UCT
eResearch
ACCELERATING RESEARCH

eResearch
Report 2015-16

UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Contents

UCT lays the groundwork for **data-intensive science**

The first decade of this century has seen a tremendous advance in information and digital technologies impacting scientific inquiry.

Megaprojects in science and engineering, sensors tracking the state of the planet – the result is vast and complex collections of data that trace the patterns and trends in human behaviour. Now, these are beginning to be creatively mined, in ways that fundamentally alter our perception of the world and empower global change. Universities that are part of this data revolution will be globally competitive in the new arena.

UCT eResearch has gone from strength to strength since its establishment in March 2014. Under the guidance of interim director Anthony Beitz in 2015, and Dr Dale Peters in 2016, the centre has not only grown its repertoire of services as well as its 'customer base'; importantly, it has also forged strong partnerships, both nationally and internationally.

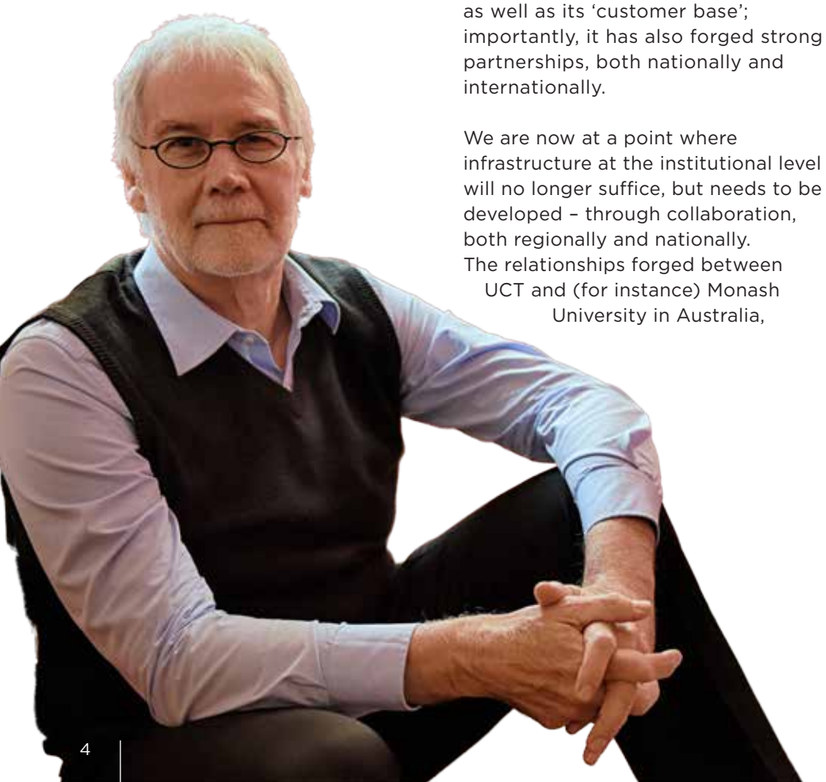
We are now at a point where infrastructure at the institutional level will no longer suffice, but needs to be developed – through collaboration, both regionally and nationally. The relationships forged between UCT and (for instance) Monash University in Australia, and the University of Oxford in the United Kingdom, are vital to ensure we remain on the cutting edge in our response to the big-data challenge. Local partnerships and collaborations, such as with North-West University, Stellenbosch University, and the University of the Western Cape, among others, ensure that nationally we are working together to provide the best possible solutions to benefit the country as a whole.

In the 2015-16 period under review, two other significant initiatives were launched that have become important players in UCT's overall big-data strategy: the Inter-university Institute for Data-Intensive Astronomy (IDIA), and a research-focused data centre: the first phase of the African Research Cloud (ARC).

All this groundwork has culminated in a UCT-led consortium, under the leadership of IDIA director, Professor Russ Taylor, and UCT eResearch interim director, Dr Dale Peters, being awarded a contract by the Department of Science and Technology to establish a Western Cape Data-Intensive Research Facility, as part of the department's National Integrated Cyberinfrastructure System (NICIS).

This facility is the beginning of a national drive to ensure that researchers have the advanced information technology (IT) support they require, and UCT is honoured to play a leading role in the early stages of what is set to be an exciting process of innovation for South Africa.

Professor Danie Visser
Deputy Vice-Chancellor:
Research and Internationalisation



Tribute to our **outgoing DVC**

In his last year as Deputy Vice Chancellor (DVC), the eResearch partners would like to recognise the foresight and vision Professor Visser showed in supporting eResearch, from a fledgling organisation in 2014 to its current position as a leader in eResearch expertise and infrastructure on the continent.

"If it were not for Danie's vision, eResearch at UCT would not have materialised. This vision not only allowed UCT to have the first eResearch centre on the continent, but set an example to other universities to follow – and therefore changed the face of research in South Africa forever."

Mr Sakkie Janse van Rensburg, Executive Director, Information and Communication Technology Services (ICTS)

"A few years ago, it was already evident to all how the very nature of research had changed, due to new technologies and the use of increasingly massive datasets. Fortunately, Danie astutely seized the opportunity to position UCT's researchers optimally within this new research space, by getting the support departments to work across the traditional silos of ICTS, UCT Libraries and the Research Office. His focus throughout was on how best to support the researchers within this new environment, enabling them to be at the helm of critical projects rather than simply producing the data."

Dr Marilet Sienaert, Executive Director, Research Office

"Under Danie's leadership, UCT eResearch has embodied the foremost research infrastructure in the higher education sector in South Africa today. He leaves our research community a vital legacy of collaborative support structures that are well integrated into the national cyberinfrastructure to accelerate global research."

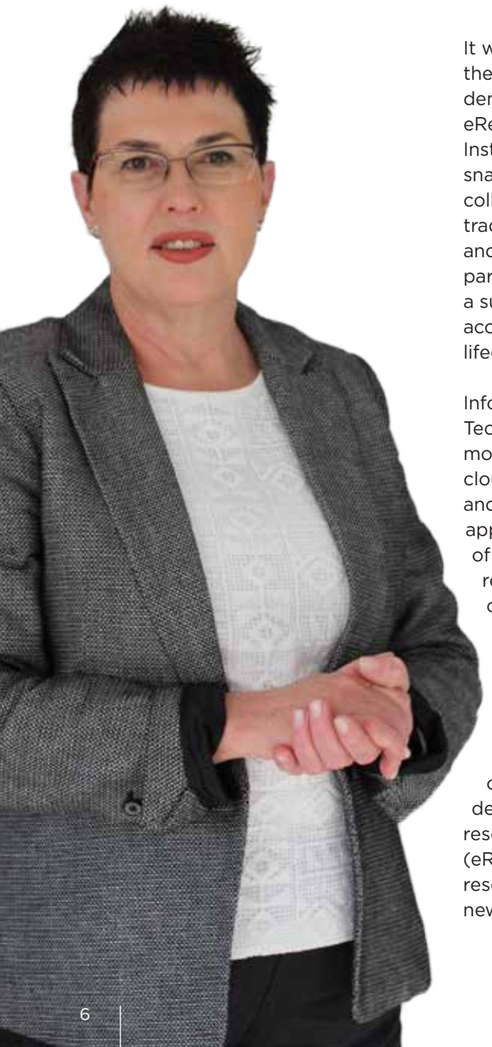
Dr Dale Peters, Interim Director, UCT eResearch

"Back in 2014, Danie had a vision for eResearch that filled us with excitement and expectation. In those early days he also supported us, as partner and player, though the ups and downs of the eResearch partnership. He saw the opportunities, and brought together a community to engage, deliberate and incubate ideas to grow this extraordinary capability – which has earned us, as a campus community, recognition as leaders in this arena."

Ms Gwenda Thomas, Executive Director, UCT Libraries

The realisation of **shared resources**

We are currently witnessing the rise of large-scale, distributed global research collaboration, along with greater application of the principles of openness: open access, open data, and open science. These changes mean a greater need for distributed and shared technologies that fundamentally alter the manner in which scientists carry out their work; in the tools, applications and workflows they use, and in the manner of communication arising from their collaboration.



It would be impossible to showcase all the projects and collaborations that demonstrate the special value that eResearch brings to UCT scholarship. Instead, this report provides a snapshot of our achievements and collaborative engagements that bridge traditional disciplinary boundaries and institutional silos. The eResearch partners have brought to fruition a suite of end-to-end services that accelerate and support the research lifecycle, from proposal to publication.

Information and Communication Technology Services (ICTS) has moved decisively into the era of cloud computing, or the storing and accessing of data and software applications over the internet instead of on a local server. This gives researchers the ability to develop collaborative research environments in which they can share data, computational capabilities and other resources, unimpeded by the restrictions of time and space.

The Research Office has contributed services in the deployment of the electronic research administration system (eRA); in hosting the emerging-researcher programme to introduce new approaches, in a workshop series on eResearch capabilities; and in an effective communications and marketing service, to disseminate the impact of technology-enabled research by the research community at UCT, both nationally and internationally.

UCT libraries have contributed an innovative open-scholarship service, enabling access to (and visibility of) UCT scholarship in the implementation of the open-access policy and of OpenUCT, the institutional repository. In late 2015, the libraries developed a research data-management service to assist researchers in maintaining research integrity – ensuring that relevant data is archived, accessible and citable, so that results can be verified and data reused in future.

Poised at the leading edge of data-intensive science, the future of UCT eResearch will continue to be shaped by the products of the excellent research conducted by our academics. I should like to take this opportunity to thank all our collaborators and supporters for sharing in that spirit of curiosity and innovation in this past year – we are equally excited by the road ahead!

Dr Dale Peters
Interim Director, UCT eResearch

Raising awareness for **eResearch support and services**

UCT eResearch was developed to help researchers deal with the data deluge they are increasingly compelled to face, at several points in the research lifecycle. Of course, this will only help if researchers are aware of the support on offer, and are provided with a forum in which they can be given guidance on how best to make use of it. Raising this awareness and providing such opportunities is one of the most important tasks for the Research Office, as a partner in the eResearch project.

A series of eResearch workshops has now been built into our researcher-development programme, structured around the challenges researchers may face at various stages of the research lifecycle, and the support provided by eResearch to help them answer those challenges.

The series was initially aimed primarily at emerging researchers, with the idea that by talking to researchers early in their careers, and – ideally – during the planning stages of their research projects, the trauma of suddenly being overwhelmed by the complexity and costs of managing their data could be prevented. However, the workshops have also been embracing researchers who are established in their fields, but new to data-driven challenges and interested in learning about eResearch, and who are finding that their disciplines are being reborn in the digital age.

These workshops are an excellent example of the work that has taken place across a number of departments in breaking down silos and streamlining the support they offer. This is most visible in the research support hub, a new website developed by the Research Office and launched this year, intended to provide a 'one-stop shop' for researchers. UCT eResearch is integrated into this hub – as is the new research portal, where researchers can create and manage CVs, draw on and apply for grants and funding using the new and powerful Converis platform, and create a publicly visible profile.

The Converis platform is the cornerstone of the electronic research administration (eRA) programme that is being introduced over the next two years, to streamline and automate research-management processes. Both the eRA and the developments in eResearch are indications of the rapidly changing research landscape, and the crucial need for UCT to rise to the challenges it presents. The Research Office is proud to play its part in this.

Dr Marilet Sienaert
Executive Director,
Research Office

UCT Libraries in the **digital age**

The UCT Libraries are at the heart of the academic community, providing support to students and researchers at every step of the research lifecycle. Our role is also to make eResearch services available to this community. Simply put, this involves making data visible, discoverable and accessible – not only now, but into the future. A hundred years from now, the validation, re-use and visibility of data will be central to the role of the libraries.

The digital age has fundamentally shifted the research lifecycle. In 2015, we upgraded our facilities, upskilled our teams, created an online presence and introduced new data-management services in order to better support researchers.

Increasingly, funding agencies require a data-management plan, and potential publishers want access to datasets to validate and confirm the results the researcher is claiming. This is the world we're moving into, and demand for eResearch support will grow. UCT's Research Strategy 2015-25 identifies the important role that the libraries play in providing digital services, including open access and digital scholarship. Much of the 2015-16 period has been spent designing a new organisational structure to meet these new researcher needs, without compromising our long-standing practices and services.

Innovation is a key part of this process. One such innovation, in 2016, was the refurbishment of the Hlanganani Junction, and the introduction of the visualisation wall. As a neutral space, the libraries can accommodate this new multi-purpose-designed area, which will facilitate researcher engagement, data visualisation and data management training for the campus community.

Digital Library Services team provides support services for digital scholarship, which involves the curation of a wide range of digital collections. We've also purchased legacy equipment – such as a video recorder and cassette players – so that we can migrate collections stored in old technologies (such as celluloid film) to digital formats.

Maintaining international standards is important, and we work closely with the Network of Data and Information Curation Communities (NeDICC) in South Africa to make sure our systems are institutionally accessible, interoperable and visible to researchers across the world.

Students and researchers everywhere in Africa should be able to access information, and we will continue to actively engage our partners through national organisations such as NeDICC, and international agencies including CODATA (the International Council for Science: Committee on Data for Science and Technology) and standards such as the RDA (Resource Description and Access) to achieve this strategic goal.

Gwenda Thomas
Executive Director, UCT Libraries

Strong partnerships are **strong foundations**

Looking back on the period from June 2015 to June 2016, it is clear how UCT eResearch transitioned – under the guidance of 2015 interim director Anthony Beitz – from a planning phase into an implementation phase; while in 2016, with the fledgling unit under the directorship of Dale Peters, we are witnessing a period of stabilisation and growth.

In this process, we have seen clearly how eResearch has moved out from under the ICTS umbrella to become a university-wide effort, as the need for more advanced information technology (IT) support for researchers becomes more widely acknowledged. In this 12-month period we have seen a substantial increase in the number of of researchers seeking support from eResearch, which is indicative of a growing confidence in this centre from the research community.

The success of this process of maturation is testament to the support offered by UCT Libraries, under the guidance of Gwenda Thomas, and the Research Office, led by Marilet Sienaert. The active support from these two executive directors in the eResearch partnership meant that despite the challenging financial times faced by the university, eResearch could continue to flourish and grow. In the year ahead I am confident the three key eResearch partners – ICTS, the Research Office, and UCT Libraries – will remain committed to further developing our eResearch capacity across the university.

Partnerships continue to remain key to the eResearch initiative, and in the 2015-16 year we worked to maintain existing relationships and forge new links, both locally and abroad. In late 2015, we hosted the Director of the Oxford University e-Research Centre, Professor David De Roure, strengthening ties with that university.

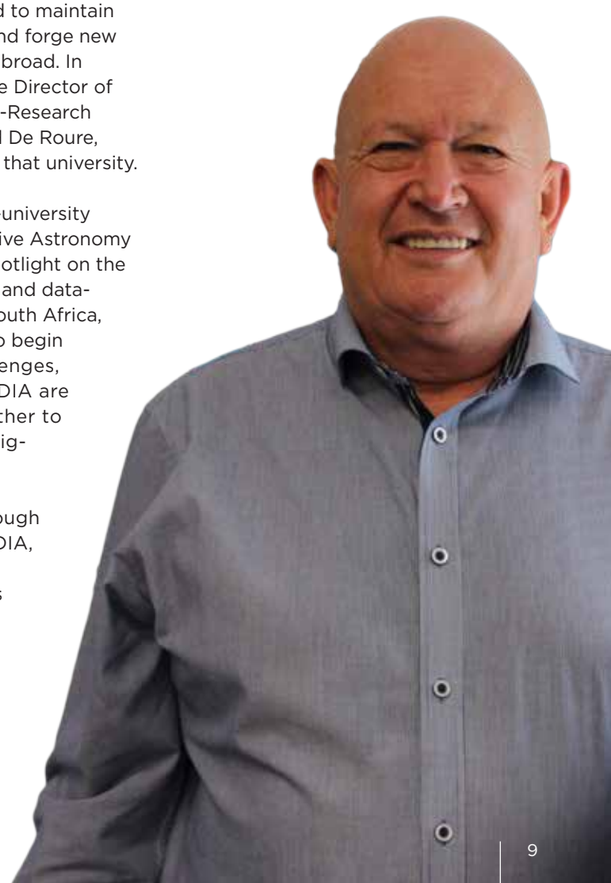
The launch of the Inter-university Institute of Data-Intensive Astronomy (IDIA), really put the spotlight on the challenges of big-data, and data-intensive research, in South Africa, Africa and the world. To begin to address these challenges, ICTS, eResearch and IDIA are actively working together to find solutions to the big-data questions.

In this period, and through our partnership with IDIA, we began to build a network of universities across South Africa, which began with a memorandum of understanding (MOU) with North-West University.

This MOU was the first step towards the development of the African Research Cloud (ARC): a highlight of the year for eResearch and ICTS.

The forging of these partnerships – within the institution, locally and abroad – signifies the readiness of the UCT research community to rise to the challenges of the big-data era.

Sakkie Janse van Rensburg
Executive Director,
Information and Communication Technology Services (ICTS)



Big developments in big data: **How astronomy is driving data science in Africa**

The Square Kilometre Array (SKA) astronomy project will produce data at a rate comparable to that of global internet traffic. But if we don't have the infrastructure and skills to deal with it, the data will go offshore; Africa will lose this stellar science and business opportunity. In September 2015 the Inter-university Institute for Data-Intensive Astronomy (IDIA) was formed, with the express goal of ensuring that Africa is able to meet this challenge – with benefits for the continent that go far beyond astronomy.

"The data revolution is set to be a globally transformative phenomenon – if you don't ride the wave, you're going to be flooded by it," says Professor Russ Taylor, director of IDIA, the culmination of a partnership between UCT, the University of the Western Cape, North-West University (NWU) and the University of Pretoria.

Big-data refers to the large, complex data sets – created and collected through technology – that are set to affect every aspect of life. Even by the standards of this world of big-data, however, the SKA poses a particular challenge. Tasked to collect data from deep space dating back to the very start of the universe 13 billion years ago, the SKA will collect around 1.5 exabytes of data a year – that is, roughly one and a half billion gigabytes.

Image courtesy SKA South Africa

South Africa, as co-host (with Australia) of the SKA, is thus uniquely placed to lead the global response to big-data – an opportunity we dare not miss.

UCT Vice-Chancellor Dr Max Price notes that while we have the geographical advantage of the southern skies, we don't – yet – have an advantage when it comes to analysing the data. "The risk is that we may become servants of scientists around the world, and not a team and a country that can generate its own knowledge and play in the big league."

Preparing for data sharing

It is this precise situation the IDIA seeks to avoid. The SKA big-data challenge begins with the MeerKAT radio telescope, which will make up about 1% of the total SKA project. MeerKAT will begin generating data in early 2017, for the phase known as 'early science'. That capacity will quadruple the following year.

"At IDIA, we are essentially laying the groundwork – in terms of both infrastructure and human resources – to be ready when the SKA turns on," says Taylor.

The real challenge, explains Taylor, is not just to build a big pipe to manage the data, but to store it in a way that enables the global collaboration required for a project of this magnitude. "Teams in Africa, Europe, Asia, Australia, and North America all want to work together on this data. So the issue is not only how to store and manage the data, but how to enable collaboration on a big-data set that nobody can actually have on their desktop," he says.

Africa Research Cloud

The first of these platforms is the Africa Research Cloud (ARC), the first phase of which is housed in UCT's cloud-based data centre, launched in the same week as IDIA. A second ARC node is now in place at IDIA partner NWU.

Cloud computing, which simply means storing and accessing data and programs over the internet instead of on a local computer or server, is already revolutionising the business and research world. The ARC gives researchers the ability to develop collaborative research environments in which they can share data, computational capabilities and other tools, unimpeded by the restrictions of time and space. UCT's distributed big-data research cloud will be prototyped by the IDIA partner universities, in collaboration with South African organisations such as SKA South Africa and the Centre for High Performance Computing in Cape Town.

The ARC is envisioned to grow to include the eight African partner countries on SKA, and a number of SKA partners in Europe. Private companies Dell and Canonical have also worked on the development of the ARC. In time, the ARC will service the larger research community, allowing researchers to share data and collaborate on a number of other 'big science' projects, such as bioinformatics.

eResearch and IDIA

UCT eResearch is a key partner to IDIA, supporting the institute in terms of both expertise and infrastructure. eResearch is integral to IDIA's strategy to build data capacity for the SKA in South Africa in two particular areas: the development of innovation in computing architectures for data-intensive science; and the development of software platforms to provide web-based access to the data and services researchers will require to analyse and process the huge data sets from anywhere in the world.

"Members of the eResearch staff are core to our cyber-infrastructure research and development programmes," says Taylor. "Two of the IDIA software development team were hired through eResearch, and have been seconded from eResearch to IDIA. This hiring process was entered into deliberately, to formalise the partnership between IDIA and eResearch."

Taylor says that, in time, the collaboration between IDIA and eResearch will be useful for other data-intensive research domains, such as bioinformatics and climate-change research.

"As we move forward, other researchers will begin to benefit from the innovations that arise from this collaboration."

Future perfect: **Data-intensive science in the Western Cape**

Image: Stephen Williams

UCT, through the Inter-university Institute of Data-Intensive Astronomy (IDIA) and UCT eResearch, is to lead a consortium to build South Africa's first regional data node of national cyberinfrastructure.

By early 2017, the MeerKAT telescope – part of the larger Square Kilometre Array (SKA): a multi-radio telescope project – will begin to produce massive data sets that need to be processed and mined for science by researchers around the world. The completion of MeerKAT will mark the beginning of exponential growth in radio-astronomy data over the next decade. Astronomy is not alone in facing the challenges of big-data. In bioinformatics the growth of rapidly advancing gene-sequencing technologies offers similar challenges.

The scale of the infrastructure required to respond to these data challenges is beyond the capabilities of a single institution, or even a group of institutions; it requires a national response.

National Integrated Cyberinfrastructure System

The Department of Science and Technology (DST) – in recognition of the size of infrastructure investment needed in data-intensive research to ensure that South Africa continues to play a significant role in cutting-edge scientific research across disciplines – has developed the South African National Integrated Cyberinfrastructure System (NICIS) to support and advance national strategies for science, socio-economic impact, and global competitiveness.

The mission of the NICIS is to enable, promote and support research innovation and knowledge creation. To this end, a tiered data infrastructure is to be established as part of the overall strategy to support data-intensive research in South Africa. Tier 1 is national-level infrastructure; Tier 2 is regional; and Tier 3 institutional.

NICIS Tier 2 Node awarded to Western Cape

In the first half of 2016, UCT, the University of the Western Cape (UWC), the Cape Peninsula University of Technology, Stellenbosch University, and SKA South Africa agreed to put forward a proposal, in response to a call by the DST, to form a consortium to build the Western Cape Data-Intensive Research Facility (DIRF) as the first 'node' in the NICIS Tier 2 data infrastructure. Since then the partnership has grown to include the newly established Sol Plaatje University in the Northern Cape.

The DST describes these data nodes as remotely accessible (which means researchers can access the data and software from anywhere) cyberinfrastructure that forms part of a greater network of similar nodes. This infrastructure will enable not only data storage but also data sharing, and will allow for easy research collaboration.

The UCT-led consortium, through IDIA and UCT eResearch, proposed to establish and operate a data-centric high-performance computing facility for data-intensive research.

The proposal leveraged the already considerable investment made in data-centre capacity by the Information and Communication Technology Services (ICTS); the expertise of eResearch teams at UCT; and the investment in infrastructure and personnel for technical research and development by the IDIA partners.

The proposal was successful; in the second half of 2016, the DST awarded the contract to the Western Cape consortium.

"DIRF will be a platform for developing innovative approaches to research with big-data that will enable South African researchers in astronomy and bioinformatics to compete with the best in the world," says project leader Professor Russ Taylor, who is a SKA research chair at two of the consortium universities (UCT and UWC), and director of IDIA.

"The expertise of the eResearch team will be critical to the implementation and sustainability of the Tier 2 facility research operations and support," says eResearch interim director Dr Dale Peters.

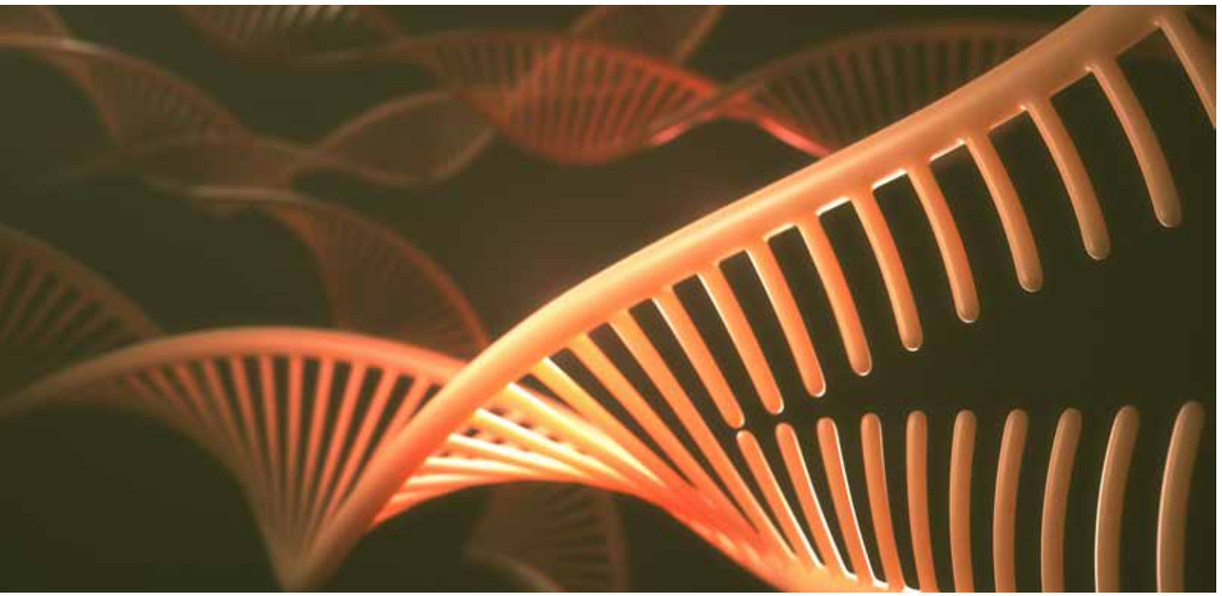
The winning of the award is only the start: a great deal of work lies ahead for the consortium, as they prepare to put systems in place to build this first regional data node in the national integrated cyberinfrastructure.

The Western Cape Data-Intensive Research Facility

As currently proposed, the Western Cape (WC) Tier 2 DIRF will have at its core a data-centric computing system that will provide computing and data storage for both data-intensive research projects and technical research and development programmes. Other elements of the WC DIRF include training and human-capital development – already, the consortium partners offer a number of postgraduate qualifications in the fields of data science and digital curation – and the development of new and the strengthening of existing international partnerships.

"It is gratifying to know that UCT has been able to band together successfully with partner organisations to be part of the solution to the many data challenges our researchers face," says deputy vice-chancellor for research and internationalisation Professor Danie Visser. "It is important to remain mindful of the fact that, as daunting as the big – or even massive – data challenge may be, it means we will in time be able to reach depths of scientific understanding never previously imagined possible."

Virtual machines enhance the **peer-review process**



Researchers in all disciplines are using advanced technologies to collect massive data sets previously unimagined. But collecting the data is only the start. It needs to be interpreted, and when it comes to data analysis a picture is indeed worth a thousand words. eResearch is working on solutions to allow researchers to look at their data-sets in a way that is visually pleasing and easy to understand.

In 2014, the journal Nature Methods noted that a growing proportion of original research submitted to them relies on custom algorithms and software developed by the authors. Noting that these algorithms and software are an important methodological component of the manuscript, the journal published detailed guidelines for authors to include software or source code in their submissions. Over a year later, the same journal published an editorial outlining the challenges inherent in reviewing papers that are "computational at heart".

"Too frequently though, our guidelines notwithstanding, papers reporting a computational method are initially submitted without either [software or code]. Separately, referees quite frequently have trouble running the provided software, which they may report to us during the review process or at the end of it. Both of these problems can lead to delays, or even the rejection of a paper."

It was with these challenges for the peer-review process and publication in mind that a team of researchers in the BTRI, led by Mhlanga, collaborated with UCT eResearch to deploy a virtual machine, which allowed referees access to both the computational code and the data central to their methodology.

"In a way, a virtual machine is a precursor to cloud computing," says Dr Jason van Rooyen, eResearch analyst. "Instead of having individual computers in a laboratory with all the required software and data loaded, we at UCT eResearch run a server in our facilities that offers remote access to the software and data to anyone who needs it, be it collaborators or – in this case – referees."

The paper in question was submitted to the journal *Cell* as a resource article, the purpose of which is to highlight significant technical advances.

Benefits of using **a virtual machine**

- Quick to deploy, because no hardware needs to be built or ordered
- No maintenance for the researcher, because it is part of the eResearch infrastructure and is maintained there
- Scaleable, so can be made bigger or smaller depending on needs
- Researchers get access from anywhere, through any device
- Cheaper in the long run, because no actual machines need to be bought or replaced

In their research, Mhlanga and his team developed a completely novel approach to the measurement of spatial patterns of RNA on a sub-cellular level.

"RNA (ribonucleic acid) is known as the 'origin of life' molecule. It is believed to predate DNA and protein," says Mhlanga. Part of his research is to understand the spatial distribution of RNA in cells.

"This spatial distribution – while very poorly understood – is important for several cellular processes, including development, immune response, and learning and memory," he says.

So the group embarked on a systems biology approach, to try to understand how RNA is spatially localised.

"RNA (ribonucleic acid) is known as the 'origin of life' molecule. It is believed to predate DNA and protein," says Mhlanga. Part of his research is to understand the spatial distribution of RNA in cells.

"To do this we developed a mathematical approach, combined with imaging and microscopy, which led to fundamental insights into how cellular organelles influence the spatial distribution of a broad class of RNAs," he explains. Because of the novel tools used to gain what Mhlanga calls "a provocative new insight", it was critical the referees could access the data and run the code.

One challenge is the requirement for anonymity of referees, which for Mhlanga is why the collaboration with eResearch was so critical. "We gave anonymous accounts for access to the virtual machine to *Cell*, which meant referees could assess the code and the data, comfortable that their anonymity would be maintained," he explains.

"This step towards allowing referees easy access to review code and to run software has the potential to add a new depth to the peer-review process, and help ensure more reproducible science," says van Rooyen.

Opening the libraries: **Open scholarship and digital library services at UCT**

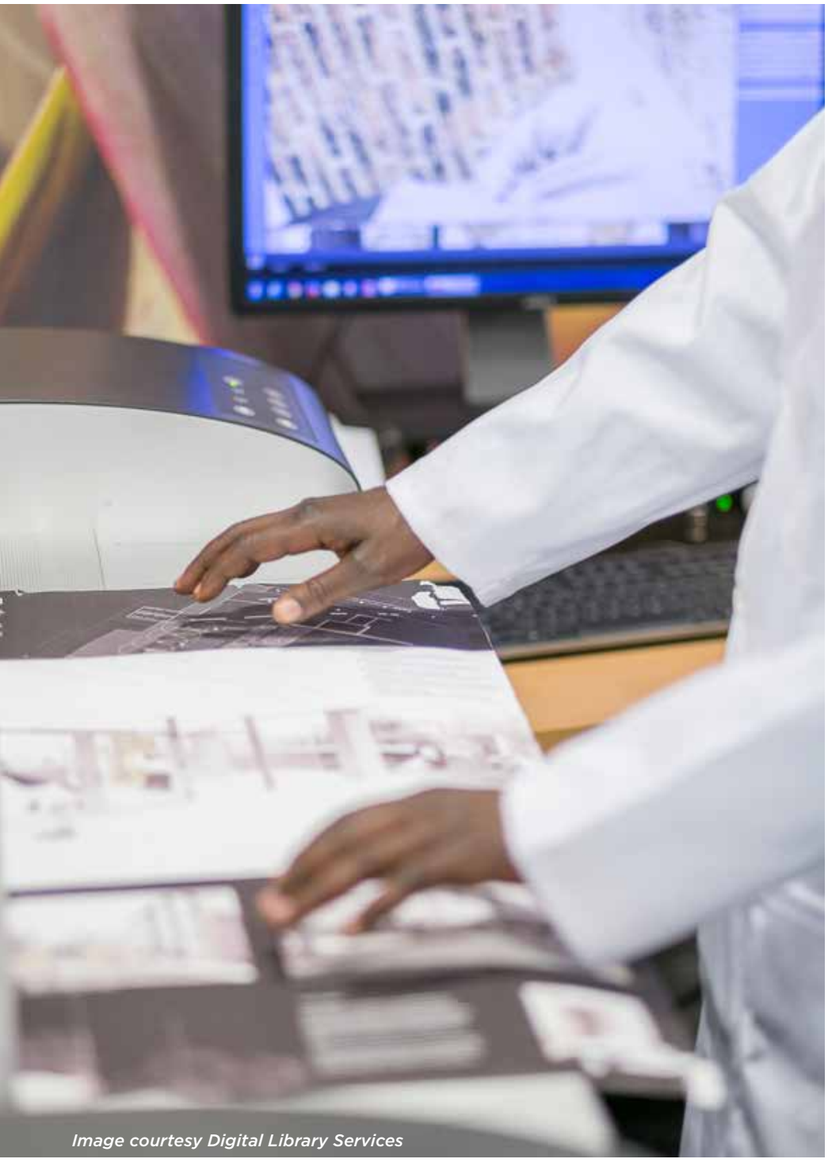


Image courtesy Digital Library Services

"Digital technologies have created more than one revolution. Let's call this one the access revolution," writes Peter Suber in the seminal book *Open Access*. For UCT, the access revolution is now entrenched in the UCT Libraries.

Between the newly formed unit, Digital Library Services, and UCT's OpenScholarship service – which encompasses the institutional repository, OpenUCT, and open publishing services – the libraries are ticking every box to ensure UCT's scholarship is searchable, findable, accessible and re-usable; from digital conversion and research data management, all the way to open-access publishing.

UCT passed its open-access policy in June 2014, which mandated the UCT Libraries to implement this policy. The libraries' focus extends beyond just research outputs but also on the research data that directly support or substantiate published research findings, to allow for greater accessibility, discovery, re-use, preservation and sharing of UCT scholarship.

"Research needs to be publicly available; because more often than not, it's publicly funded, and should be for the public good," explains Jill Claassen, access and visibility manager at UCT Libraries. It is for this simple reason that there is a global shift towards greater openness: a global shift that offers a payoff to ordinary citizens who would not otherwise be able to access that knowledge.

Research data management services

"New technologies are changing the very nature of how research is conducted – how researchers collaborate, and even how research publications are reviewed," says Dr Dale Peters, eResearch interim director and deputy director in the Libraries.

UCT Libraries, through its partnership with UCT eResearch, seeks to respond to this challenge of openness at every point of the research lifecycle. This means upskilling the staff to be able to respond to researcher needs, and implementing the infrastructure – such as online repositories, tools, and customised software – to overcome the challenges in our South African context.

Digital Library Services

"Digitisation, research data management and digital curation are very much overlapping services," explains Niklas Zimmer, head of Digital Library Services (DLS) at UCT Libraries. The digitisation unit within DLS offers digital conversion, project management, and curation and preservation services for a variety of audio-visual, photographic and paper formats, to ensure long-term preservation of and access to digital collections. This goes significantly beyond conventional scanning services, stresses Zimmer, as digital objects are created and curated according to international archival preservation and access standards.

"Special collections materials, such as thousands of negatives of biological specimens, or audio files of ethnographic interviews, are primary research materials," explains Zimmer. "These materials need to be carefully and consistently curated to ensure that they are not only correctly preserved, but can also be found and re-used by researchers in the future."

According to the Digital Curation Centre (DCC), digital curation involves "maintaining, preserving and adding value to digital research data throughout its lifecycle". This active management of the research data ensures the data's long-term research value, and makes it available for further high-quality research. It is here that DLS seeks to support the research community in the choice of data planning, equipment, naming conventions, file formats, metadata standards, and research-data collaboration tools – in line with international trends and best practice.

Open access and OpenUCT repository

In January 2015, OpenUCT – the UCT open-access institutional repository – became a service of the UCT Libraries. OpenUCT makes available and digitally preserves the scholarly outputs produced at UCT, including theses and dissertations, journal articles, book chapters, technical research reports, and uniquely, open educational resources.

"One of the things that makes our open-access policy different," says Claassen, "is that we are interested in all scholarship, including open education resources." This makes UCT scholarship all the richer, as anyone around the world can see what UCT is doing in terms of teaching and learning.

"One of the things that makes our open-access policy different," says Claassen, "is that we are interested in all scholarship, including open education resources." This makes UCT scholarship all the richer, as anyone around the world can see what UCT is doing in terms of teaching and learning."

This strategy has also paid off in other ways: in the Ranking Web of Repositories of July 2016, OpenUCT was ranked 7 out of 20 in South Africa, and 8 out of 73 repositories in Africa. (See 'OpenUCT Rankings' below.)

Library as publisher service

The great coup for libraries in the 2015-16 year was open-access publishing, says Claassen. UCT Libraries now offers an open-access publishing service that includes open-access publishing of journals and monographs (books).

The first journal to be published in the Libraries was *UR@UCT,* a journal of undergraduate research at the university. This was followed by three more journals. Without a pause for breath, the team in the libraries began to publish monographs – the first, a pilot, was *The Quest for Deeper Meaning of Research Support,* on the changing role of librarians. This was followed by four more monographs. the open-access publishing service is free to both reader and author.

"We are very proud of our institutional repository," says Claassen, "but it is in open-access publishing that I really believe UCT is taking the driver's seat. We are the first institution on the continent to publish open-access monographs. Here, we have pushed the boundaries of openness, and of the services in which the libraries are involved. And very soon we will be cutting our teeth on publishing open-access textbooks."

"UCT's open-access policy has maximised the visibility of UCT scholarship," adds Peters. "In addition to this momentous step, the research data management policy, currently under consideration, will transform the way research is conducted at UCT; by accelerating discovery, increasing the value of research decision-making, and catalysing innovation for social and economic development.

"The direct response of the libraries has been remarkable in developing leading-edge services that support the research community in achieving this goal."

OpenUCT rankings

January 2015

IN SOUTH AFRICA	IN AFRICA	IN THE WORLD
Ranked **9th** out of **22** repositories	Ranked **12th** out of **57** repositories	Ranked **470th** out of **2 154** repositories

July 2015

IN SOUTH AFRICA	IN AFRICA	IN THE WORLD
Ranked **8th** out of **22** repositories	Ranked **10th** out of **65** repositories	Ranked **423rd** out of **2 275** repositories

January 2016

IN SOUTH AFRICA	IN AFRICA	IN THE WORLD
Ranked **8th** out of **22** repositories	Ranked **8th** out of **70** repositories	Ranked **370th** out of **2 297** repositories

July 2016

IN SOUTH AFRICA	IN AFRICA	IN THE WORLD
Ranked **7th** out of **20** repositories	Ranked **8th** out of **73** repositories	Ranked **495th** out of **2 275** repositories

DMPonline @ **UCT**

The submission of a data management plan (DMP) is increasingly becoming a requirement in the international research community.

Many funders now require a DMP at one or more of the following stages:

1. grant proposal;
2. initial project implementation;
3. before final review.

It is recommended that a generic DMP is used, as a point of best research practice, even when it is not mandatory.

DMPonline was developed by the Digital Curation Centre (DCC) as a tool to enable researchers, data managers and principal investigators to work collaboratively in completing their DMPs.

UCT Libraries manages its own instance of DMPonline, so that your DMP information is stored and managed locally. This also allows UCT to create changes to DMPonline that are useful to our local user community, and to respond timeously to requests for uploading specific funder and institution templates. DMPonline at UCT currently also provides two unique institution templates:

- A UCT-generic template with relevant UCT guidance and links to UCT resources;
- A UCT-Clinical Research Centre (CRC) generic template, developed with data managers in clinical research at UCT.

For more information about research data management (RDM) and DMPonline, please visit the Digital Library Services website.

Boosting research with high-performance collaboration


Image by Stephen Williams

Developing software to support the new Square Kilometre Array (SKA) and probing the structure of conjugate vaccines are two of the diverse fields that drive the work of the Department of Computer Science's Associate Professor Michelle Kuttel.

Kuttel's research in computational science involves high-performance computing (HPC), which uses simulations and intensive calculations to answer scientific questions.

"I need to use high-performance computers for my collaborative research – whether it be to help astronomers develop more efficient searches for astronomical bodies in SKA data, or to explain clinical results by building virtual models of vaccine components, and then investigating their structure and dynamics," explains Kuttel.

The key is HPC

Together with innovative algorithms and analytical methods, Kuttel relies on the exceptional speed of HPC to power her work. In particular, the specialised 'mini-supercomputer', incorporating graphical processing cards, enables Kuttel to get results more quickly, and to solve problems that are simply not tractable on an average personal computer.

"These computers work so much faster than normal computers. I'm doing in a year what would normally take 10 years to do," says Kuttel.

Kuttel was one of the first researchers at UCT to build and use parallel computer processors, but she concedes it was a challenge.

"Years ago, a donation of around R250 000-worth of machinery landed in my office. It sounded wonderful at first; but I had to house the machines, organise power, and install and maintain them. As an academic working in scientific computing, being the systems administrator as well as the researcher is a bit of a nightmare," says Kuttel.

Centralising HPC capacity in eResearch

But UCT has found a way to free up time for researchers like Kuttel; they can now focus completely on their research, while UCT eResearch's dedicated HPC team takes care of the systems administration.

"It's been fantastic," says Kuttel. "They are extremely helpful. eResearch supports the scientists in many ways – from assistance with the purchasing process, to installation and maintenance of the machines and software. They solve technical problems very rapidly."

HPC specialist Andrew Lewis says the team members are on hand to set up systems and follow up with help.

"A large part of our job is administering the cluster – checking on security loopholes, new versions of software, installation of software packages, and maintaining the system."


Image by Stephen Williams

"We could easily double the size of the cluster, and we'd use it. We just need to ensure that the output is commensurate with the investment. But we're tracking citations carefully, and so far we're doing very well."

He says ICTS had been considering HPC facilities since about 2010. It also needed to get the researchers on board.

"The idea was to get a central unit, staffed by professionals and trained specifically in HPC – and then convince sceptical researchers that we could support their research."

The initiative took off. ICTS received seed money to buy hardware, and then worked with researchers to expand on this. The unit needed buy-in from researchers generally, as the infrastructure is expensive; it's also unlikely that a single research group would able to use it all the time, and to its full capacity. Today the HPC team are located under the eResearch umbrella.

Sharing the resources

Kuttel spotted the opportunity, and channelled research funding into helping to buy a shared resource. While it's configured to give the lion's share to Kuttel's research, other researchers are also able to use it.

Kuttel's collaboration with eResearch HPC has worked very well, and is the sort of cross-pollination that the university is keen to encourage further.

For Lewis, working in partnership with researchers has been very rewarding. "It's good to be at the forefront of research, and to know that we are able to give researchers a stable, well-maintained computing facility."

HPC facilities widely used across disciplines

It's also good to know how research such as Kuttel's can be boosted through HPC. In fact, the HPC is used across faculties and research fields at UCT.

In humanities, people can visualise ancient ruins and monuments. In oceanography, researchers can do ocean-temperature modelling. The mathematics department has a project monitoring sardine stocks. Civil engineering is modelling heart valves, and using HPC to design landmine-proof vehicles. HPC is also being used to develop vaccines, and to investigate cancer.

Lewis and his team offer training to researchers as well as back-up; and with interest growing in HPC, it looks set for an exciting future.

"We could easily double the size of the cluster, and we'd use it. We just need to ensure that the output is commensurate with the investment. But we're tracking citations carefully, and so far we're doing very well," he says.

Seeing the data
eResearch's first steps to a visualisation service

Researchers in all disciplines are using advanced technologies to collect massive data sets previously unimagined. But collecting the data is only the start. It needs to be interpreted, and when it comes to data analysis a picture is indeed worth a thousand words. eResearch is working on solutions to allow researchers to look at their data-sets in a way that is visually pleasing and easy to understand.

Images courtesy Zamani Project

Oculus Rift

Image courtesy of Occulus Rift

You can't help but feel a palpable sense of excitement as you move through the narrow passage between two great walls of rock, knowing that just around the corner is one of the world's most remarkable archaeological sites: the ancient city of Petra, in Jordan. As you come around the corner, the intricacy of the exterior, carved out of rose-red rock, is striking in its beauty. You walk up the steps and through the great archway, to enter a massive room that may once have been a marketplace where traders hawked their wares, as far back as approximately 300 BC.

This whole experience takes place from an office chair in the Menzies Building on UCT's Upper Campus, with an Oculus Rift strapped to the head. In fact, even if you were to travel to Jordan to visit Petra, you would only be able to stand outside to admire the ancient building, as this is a restricted site, with access granted to only a handful of scientists, officials and dignitaries. Fortunately, the Zamani group were given full access to the site when they visited it as part of their larger project to record, map and create three-dimensional (3D) models of the world's heritage sites, with a particular focus on those in Africa.

As part of his master's degree under the supervision of Professor Heinz Ruther (principal investigator of the Zamani Project), Stephen Wessels used Unity 3D game-engine software to take the digital 3D models the group had already created of Petra, and create a walk-through environment of the archaeological site.

"On a two-dimensional platform, viewers are limited to looking only at what is displayed on the screen; but wearing the Oculus Rift, you can look around and see all there is to see," says Wessels.

Some might spot the enormous camel standing on the left as you ascend the steps to enter the building. Others may be more interested in the intricate carvings on the outside of the building, and want to examine this in closer detail. Or perhaps the experience of standing inside Petra looking up at the beautiful red rock out of which the city was carved is so breathtaking, you want to stay in a single spot for 10 minutes. The Oculus Rift allows for this full range of experience. Watch the video.

With the Petra project behind him, Wessels is now working (as part of his PhD) to create a fully immersive display of San rock art – one that offers an additional educational element, with a description of each piece of art popping up when a viewer looks closely at it.

The Zamani group's hope is that they will not only preserve the continent's heritage sites, but open up access to everyone, all over the world.

The Iziko planetarium goes digital

The Iziko planetarium is set to enter the next era of technological development, thanks to a partnership between UCT, the Department of Science and Technology, the University of the Western Cape, and Cape Peninsula University of Technology, among others.

Up to mid-2016 the much-loved planetarium still sported its original 'Star Projector' technology that dates back to the early 1980s. But this is about to change as the Iziko Planetarium Digital Upgrade Project gets going, culminating in a state-of-the-art digital full-dome immersive theatre facility by early 2017.

This full-dome digital technology has become the norm for planetaria worldwide. There is also a move towards using planetaria for more than just edutainment; they can be data-visualisation facilities for researchers. The full-dome capacity will make the rendering and visualisation of big-data possible – providing a revolutionary tool for scientists in diverse fields to navigate through their data (especially large data sets) and interrogate it simultaneously. It will also contribute to developing skills to respond to both local and global digital transformation challenges.

By combining motion with 3D (effectively simulating 4D), researchers can 'fly' virtually through multi-dimensional visualisations, opening up whole new avenues of exploration in everything from planetary and solar-system science to geology, oceanography, climate and earth science, medical science, molecular chemistry and biochemistry, and even town planning.

Image courtesy of Iziko Museums

Hlanganani Junction's visualisation wall

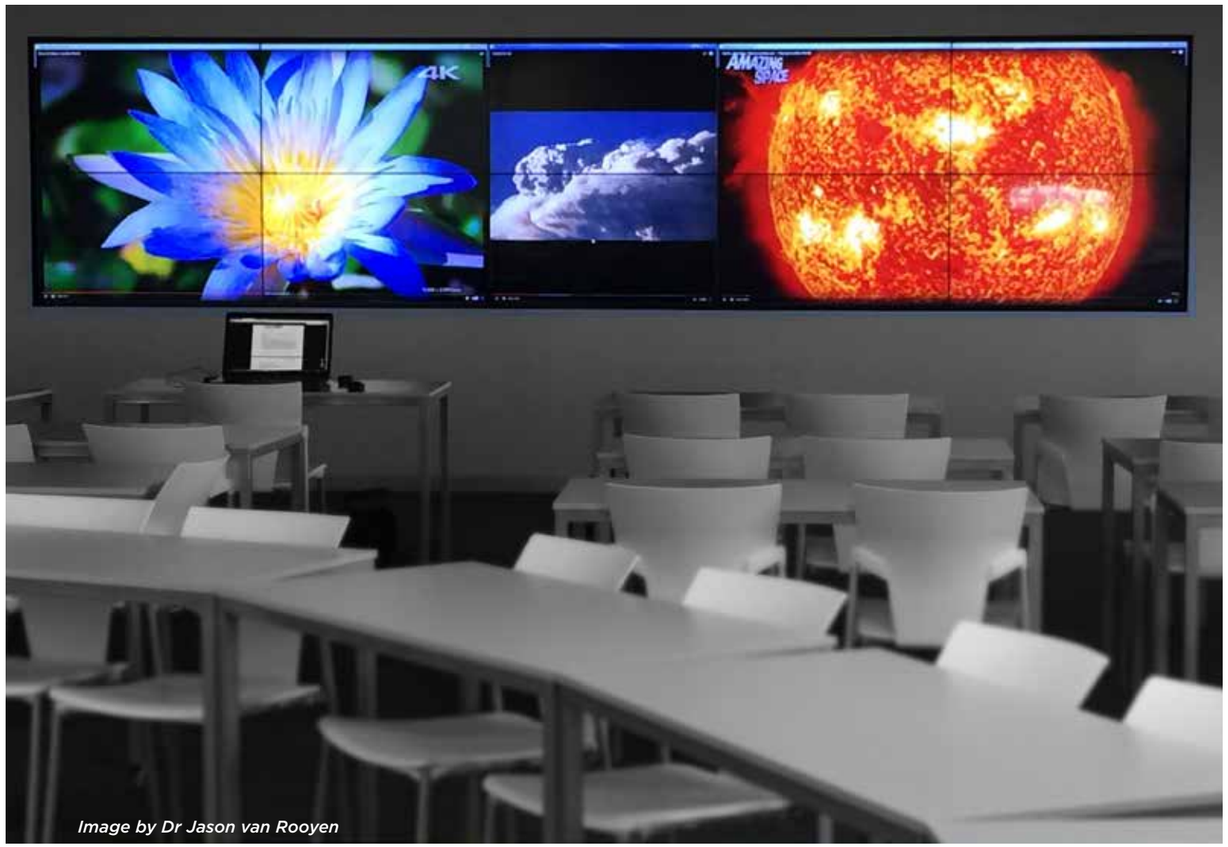


Image by Dr Jason van Rooyen

UCT eResearch, in close collaboration with UCT Libraries, has recently launched a visualisation wall in the Hlanganani Junction in the Chancellor Oppenheimer Library. This space – which boasts a 20-million-pixel research video wall – is to serve as a collaboration and engagement space for researchers in the libraries.

Researchers using the space will have access to Sage 2 collaborative software which allows researchers to bring their own devices – phones, tablets, laptops – and display their data on the video wall.

The software allows multiple researchers to display their data, extracted information, visualisations and animations at the same time on a very large-format screen, and to use their devices to interact with the screen displays.

"This space can be used by researchers for interrogating data visualisations or having lab meetings, and will allow them to have joint poster sessions, display architectural diagrams, and even host collaborative art displays and music events," says eResearch analyst Dr Jason van Rooyen.

Accessing our past: **UCT eResearch and the Five Hundred Year Archive**

The Five Hundred Year Archive project, led by DST/NRF SARChI Chair in Archive and Public Culture, Professor Carolyn Hamilton, seeks to stimulate research into the neglected eras of the southern African past before the advent of European colonialism.

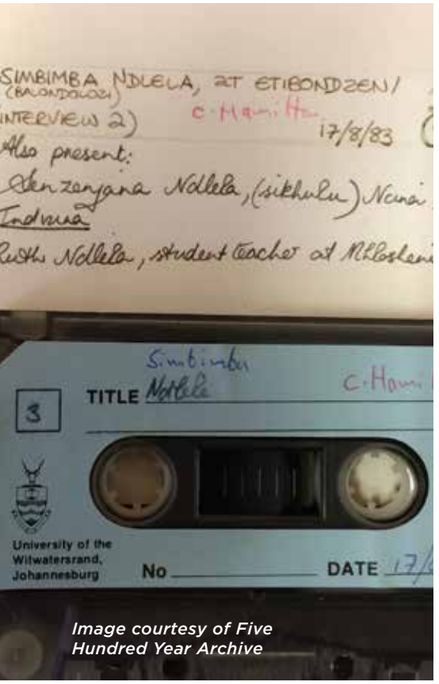


Image courtesy of Five Hundred Year Archive

Research in this area is hampered by the absence of archival material.

Researchers can make use of non-textual materials, yet much of this material concerning the remote southern African past is misidentified, often undated, lost, or dispersed in institutions across the world. This is a challenge that the Five Hundred Year Archive (FHYA) project aims to address in order to stimulate and facilitate further interest, research and enquiries into the southern African past.

One of the outputs of the project is a digital archival exemplar capable of virtually convening visual, textual and sonic materials, located across the world, relevant to the last 500 years of the southern African past. As an exemplar it is not an archive that will exist in perpetuity in its own right, but rather a sample or prototype designed to show what is possible.

UCT eResearch has been working with the FHYA project team to develop the exemplar.

"We faced a number of technical difficulties; and it was here that the assistance of eResearch was invaluable," says Dr Grant McNulty, postdoctoral fellow of the Archive and Public Culture Research Initiative, through which the FHYA project is managed.

One concern, for instance, was the multi-institutional nature of the FHYA project: the group at UCT partnered with a number of South African institutions, including the KwaZulu-Natal Museum, Historical Papers at the University of the Witwatersrands, and the Campbell Collections of the University of KwaZulu-Natal, as well as a number of international institutions including the University of Cambridge.

"In order to correctly represent our partner institutions and their collections and contributions on the FHYA website and portal, we could not have a domain name ending in .uct.ac.za," explains McNulty. "But we needed the website to be hosted cost-effectively and securely within UCT."

One challenge was the configuration of the unique archival software required to work across the multiple media formats. The FHYA seeks to take a wide range of material and bring it into a single searchable framework to allow researchers and the general public easy access to the materials. In order to do this, the FHYA project identified unique archival software called AtoM (Access to Memory), as the best software solution. AtoM, built with support from the International Council on Archives, is open-source software, which means not only that it is free, but also that the users are able to modify it according to their specific needs.

"eResearch was able to host the website for a minimal cost, install and configure the AtoM software, and also help us with a number of other technical issues such as centralised data storage to back up important project materials and setting up a file transfer protocol to make data sharing easier among the partners," says McNulty. "Our project now has in-house support in terms of software, digital storage and website hosting and security."

eResearch offers a range of **training options**


Image by Natalie Simon

"Training in the use of new technologies is the on-ramp to data-intensive research," says Dr Dale Peters, interim director of UCT eResearch. "The need to equip staff and students to become data analysts is critical to research success."

eResearch emerging researcher series

These seminars, run through the Emerging Researcher Programme in the Research Office, are spread over eight weeks, and seminars follow the research project lifecycle. The first seminar – 'What is eResearch?' – explains the various services and support eResearch offers throughout the research lifecycle. Other topics include tips on proposal planning and helping researchers cost their technical requirements for a research project; data gathering; research collaboration; data analysis; data visualisation; data management; and open-access publishing.

High-performance computing (HPC) training

The HPC team at eResearch offer both basic and advanced training programmes in the Linux operating system and using the HPC facilities, as well as domain-specific courses.

- Beginner: The basic course provides an introduction to Linux, the operating system through which the HPC facilities are run. In this course you will get an overview of what kind of HPC facilities the UCT eResearch offers, learn the difference between a CPU (central processing unit) and a GPU (graphics processing unit), and learn how to submit jobs to the HPC scheduler.
- Advanced: This course runs through more advanced concepts in HPC, such as interactive job submission and parametric jobs; as well as offering insight into software and storage.
- Domain-specific: In 2016, eResearch ran an on-demand course for postgraduates and postdoctoral fellows in the human genetics field of research. eResearch partnered with domain experts to focus on domain-specific training. This is a service available to groups that can muster a minimum number of participants.

Scientific computing: Software Carpentry workshops

Since 1998, Software Carpentry (today the Software Carpentry Foundation) has been teaching researchers around the world the computing skills they need 'to get more done in less time, with less pain'. In July 2015, UCT eResearch co-hosted their second of many planned SWC workshops.

The SWC workshop offered researchers a grounding in the Python and R programming languages, as well as version control, and basic shell commands to allow them to conduct data analysis through software that is both freely available and commonly used to improve the reproducibility of data analysis.

"Training in the use of new technologies is the on-ramp to data-intensive research," says Dr Dale Peters, interim director of UCT eResearch.

Newsbytes

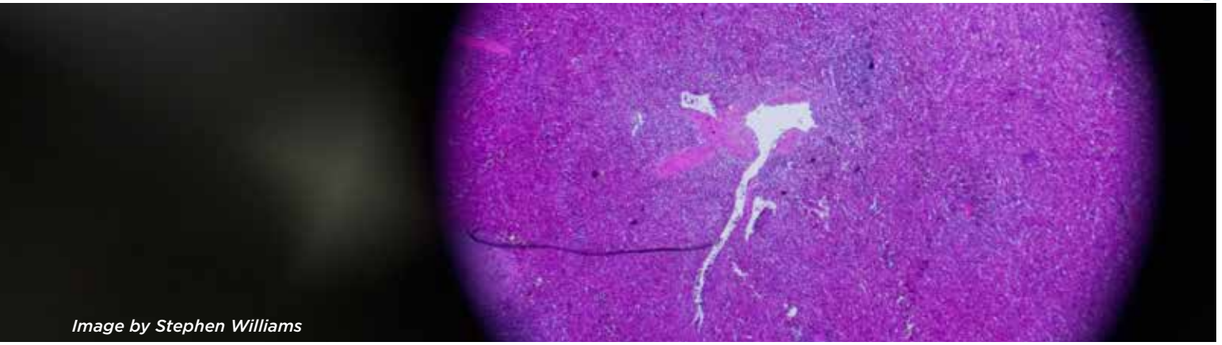


Image by Stephen Williams

Supporting clinical research data management at UCT

Effective clinical data management is key to generating high-quality, reliable and statistically sound data from clinical trials and other studies. It is the process of collecting, cleaning and managing clinical data in compliance with regulatory standards. A researcher does not want to run the risk of having work disregarded because of missing or inaccurate data.

A number of software tools are available for clinical data management, each with its own strengths and weaknesses. This is why the Clinical Research Centre (CRC), in collaboration with eResearch, supports a number of clinical research data-management software tools.

UCT eResearch works with the CRC to support end users of the software. “The eResearch team mainly do the back-end work to install and maintain the systems, while the CRC provides front-end support,” says Annemie Stewart, data analyst at the CRC. “If a system has a bug, for instance, eResearch will contact the developer, install a newer version, or otherwise help resolve the issue. But when a user needs help navigating the system, creating a project within the system, we help from the CRC side.”

She says the partnership with eResearch has been of great value to the CRC: “eResearch is always willing to help find solutions to unique situations that researchers may face in their different projects.”

Automating data management **in drug discovery**

“Processing data takes a long time,” says Ronnett Seddon, screening technician for the Drug Discovery and Development Centre (H3D), who recently enlisted the help of UCT eResearch to automate parts of her data management process, reducing tasks that would have taken her an entire day to one or two hours.

Seddon runs biological tests, known as assays, to screen compounds from various research projects. “On any given day I run around 200 assays,” she says. “Each run has at least 30 plates, which represent around 210 compounds. Each compound has at least 12 data points. So that is a lot of data.”

In order to process the data, Seddon must upload it to an online database. But before she does that, she has to convert the data into the correct format to be uploaded. Previously, Seddon had to copy and paste information repeatedly into a spreadsheet to obtain pre-processed data to feed into the online database, a laborious and time-consuming process with a large margin for error.

Using Python, a programming language popular in the scientific community, the eResearch team wrote a script to automate that manual pre-processing step.

For Seddon, the eResearch intervention has made a world of difference. “This automation has resulted in a minimum of 75% of my time saved.”

The UCT eResearch team

Back row (l-r): Timothy Carr – Senior eResearch Technical Specialist | Niklas Zimmer – Manager, Digital Library Services Andrew Lewis – Senior eResearch Technical Specialist | Chris Mtshengu – Research Support Specialist | Gwenda Thomas – Executive Director, UCT Libraries | Coline Laubscher – Business Analyst | Heine de Jager – Senior eResearch Technical Specialist Dale Peters – Interim Director, eResearch | Jess Oosthuizen – Senior Communications Officer, Research Office David Aikema – Data Analyst | Erika Mias – Data Curation Officer | Elizabeth Moll - Librarian | Adrianna Pinska – Data Visualisation Officer | Prof Ed Rybicki – Senior Research Consultant | Ashley Rustin – Senior eResearch Technical Specialist
Front row (l-r): Jill Claassen – Manager, Access and Visibility Services | Lena Nyahodza – Scholarly Communications Librarian Amina Adam – Research Commons Librarian | Andiswa Mfengu – Librarian | Kayleigh Roos – Data Curation Officer
Absent: Jason van Rooyen – eResearch Analyst.

HPC 2015-2016 (June to June)

27 CITATIONS

1458 CORES

219 TB STORAGE

260,678 JOBS

270 USER UPTAKE

5.5 M CPU HOURS

HPC growth since 2012

Year	Citations	Cores	Storage	Researchers supported (HPC)	Jobs	CPU hours
2012	8	232	9TB	31	75 466	364 343
2013	13	292	56TB	71	111 081	1 496 211
2014	19	1 142	169TB	124	985 216	4 163 686
2015	26	1 458	219TB	209	152 690	5 997 849

eResearch activities

Clinical data management systems (CDMS)

Clinical trials 2015-16 at UCT

23 RedCap

11 OpenClinica

2 OpenMRS

Research data storage

780 Users access research data storage

TOP 5 Research groups using our Research data storage

1. **SA Tuberculosis Vaccine Initiative**
2. **Magnetic Resonance Imaging Unit**
3. **Department of Astronomy**
4. **Zamani Project in the Division of Geomatics, Department of Architecture, Planning and Geomatics**
5. **Department of Computer Science**

Editorial: Natalie Simon
Pg 18 & 19: Boosting research with high-performance collaboration by Kim Cloete
All other features by Natalie Simon
Design and layout by Rothko
With a special thanks to all researchers and staff members featured in this report.
Visit our website (www.ereasearch.uct.ac.za) to keep up to date with the latest UCT eResearch news and updates.