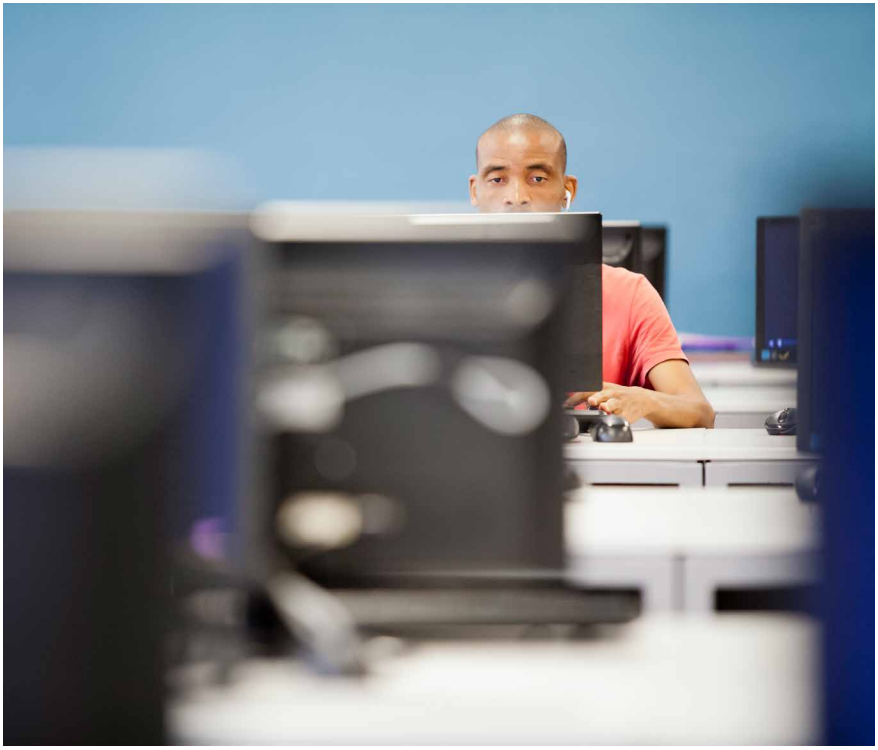
In 2019 UCT eResearch began to consolidate its training into focused seminar presentations covering a range of discipline-specific challenges faced by researchers.

"While UCT eResearch has always offered training in response to specific requests, this year we have worked to streamline our training into a mainstream service, tailor-made for specific faculties, disciplines and departments," says eResearch analyst Renate Meyer, who has been managing and coordinating the training.

"The seminar series usually comprises six sessions, developed in consultation with the relevant department to ensure we are meeting their needs." One example from the pilot programme included a collaboration with Professor Raj Ramesar, head of the Division of Human Genetics, to develop a series for postgraduate students and staff that addressed the research lifecycle specific to human genetics.

"There is great pressure on researchers and supervisors to not only stay on top of changes and new issues in their own fields, but also broad shifts in their institutional, national and even international research environments," says Ramesar. "It is a great relief that we can draw on eResearch as a resource to help us stay on top of these shifts, highlighting not only new demands but new support services."

UCT eResearch is running this collaborative training initiative in tandem with another eResearch project working towards consolidating training calendars across research support divisions. "Through these two projects we aim to firstly address the specialised needs of researchers who are working in a digitally demanding environment, and secondly provide a single platform for them to find training opportunities on campus," says Meyer.



Right: UCT eResearch is offering training to address the specialised needs of researchers working in a digitally-demanding environment. Photo by Stephen Williams.


UCT
eResearch
ACCELERATING RESEARCH