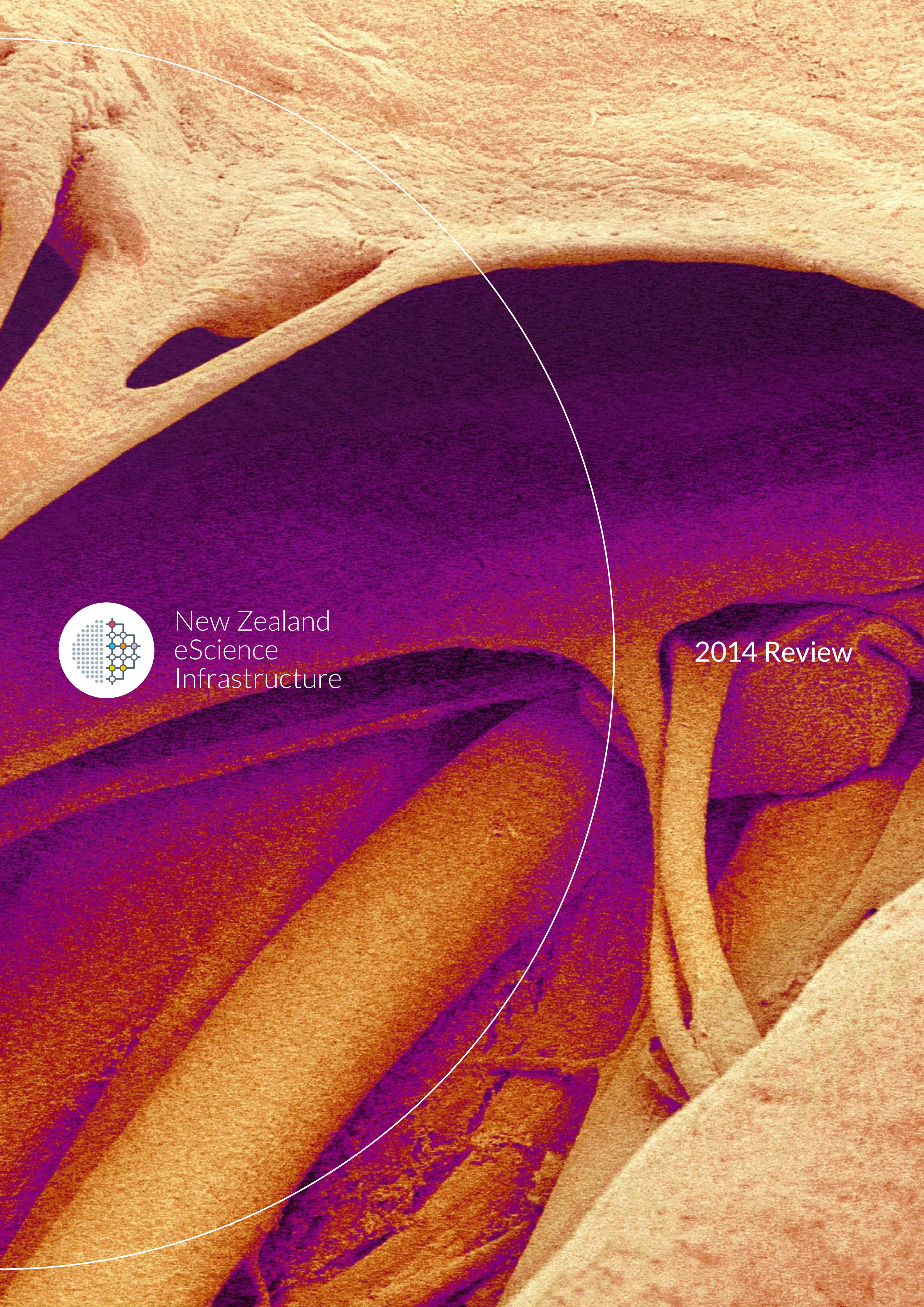


New Zealand eScience Infrastructure

2014 Review

New Zealand eScience Infrastructure

In 2014 NeSI Supercomputer power ran Computational Fluid Dynamic (CFD) modelling to quantify the haemodynamics for various stent designs. Improving stent design by measuring the impact on blood flow may significantly reduce stent failures and ultimately save lives.

See case study, page 16

2 Introduction

4 Chair's report

6 Director's report

8 – 11 Team

- Board of Directors
- Management team
- Delivering value through eScience services
- Computational Science team

12 – 15 Realising sector goals

- Advanced, scalable computing infrastructure
- Provide for the best possible uptake and return
- Encourage a high level of coordination and cooperation
- Contribute to high quality research outputs

16 – 27 Case studies

- Improving treatment of heart disease
- Securely transferring health study data across the country
- Protecting New Zealand's biodiversity with collaboration
- Genetics & metabolic syndrome
- The formation of surface archaeological deposits in arid Australia
- Modelling ash & gas dispersion for the Te Maari eruption

28 Strategic direction

VISION

New Zealand alongside the worlds' leading nations in accelerating research to advance prosperity.

PURPOSE

To grow the computing capability of researchers to ensure New Zealand's future prosperity.

OUR RESEARCH COMMUNITIES

NeSI supports research outcomes in key research areas for New Zealand across a broad range of disciplines and institutions, focusing on research underpinning New Zealand's economic, social and environmental wellbeing.

The power behind researchers

THE NATIONAL ORGANISATION

NeSI is a collaboration of five institutions coordinating investments in partnership with the Crown. This collaboration is constituted through a legal agreement between the University of Auckland, the University of Canterbury, the National Institute of Water & Atmospheric Research (NIWA), the University of Otago, and Landcare Research, with a single host institution, the University of Auckland, as the legal contracting entity with the Crown. A set of operating principles defined in the Collaboration and Services Agreements forms the basis for governance, management and operations.

NeSI is a peak eScience organisation in New Zealand, providing thought leadership on opportunities to innovate. We identify infrastructure & skill gaps that, if addressed, will significantly increase the sophistication & impact of e-infrastructure capabilities on research.

A year of intense activity and transformational change

Rick Christie
Chair

CHAIR'S REPORT

2014 was undoubtedly the busiest year ever in NeSI's four year history. Not only did usage grow significantly, but also, following the 2013 MBIE Review we have set about transforming our business model to a fully integrated, centralised and capability-focused shared services organisation.

From the outset I must thank Nick Jones and his Management team, the Collaborators and my Board for their massive effort to achieve this, rewarded by a renewal of NeSI's funding both from the Crown and Collaborators.

We now look to the future with renewed confidence, a strong strategic focus on delivering better research outcomes, and a capability set which is fit for purpose to tackle future science priorities, especially around the National Science Challenges and the next generation of CoREs.

The next few years will see increased emphasis on New Zealand's research infrastructure and eResearch in particular, and NeSI is a vital element in this mix, together with our partners REANNZ and NZGL. In last year's Annual Report I mentioned the formation of a Joint Working Group to facilitate a closer working relationship between the three organisations, and that work matured further during 2014 resulting in our joint commitment to the eResearch 2020 initiative, the eResearch NZ annual conference, and forthcoming challenges in managing and analysing both existing and future big data sets for the science sector. This work will accelerate during 2015 under the guidance of a Governance group comprising senior Directors from our respective organisations.

New Zealand's eResearch infrastructure has matured and grown significantly over the last ten years, both in terms of hardware assets and the technical support for scientists to use HPC to solve ever more complex research projects. Whilst NeSI's emphasis over the past year has been on the latter, 2015/16 will see the implementation of a National Platforms Road Map. This will comprise an HPC technology review, a needs analysis with the research community and collaborators, procurement process development and engagement with vendors, and a cloud services development plan. The implementation and execution of the overall plan and strategy will be approved and monitored by the NeSI Board, with the support of a strong and technically competent Management team.

As a result of changes in structure, 2015 has already seen an enlarging of the ranks of NeSI Collaborators, and this work will be ongoing for the balance of the year with changes to our consultation, access and Governance arrangements to suit. Overall we should see the research sector continue to thrive and grow with the changes in infrastructure undertaken during 2014.

Rick Christie

Rick Christie
Chair, Board of Directors

Growing research that computes

Nick Jones
Director

DIRECTOR'S REPORT

With the purpose of growing researcher capabilities, NeSI harnesses energy and intellect from across the sector to deliver specialised support for researchers.

Now in its fourth year, NeSI takes a refreshing approach to meeting sector needs, building on collaboration as a foundation and integrating existing capabilities at five research institutions.

NeSI is now an established high-tech infrastructure supporting high-performance computation and analysis for researchers. NeSI designs platforms tuned to the complexity of research that models cosmology, climate, natural hazards, biological systems, the earth and it's species, among other things.

Our people work with researchers to understand their research directions, and to acquire technologies to enhance their workflows, data analysis, computation and simulation. During the last year we have supported hundreds of researchers across almost all research institutions, and most disciplines. We're improving both the productivity by which research is delivered, and the possibilities of the questions being asked.

Our case studies from this most recent year highlight research powered by NeSI: applying computational fluid dynamic techniques to stent development; investigating the possibilities for new metabolic therapies for new antibacterial and antifungal agents; validating current methods for informing disaster response efforts during volcanic eruptions.

As we reviewed our past performance and future prospects over the last year, we identified areas where we felt we could improve. We've put in place strategic plans for the coming four years, articulated in our business case for the same period. The case outlines how we'll optimise our use of existing assets while adapting to the opportunities of cloud computing. We'll approach this as a nationally integrated team, and strengthen collaboration and partnerships across the sector.

NeSI's sector outreach to researchers consistently measures as high quality and meeting needs. Meanwhile, institutions are starting to work more strategically with their research leaders, putting in place joint planning with those responsible for their investments in enabling research, such as their libraries and IT services and research administration groups. Significant opportunities are emerging in applying infrastructure capabilities and services to enhance research effectiveness. We feel more can be done – so we will broaden our engagement beyond researchers, working to increase alignment with institutions sector wide.

We look forward to continued growth across the sector, and to supporting research that underpins New Zealand's economy, society and environment.

Nick Jones
Director

Board of Directors

The NeSI Board is responsible for strategy, policy, approving major initiatives and investments, as well as monitoring the NeSI risk register. Three of the collaborator institutions appoint a Director alongside an independent Chair and another independent member with expertise in the field representing the research sector at large. All Board members are focused on the interests of NeSI, being the effective delivery of national research infrastructure services.

Rick Christie
Chair, Independent Director, Professional Director
NeSI

Dr Murray Poulter
Chief Scientist, Atmosphere, Hazards & Energy
NIWA

Prof. Andrew Rohl
Independent Director & Professor of Computational Science
Curtin University

Prof. Steve Weaver
Deputy Vice-Chancellor (Research)
University of Canterbury

Stephen Whiteside
Director, Organisational Performance & CIO
University of Auckland

Crown Observers

As the Crown is a key partner supporting NeSI, the Crown appoints observers to Board discussions :

Anne Berryman
Observer Manager National Science Development Team
Ministry of Business, Innovation & Employment (MBIE)

Paul Bruere
Observer, Senior Investment Manager, Ministry of Business, Innovation & Employment (MBIE)

Management team

NeSI has a Management team, led by a Director, who are responsible for executive management, planning and overseeing day-to-day operations. Activities are coordinated through nationally distributed functional teams across each of the investing institutions. NeSI also maintains Advisory Committees and Panels, including an independent Access Policy Advisory Committee to advise the Director on the ongoing development and periodic review of access policy.

Kirsten Brown
Team Administrator & Executive Assistant

Malcolm Fraser
Transition Programme Manager

Marcus Gustafsson
Service Delivery Manager, NeSI Auckland

Nick Jones
Director

Sat Mandri
Tuakiri Service Manager

Kieron Mottley
Business Manager

Ella Rokotyan
Project Coordinator & Reporting Analyst

Dan Sun
Service Delivery Manager, NeSI Canterbury

Michael Uddstrom
Service Delivery Manager for HPCF, NIWA

Delivering value through eScience services

HIGH PERFORMANCE COMPUTING

NeSI provides a complete HPC service to the NZ Research and Education sector, Government initiatives/projects and Government Agencies, and industrial research. The key components of this service are both the platforms and people providing high performance computing services to the research community. Researchers received allocations on HPC platforms and were managed through allocation classes specific to various national communities, with allocation classes including Research, Proposal Development, Postgraduate, and Institutional Collaborator investors, as well as the development of Institutional Subscribers.

HPC CONSULTANCY & TRAINING

NeSI's HPC service also provided consultancy and training via its Computational Science team. Members of this specialist team provided expertise to support research projects with demanding HPC requirements as well as leading in the development and delivery of workshops and training sessions. A significant outcome of this service is to develop computational science capability to deliver on strategic collaborations with the research sector on projects of national significance.

OTHER ESCIENCE SERVICES

In 2014 NeSI delivered the following eScience services in collaboration with infrastructure providers and institutions:

- Data sharing allowed researchers to store, share and transfer data via the Data Fabric.
- High-throughput data transfers done in collaboration with REANNZ and Globus facilitated transfer of large datasets for users. This included improvements to transfer endpoints at NeSI sites and for endpoints at other institutions.
- Tuakiri, New Zealand's Access Federation, provided a common shared ICT service providing a secure authentication framework (web-based SAML Single Sign-On) that supports collaboration across the sector.

All of these services made use of NeSI's infrastructure hosting.

Computational Science team

The Computational Science team provides consultancy support to approved Projects. User training ranges from use of the HPC and data platforms to best practice software development methodologies.

Dan Sun

Computational Science team Lead: Dan has a Masters of Science in Computer Science and has background in software development, professional service & service management. He currently manages the operations of NeSI's Computational Science team and NeSI's service desk.

Sung Bae

Sung is NeSI's education and training coordinator and a certified Software Carpentry instructor. He has a PhD in algorithm and data structure optimisation and is interested in making scientific programming and HPC more accessible to researchers.

François Bissey

François gained a PhD in Particle Physics (University of Adelaide) during which time he was involved in lattice QCD(1), an early application of HPC. His expertise is in physics and quantum mechanics in particular.

Jordi Blasco

Jordi has a Licentiate in Physics with specialisation in Computational Physics (University of Barcelona). He has many years' experience in HPC in industry and academic environments, as a Solutions Architect, and in leadership & coordination of cutting edge HPC projects.

Peter Maxwell

Peter is the only member of the Computational Science team in Dunedin, and the only biologist. He has an MSc from the University of Auckland and spent a decade in Australia working in the bioinformatics field as a programmer, a consultant, and occasionally a tutor.

Ben Roberts

Ben completed his B.Sc. (Hons) in Chemistry (University of Canterbury) and obtained his PhD in Medicinal Chemistry (Monash University) before joining NeSI in 2012. Ben supports research teams with a focus on computational chemistry and materials science and is technical lead for the NeSI service desk.

John Rugis

John holds a PhD in Computer Science and a BSc in Mathematics (Auckland). He has extensive teaching and research experience in electronic engineering, physics and earth science. He collaborates with research groups focusing on scientific data visualisation and HPC implementation.

Gene Soudlenkov

Gene is trained as an applied mathematician and systems programmer with a development background in scientific programming, network security, and system level development. Gene prefers algorithmically intensive challenges.

12 – 15

NeSI has four objectives that define the shared intent of the Crown and its investors. These objectives underpin the rationale for NeSI's existence.

OBJECTIVES

OBJECTIVE 1
Create an advanced, scalable computing infrastructure to support New Zealand's research communities.

OBJECTIVE 2
Provide consistent access across HPC platforms, and research tools and applications, supported by basic data management tools, user-support, and sector engagement.

OBJECTIVE 3
Encourage a high level of coordination and cooperation across the research sector.

OBJECTIVE 4
Contribute to high quality research outputs from the application of advanced computing and data management techniques and associated services, which support the Government's published priorities for science.

Advanced, scalable computing infrastructure

OBJECTIVE 1

Create an advanced, scalable computing infrastructure to support New Zealand's research communities.

PROJECTS

Over 300 projects were run on NeSI HPC platforms in 2014 from approximately 50 science disciplines including including mechanical engineering, oceanography geology, human biology, statistics, computer science, and chemistry.

UTILISATION

More than 400 researchers accessed NeSI HPC platforms in 2014 using a total of 83,747,436 CPU core hours.

AVAILABILITY

NeSI's HPC platforms achieved over 98% availability in 2014.

RENEWED INVESTMENT

NeSI successfully completed its renewal with its collaborators and the Crown. An agreed business case and funding contract represents investment that will enable NeSI to continue to provide advanced computing services to enable researchers to enhance our national prosperity.

TRANSITION TO A NEW MODE OF DELIVERY

The Management team and stake holders played a key role in designing and implementing key changes to the way NeSI will deliver services over the coming four years. This proactive, team-led approach is a good indicator that NeSI is evolving to meet the goals outlined in its business case with the Crown.

HPC PLATFORMS

NeSI increased its compute capacity by 20% (or 1000 more cores, bringing the total to over 6000) on its x86 platform to address a shortage of Intel x86 based core hours available to researchers. The addition now makes a total of 53 million core hours per annum available for researchers on this platform, potentially catering for another 100 researchers.

SERVICE DELIVERY MATURITY

In 2014 NeSI developed processes for establishing a coherent service portfolio and roadmaps for each of its services. The wider programme aims to improve project management, organisational development, and change management. A service delivery design initiative defined improvements to core aspects of NeSI's services delivery. NeSI now has a capability development framework underpinning a Services innovation programme which sits across all of NeSI's services.

HPC CONSULTANCY

The Computational Science team contributed to NeSI's training and outreach programme throughout the year, as well as collaborated with researchers to lift the HPC capabilities of the research sector.

Provide for the best possible uptake and return

OBJECTIVE 2

Provide consistent access across HPC platforms, and research tools and applications, supported by basic data management tools, user-support, and sector engagement.

INSTITUTIONAL SUBSCRIPTION

NeSI is reducing barriers to entry to HPC by developing a new service offering for institutions to centrally fund access for their researchers to NeSI services through an institutional subscription. This approach will be piloted with Massey University, focusing on the HPC service, and Plant and Food Research, where the focus is on NeSI's data services.

RESEARCHER CAPABILITY DEVELOPMENT

NeSI ran training workshops throughout the year routinely or in response to users' requests. Workshops were conducted at Massey University, University of Otago, University of Auckland and University of Canterbury. A gap analysis was also completed this year to identify the research sector's need for HPC training. The outcome of this analysis informs 2015's annual planning process to ensure NeSI addresses these sector requirements.

TUAKIRI FEDERATED ACCESS

In 2014 Tuakiri facilitated 101,000 authentication sessions for users. The 2014 income forecast was met and membership maintained at all eight NZ universities, seven Crown research institutes, and one ITP institute. There were thirteen new Service Providers registered, a mix of national and international service providers and one new Identity Provider registration – AgResearch. Tuakiri maintained a highly available platform with 99% availability for 2014.

HIGH-THROUGHPUT TRANSFER CAPABILITY

NeSI worked with the Globus team to implement a pilot version of their high performance data transfer service. This service was then Tuakiri-enabled to provide single sign-on access using identities from NZ research institutions that are part of the Tuakiri Federation.

DATA SERVICES

A project was completed at the University of Otago to provide simple access to high performance data transfers between NeSI environments in Auckland and Otago, and to enable researchers to store, manage and share data with colleagues located at other institutions.

Encourage a high level of coordination and cooperation

OBJECTIVE 3

Encourage a high level of coordination and cooperation across the research sector.

PROJECT MANAGEMENT APPROACH

During 2014 NeSI formalised its project management approach and started the implementation of the approach to become a project-led organisation. A knowledge base has been developed that aligns with NeSI's capability framework, providing a clear roadmap for the team for the coming years. A training programme was initiated with the implementation of the project management approach and training programme to be continued in 2015 with the aim of increasing overall maturity.

NATIONAL ENGAGEMENT

- A NeSI/NZGL pilot project began exploring partnership in services delivery. A successful full day kick-off workshop was held on November 5th in Dunedin with plans for a joint training workshop early in 2015.

- The eResearch 2020 programme completed a draft report back document informed by the discussions at an eResearch 2020 hosted workshop, attended by leaders of the National Science Challenges and Centres of Research Excellence – this was notably the first workshop where the emerging leadership group were all together. This community will become a major stakeholder group for NeSI as their relevant infrastructure needs are identified.

- With host University of Waikato, NeSI organised the annual eResearch NZ conference held in early July 2014. The NeSI team were significant contributors, hosting the full day NZ HPC Applications workshop and contributing to many and varied sessions throughout the event.

INTERNATIONAL ENGAGEMENT

- NeSI presented at GlobusWorld 2014 on its work building the structure for providing high performance data transfer services to NZ researchers.

- In 2014 NeSI joined the Research Data Alliance as an Organisational Member, and NeSI's Director joined the Organisational Advisory Board. This allows NeSI to participate in a global community driving convergence of research data standards, policies and technologies.

- NeSI's Director gave an invited panel presentation at the International Conference on Research Infrastructures, a biannual event that brings together global leaders in research infrastructures, to share knowledge and identify opportunities for closer collaboration.

- NeSI presented at the eResearch Australasia conference in late October, and renewed many connections across the Australian research system. Groundwork was laid for 2015 collaborations.

Contribute to high quality research outputs

OBJECTIVE 4

Contribute to high quality research outputs from the application of advanced computing and data management techniques and associated services, which support the Government's published priorities for science.

NATIONAL RESEARCH COLLABORATIONS

NeSI's Computational Science team participated in several research projects during 2014, including:

- Parallelising existing workflow to increase throughput – the project aim is to parallelise CellML compilation and computational runs utilising as many cores as possible to minimise the required computational time by the factor of available CPU resources.

- Numeric modeling – collaborating with Professor James Sneyd from the University of Auckland's Department of Mathematics on his project "Multi-Scale Modeling of Saliva Secretion". The goal of the current project is simulation of salivary cell function using physically realistic cell models.

RESEARCH OUTPUTS

Over the course of the year more than 50 papers were identified that acknowledged NeSI. Other outputs were also recorded including seminars, conference proceedings and talks.

Susann Beier, Faculty of Medical & Health Sciences University of Auckland

16 – 27

NeSI contributed directly to New Zealand's innovation system in 2014. On the following pages is a selection of projects that have produced significant research outcomes during the year. These represent some key research done using NeSI's platforms and services across several science domains.

CASE STUDIES

CASE STUDY 1
Improving treatment of heart disease

CASE STUDY 2
Securely transferring health study data across the country

CASE STUDY 3
Protecting New Zealand's biodiversity with collaboration

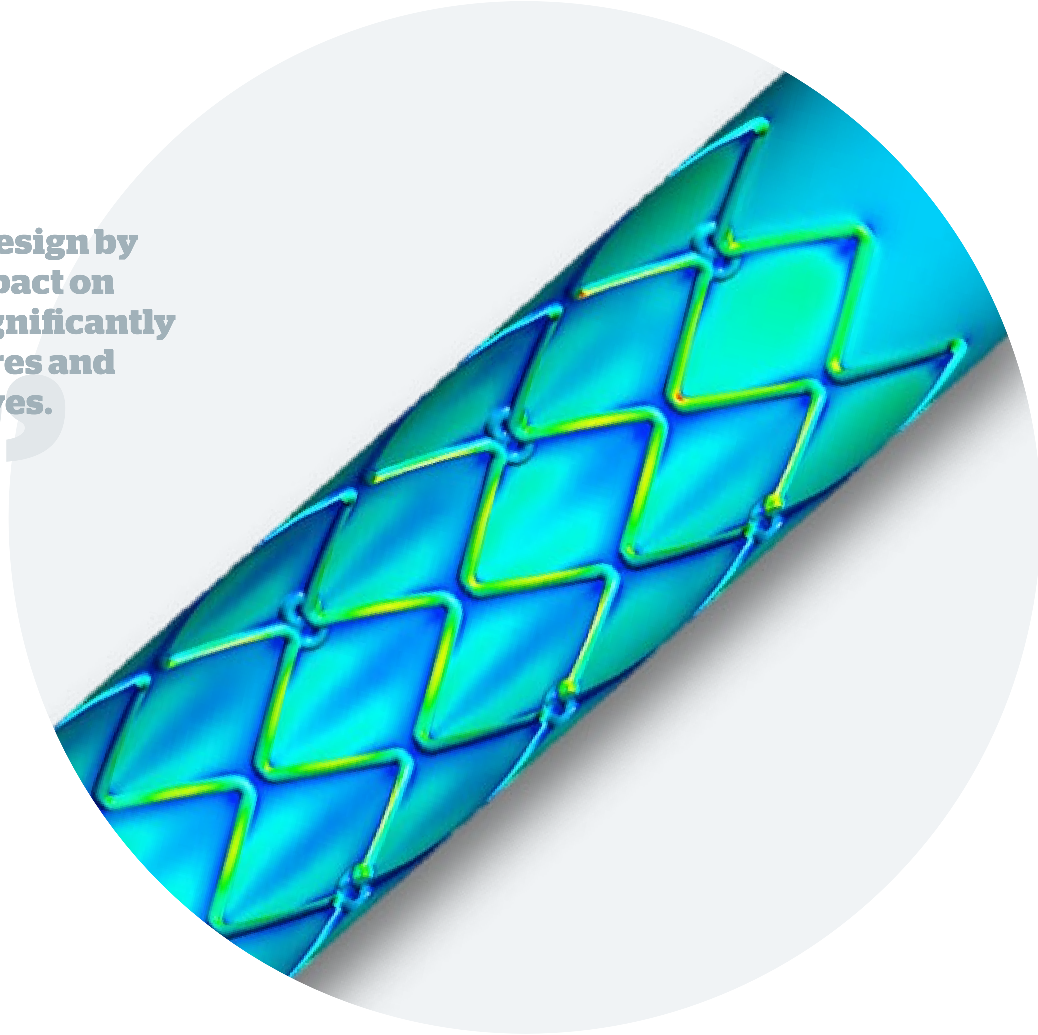
CASE STUDY 4
Genetics & metabolic syndrome

CASE STUDY 5
The formation of surface archaeological deposits in arid Australia

CASE STUDY 6
Modelling ash & gas dispersion for the Te Maari eruption

Improving treatment of heart disease

Improving stent design by measuring the impact on blood flow may significantly reduce stent failures and ultimately save lives.



Coronary artery disease is the most common killer in the Western world. One in four people will die from a blockage in one of the coronary arteries, preventing supply of blood and oxygen to the heart. A common treatment for narrowed arteries is 'Percutaneous Coronary Intervention' or PCI, where a wire mesh tube, or 'stent', is inserted into the narrowed blood vessel and expanded to hold it open.

FOR FURTHER READING ON THIS CASE STUDY, PLEASE REFER TO **nesi.org.nz/case-studies**

The expansion of the stent compresses the abnormal build-up of atheroma narrowing the artery. This is a highly effective and relatively low risk treatment, it avoids surgery, and enables rapid improvement in the patient's condition. Unfortunately, stents fail in 25% of patients, when the previous narrowing reoccurs within the stent.

New stents which release medication (known as 'drug-eluting' stents) promised a solution to this problem, with the slow local release of drugs supressing the regrowth of the tissue responsible for re-narrowing. A short-term improvement is achieved, but late failure rates still carry the risk of causing sudden death. It is known that the presence of the stent changes blood flow, which in turn alters the growth of tissue, and thereby affects treatment outcome. Currently, there are more than 250 different stent designs available, each of which has a different effect on the 3D pattern of blood flow. The stent design is therefore critically important in determining treatment success or failure. Clinical trials with groups of patients have been unable to define the stent design features which determine success or failure due to the complexity of the problem. More detailed quantitative analysis is needed to answer the question of how different stent designs affect the complex blood flow through the coronary arteries.

This research was conducted by Susann Beier, a PhD candidate in the Faculty of Medical and Health Sciences at the University of Auckland. The study aims to investigate the link between stent design features, the changes they cause in blood flow (haemodynamics), and the impact on patient outcome. Computational Fluid Dynamic (CFD) modelling was used to quantify the haemodynamics for various stent designs. This allowed the calculation of blood velocity, the stress the blood flow induces on the vessel wall, and to detect regions of rapid changes by simulating the vessel shape, circulation and blood rheology. Two commonly used stent designs were modelled to determine their respective strengths and weaknesses.

In addition to the computational simulations, a number of blood vessels were scaled up, manufactured in plastic using 3D printing and connected to a pump to simulate the heart. They were then placed in an MRI scanner to measure the blood flow with PC-MR (Phase-Contrast Magnetic Resonance). A comparison was then performed between the computational CFD and experimental MRI data with good agreement found.

Many simulation time steps were required to accurately represent the rapidly changing flow conditions in the coronaries with more than 2,500 time steps computed for a single heartbeat. To ensure full development of the flow, four heartbeats were modelled, and only the last was analysed. Each simulation contained approximately 20 million elements, and used multi-threading over six cores with 40CPU with 20GB memory each, and took around 48 hours to solve. This task would have not been possible without the use of NeSI's Pan cluster, as the memory requirements would have exceeded local desktop capabilities well before a possibly month long solution was obtained. We were able to run many simulations on the cluster, allowing us to systematically investigate more than 20 different stent design aspects, and their impact on blood flow in the coronary arteries.

Due to having access to the Pan platform we are now extending our research to more complex vessel geometries such as vessel branching. Having gained a greater understanding of the features that are critical to idealised stent design, the next milestone is to model a stented coronary artery from an actual patient. This is a highly complex undertaking but the knowledge that can be gained will have an important impact in the treatment of patients with heart disease.

Russell Butson,
University of Otago

Securely transferring health study data across the country

The Kahungunu Infant Safe Sleep (KISS) Study team used the NeSI Data Fabric to securely transfer data from the Eastern Institute of Technology in Napier to the University of Otago's Dunedin School of Medicine.

FOR FURTHER READING ON THIS CASE STUDY, PLEASE REFER TO
nesi.org.nz/case-studies

Previously, hard disk drives were physically sent around the country. Ngati Kahungunu is the nation's largest iwi by area. It spans large parts of the North Island's east coast and is primarily rural.

KISS is a randomised controlled trial of the use of the traditional Māori wahakura, or flax woven baskets, versus portable bassinet for the sleep of infants at risk of Sudden Unexpected Death in Infancy (SUDI). The researchers will look at baby sleeping patterns, mother-baby contact time, breastfeeding at night, mother-bonding, blankets covering face during sleep and the warmth of the sleeping space.

"The project team were originally going to courier hard drives across the country", explains Russell Butson from the University of Otago, "but there are security implications for that. Also, if one of the drives became damaged or lost, then we are likely to lose all of the video from that session." The nature of the data is that the interviews can never be recovered if lost.

The NeSI Data Fabric is a flexible data storage service. It allows researchers to upload files and then grant others with access to that data. Files are able to be downloaded via a web browser, or other interfaces such as WebDAV.

Files are encrypted and stored at multiple locations. Data is peered between nodes utilising the REANNZ network. This technology enables highly secure, robust method for transferring large files at very rapid speeds across the data.

The study generated roughly 5TB of data over 18 months. This included ~250 transfers of video files, each of which were roughly 20GB.

The project's own needs have been met and the experience has led to several improvements, within NeSI, which benefit all researchers.

> **"The KISS project has been a rewarding process for us in general. We've been able to improve both our user support processes and end-to-end service delivery with other infrastructure providers in New Zealand."**

KISS uncovered a number of weaknesses with firewall configurations and shed light on various networking bottlenecks. To identify and fix these bottlenecks involved cooperation between NeSI collaborating institutions, REANNZ and the KISS researchers themselves. These bottlenecks have subsequently been removed, meaning faster speeds for all researchers.

Working in a collaborative manner to fix problems for researchers has also led to improvements in other, non-technical aspects of NeSI. "NeSI is now able to be more responsive to support queries" says Kieron Mottley from NeSI. "The KISS project has been a rewarding process for us in general. We've been able to improve both our user support processes and end-to-end service delivery with other infrastructure providers in New Zealand."

Mr Butson, who is based at the university's Higher Education Development Centre, emphasises the non-technical elements when it comes to adoption: "Researchers will not change practice just because it is more efficient and effective. Training and professional development shouldn't be underestimated. With that in mind, he believes that the NeSI Data Fabric is "the best way to transfer data".

Mr Butson continues, "The project team don't care about the technical details. Staff just want to upload their data to the cloud and have it available for everyone else. We happen to be shifting terabytes of high definition video footage across the country."

Richard Hall, Matthew Peacey, Nicole Moore & Jing Wang from ESR, Dan Tompkins & Dan White from Landcare Research & Kate McInnes from DOC

Protecting New Zealand's biodiversity with collaboration

Collaborators at ESR, Landcare Research and DOC have been working together to understand the health of New Zealand's only terrestrial mammal, Mystacina tuberculata bats.

FOR FURTHER READING ON THIS CASE STUDY, PLEASE REFER TO
nesi.org.nz/case-studies

This work could also help to address the longstanding question of whether coronaviruses have evolved recently or are much more ancient.

Richard Hall and Jing Wang made use of NeSI's HPC platforms and the expertise of Sung Bae, from the Computational Science Team, to speed up their part of the project. They also worked with genetic sequencing data, attempting to identify any novel viruses in samples collected by the others in the team. They used NeSI's HPC platforms to examine the results of genetic sequencing to and see whether there was any unique genetic material.

Richard explains why a powerful computing system is necessary for this kind of work, "Finding viruses is never easy. Unlike everything else, like humans, mammals, vertebrates or even bacteria, there is no common shared genes between two viruses. That means, in order to identify a virus, we need to compare our sequence against every other known sequence. That is a huge task."

The research team uncovered a new Alphacoronavirus. This may prove to be an important result for the health of this special member of New Zealand's ecosystem. As the bats are New Zealand's only terrestrial mammal and are also quite geographically isolated, this finding may indicate that coronaviruses are indeed ancient, perhaps millions of years old - although much more research is needed to confirm this.

The study involved collecting bat guano from four sites on Whenua Hou (Codfish Island), off the west coast of Stewart Island. This material was then sequenced by NZ Genomics, Ltd. The results of that sequencing were then processed by ESR staff.

Next generation sequencing produces vast quantities of data. ESR compared the bat dataset to other known all other nucleotide sequences in Genbank. Genbank is a worldwide database of genomic information. The tool used to carry this out is called the Basic Local Alignment Search Tool (BLAST) and is available at NeSI.

SCALING BLAST

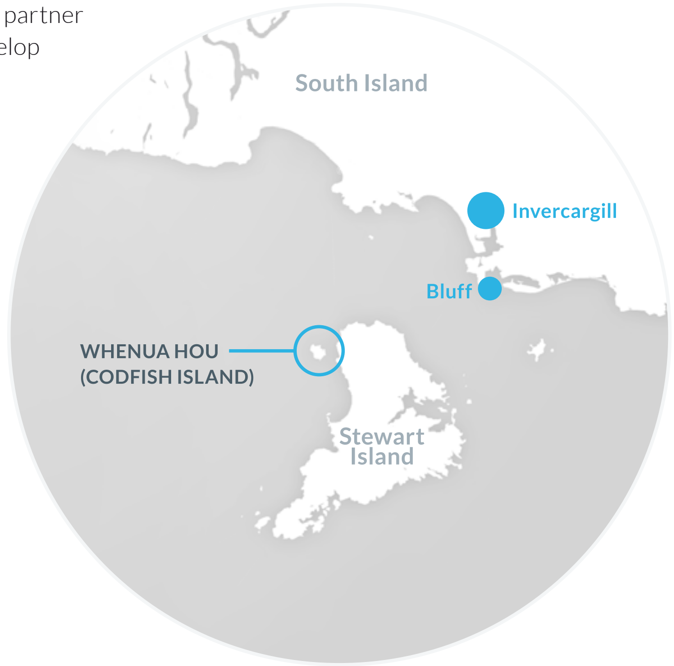
BLAST is a common tool used within molecular biology. It enables researchers to compare nucleotide or protein sequences from multiple sources. Similarities between sequences can be used to shed light on gene function, possible new gene families and evolutionary history. BLAST+ is a new version that provides significant speed improvements.

To get the most out of BLAST+, it is important to split large queries into batches. This splitting can present challenges when transitioning to an HPC system, due to the complexities of job scheduling. NeSI staff member Sung Bae worked with Richard and Jing to implement a job submission workflow that enabled linear speedup up to several thousand batches at a time.

Concurrently, the team worked to understand whether compiler options could enable the software to run faster. By profiling, benchmarking and experimenting, Sung was able to provide a 2.7x speedup without touching a line of source code.

COLLABORATIVE PROJECTS

NeSI's Computational Science Team spends time working with research groups to understand their needs and then works intensively as a partner to remove roadblocks, develop algorithms or do whatever needs doing to speed research along. The team is available nationally to every research group that has a project with NeSI under the Research Allocation Class. In one case, sitting down with the team ended up in a 1,300 - fold increase in speed per run.



Genetics & metabolic syndrome

"My research has involved investigating signatures of selection and their association with metabolic disease," says University of Otago PhD student Murray Cadzow. "By using NeSI, the length of time to complete this analysis was reduced by weeks. Initially it seemed daunting thinking about using a national compute resource with a limited technical knowledge but there was plenty of training and resources available to guide me through."

FOR FURTHER READING ON THIS CASE STUDY, PLEASE REFER TO
nesi.org.nz/case-studies

Metabolic disease, also known as metabolic syndrome, is a significant feature of many chronic diseases. It is a dysfunction of the body's ability to process and store energy, characterised by obesity, high blood pressure, diabetes and other conditions.

"The impact of this work on metabolic syndrome would give more information as to what variants in genes may have been beneficial or detrimental to a population along its history. It should give us a hint about which variants may have been selected for or against. These days, with the change in environment, these (previously beneficial) variants may now contribute to disease."

Murray is based in Dunedin, but has been able to access high performance computing platforms in Auckland via NeSI and the REANNZ Network.

The research programme seeks to provide greater insight into the genetic disposition of disease. He and his colleagues have been hunting for "signatures of selection". That is, areas of the genome that exhibit features of having been under selective pressure. These signatures are associated with phenotypes, which can be population specific. Phenotypes are physical characteristics of an organism, or could be thought of as the end result of genetics plus environment.

"As part of this, I was using publicly available data from the population data from the 1000 Genomes Project in conjunction with data from Pacific populations. For this project we created a bioinformatics workflow to output different statistics of selection."

The availability of open science data, compute resources of national scale and excellent technical support were essential to supplement the domain knowledge that Murray and his collaborators had developed through their studies. Without this synergy, the project would have progressed far more slowly, if it could have been progressed at all.

New Zealand possesses a high level of bioinformatics expertise that is supported by the collaborative and complementary efforts of its research institutions, the Virtual Institute of Statistical Genetics, NeSI, NZGL and REANNZ. Bioinformatics will be key to New Zealand's continued progress in primary industries. As Murray and his collaborators also demonstrate, it may also be a strong factor in enabling our society to combat chronic disease.

The focus on increasing the efficiency and research productivity is demonstrated by the first peer-reviewed paper that has emerged from the work. The workflow the team has developed has been made available to others via an article in Frontiers of Genomics (Cadzow, et al 2014) and as a software repository.

Murray and his team mates have invested a lot of time to make the workflow that they've developed available to others as an easy-to-use toolkit. Previous work from Pybus et al (2014) enabled the community to select a small segment of the human genome. "For researchers wishing to investigate selection in other human cohorts or populations – or other organisms – a non-trivial amount of data manipulation and subsequent computation is required in order to extract this type of information from the available data."

The toolkit integrates many tools, each written in their own programming language and utilising their own file formats, into an easy-to-use pipeline. Different tools have also made their own trade-off between accuracy and speed. The pipeline greatly simplifies the ability for researchers to detect signatures of selection across a whole genome. This computationally intensive work is able to be distributed across many cores.

Part of the research involves calculating the iHS (integrated haplotype score) values for all areas containing Single Nucleotide Polymorphisms (SNPs) within a genome. From Wikipedia: "A Single Nucleotide Polymorphism (SNP, pronounced snip; plural snips) is a DNA sequence variation occurring commonly within a population (e.g. 1%) in which a single nucleotide — A, T, C or G — in the genome (or other shared sequence) differs between members of a biological species or paired chromosomes. For example, two sequenced DNA fragments from different individuals, AAGCCTA to AAGCTTA, contain a difference in a single nucleotide."

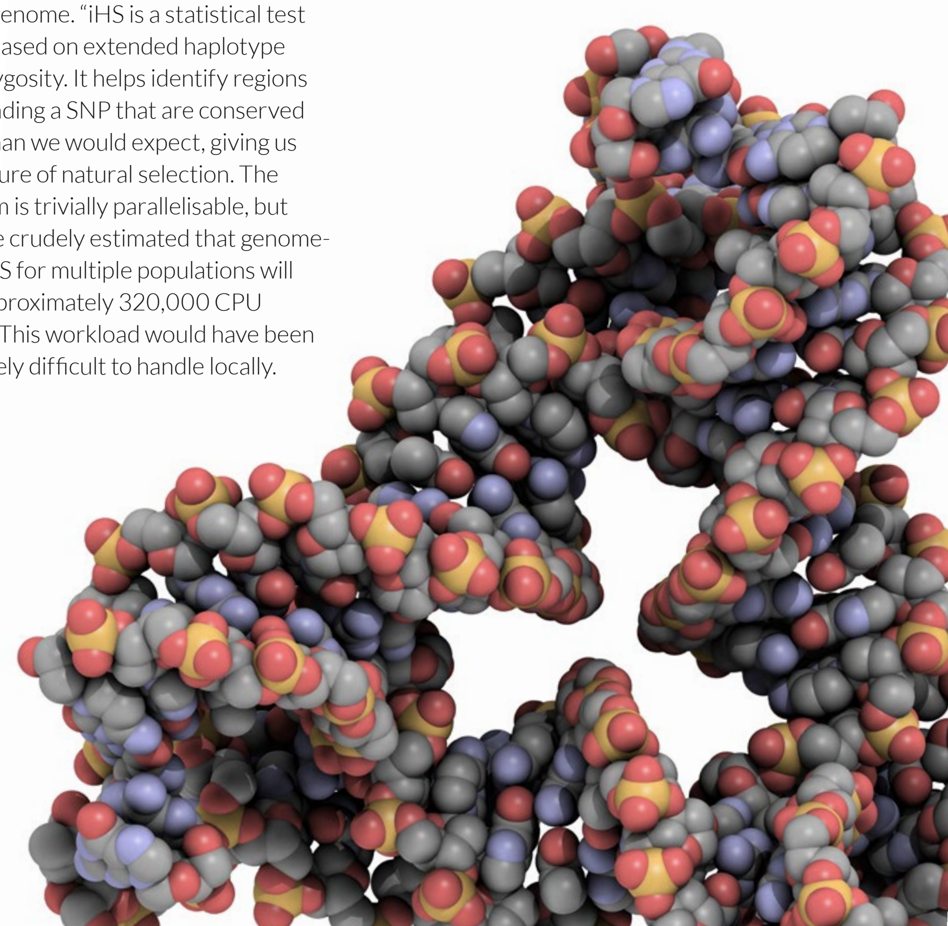
Murray explains the significance of the iHS test and the technical challenges of running the analysis across the whole genome. "iHS is a statistical test that is based on extended haplotype homozygosity. It helps identify regions surrounding a SNP that are conserved more than we would expect, giving us a signature of natural selection. The program is trivially parallelisable, but we have crudely estimated that genome-wide iHS for multiple populations will take approximately 320,000 CPU hours." This workload would have been extremely difficult to handle locally.

"At the start of the project, we also expected to develop and test improvements using MPI and C, and possibly GPU to perform the calculations." The NeSI HPC platforms provide the flexibility for researchers to experiment with the algorithm and hardware that will run best.

Access to NeSI also enabled Murray to iterate quickly on new approaches described in the literature. In his case, he was interested in exploring the feasibility of incorporating a recently-released package into his analysis. "CMS is an extended derivative of iHS combining other tests into a single test statistic. The software has only recently been released and may require some work to get running, but we again estimate roughly that it will require a further 320,000 CPU hours." NeSI was able to accommodate this request.

As Murray's PhD continues, we are sure that we will continue to see the results of this hard work. NeSI's high performance computing platforms, data services and computational support team will continue to be at his disposal as he progresses.

"By using NeSI, the length of time to complete this analysis was reduced by weeks."



Ben Davies, School of Social Sciences
University of Auckland

The formation of surface archaeological deposits in arid Australia

> **NeSI staff developed customised shell scripts that allowed us to generate thousands of simulated populations in rapid succession on the NeSI cluster.**

The deposits that archaeologists study are the product of behavioural and geophysical processes, both of which operate at varying tempos over time.

FOR FURTHER READING ON THIS CASE STUDY, PLEASE REFER TO
nesi.org.nz/case-studies

However, archaeological studies often emphasise the role of human behaviour, looking for signals of human activity within arrangements of artefacts and features. Geophysical and post-depositional processes are typically thought of as disturbances to signals of human behaviour, generating an incomplete record that archaeologists are tasked with reconstructing. Different deposits, depending on their contexts, are considered to be more or less informative based on impressions of degrees of preservation.

Ben Davies, at the University of Auckland, conducted a study to understand the formation dynamics of surface archaeological deposits in order to identify what kinds of patterns might be associated with different formational processes, both behavioural and geophysical.

Australia's arid zone is dominated by surface scatters of stone artefacts and cooking hearths. These are notoriously difficult for archaeologists to interpret due to stratigraphic mixing caused by long term erosion, and are often incorporated into settlement models based on an exclusively behavioural model. But what these deposits lack in terms of vertical integrity is made up for by the high degree of visibility over large spatial extents. This kind of record can offer archaeologists an opportunity to understand prehistoric use of space at the landscape scale, something typically hampered by the expense and logistics of excavation. Previous research at Rutherford's Creek in western New South Wales has produced a record of radiocarbon dates which have been obtained from ancient cooking hearths. When taken together, these records show a general trend of greater numbers of dates in the more recent past, as well as noticeable gaps in the occupation history for the area. The overall increasing trend may be consistent with an increasing population, while the gaps may be explainable as human absence or the effects of dispersal and congregation. Both of these patterns, however, may also be consistent with differential preservation rather than any particular activity set. Furthermore, there is some evidence to suggest that people in the past were regularly carrying away stone artefacts in fairly large numbers. This is seen in the low amount of cortex (the outer weathered surface of stone nodules) found in archaeological assemblages, indicating that the mobility of the people who manufactured the artefacts was high as cortical pieces are taken away. The spatial areas covered by the archaeological assessments, however, are small in comparison with the probable space being utilised in the past, and it is not known how different land use schemes might produce different patterning at the local scale.

To assess these issues, we constructed a set of agent-based computer simulations to evaluate the distribution of archaeological remains in space, and how general patterns of erosion and deposition through time might affect the character of surface deposits.

Spatial simulations were developed using the NetLogo software package. The first set simulates the changing visibility of surface deposits by modelling the erosion and deposition of surface sediments alongside the construction of cooking hearths. Hearths are constructed by programmed agents at a given rate, and these become hidden, visible, or destroyed by sequences of geomorphic events. At the end of the simulation, the resultant distribution of surface hearths are sampled and compared with field data. The second set of simulations resamples a database of stone artefacts made in a laboratory, and simulates their dispersal using different mechanisms (random walks, correlated walks, etc.). A GIS extension allows the simulations to be run in a spatial environment analogous to Rutherford's Creek, using different landscape features (lakes, creeks, etc.) as attractors for human occupation.

The data and code were then uploaded to the NeSI cluster, where batches of controlled experiments were automated using a customised shell script developed with the help of NeSI staff. The simulations disperse millions of artefacts in less than an hour, and generate thousands of simulated populations of radiocarbon dates in rapid succession according to different geomorphic models. Once these reach densities comparable with those recorded in the field, the results are returned from the cluster, sampled appropriately, and compared to those obtained from Rutherford's Creek. Without the speed of the NeSI cluster and the experiment automation devised by the NeSI staff, this would be a slow and gruelling procedure.

Preliminary results from this study suggest that the patterns in stone artefact assemblages and hearths are more consistent with explanations that include high levels of mobility among human groups. Changes in the levels of cortex among recorded stone tool assemblages are parsimonious with repeated, short-term occupation with little redundancy in place use. Additionally, modelling the formation of the radiocarbon record indicates that both the increasing trend in radiocarbon dates and gaps seen in the record may be consistent with large scale erosion and deposition events. These simulations can be used to develop field tests which may help to differentiate between erosion-driven gaps and those associated with human absence. This kind of insight is important for the development of informed heritage management policy as well as for developing a better understanding of the human past.

Richard Turner, Stuart Moore, Michael Uddstrom from NIWA; Natalia Pardo, Gabor Kereszturi, Shane Cronin from Massey University; Tony Hurst from GNS.

Modelling ash & gas dispersion for the Te Maari eruption

Scientists from NIWA, Massey University and GNS Science work together to test ash and gas dispersion models.

FOR FURTHER READING ON THIS CASE STUDY, PLEASE REFER TO
nesi.org.nz/case-studies

On 6 August 2012, Te Maari erupted. Te Maari is part of a geologically active area that includes Mount Tongariro and Mount Ngauruhoe on New Zealand's North Island.

The eruption provided an excellent opportunity to test and validate methods for modelling the dispersion of ash and gas. This validation strengthens New Zealand's ability to respond to future volcanic hazards. The study also revealed further improvements that could be included in the model.

The eruption produced a relatively small ash-dominated plume that covered 1,600km2 in about an hour. The plume's volume reached ~230,000m3 with a height of 8-10km. The human impact was relatively minor, but the eruption was noticed over a widespread area with gas emissions recorded for several months afterwards.

To detect the upper level (troposphere) ash movement, satellite imagery provided by NOAA was processed using computer vision algorithms on NeSI infrastructure. Information about ash volume and density was supplemented with other sources to generate a comprehensive picture of how the ash travelled through the atmosphere. These sources included samples of ash at various locations and weather data produced by the NWP model NZLAM-12.

Observational and NWP data then enabled realistic inputs to be provided to the dispersion model, NAME. NAME is produced by the United Kingdom's Met Office. It is an air pollution model, originally developed in response to the 1986 Chernobyl accident. The modelling agreed quite well with observations. For example, a split of the plume at the 10-30 micron level occurred between 2-3am. In the satellite imagery, particles of 11 and 12 microns in diameter are split into a main cluster and a smaller fringe cluster by 3:45.

While upper level ash movement was very consistent between, there was more ash fall at large distances (100-150km) from the mountain than forecast by the model. This suggests that more features need to be incorporated into the model. Candidates for inclusion are particulate aggregation and ice nucleation effects.

Events such as the Te Maari eruption enable the science response to emergencies to be rigorously tested. In order to provide timely and accurate advice to civil defence emergency management officials, the science system needs to be able to have access to high quality tools, systems and expertise.

NeSI is part of the infrastructure behind the science response. The time taken to produce a sufficiently high resolution model can mean the difference between operational decisions—such as flight cancellations—being well informed or being made in the dark. The ability to quickly model ash fall enhances New Zealand's ability to prepare for and respond to its volcanic risk.

To view the images in this Case Study, please refer to our website www.nesi.org.nz/case-study/te-mari

Strategic direction

Researchers are now dependent on advanced computing capabilities to achieve their goals. Historically, these capabilities have been funded by research projects. Developing and sustaining these capabilities is now a strategic priority for research intensive institutions and nations.

New Zealand eScience Infrastructure (NeSI) recognises and supports this change, and is New Zealand's response to deliver on these strategies. NeSI adapts advanced computing technologies to meet the varied needs of New Zealand's research communities, providing essential research services and capabilities that enable and empower research. NeSI has served over 150 million hours of computing since its inception less than 4 years ago, working across hundreds of research projects each year to enhance research productivity.

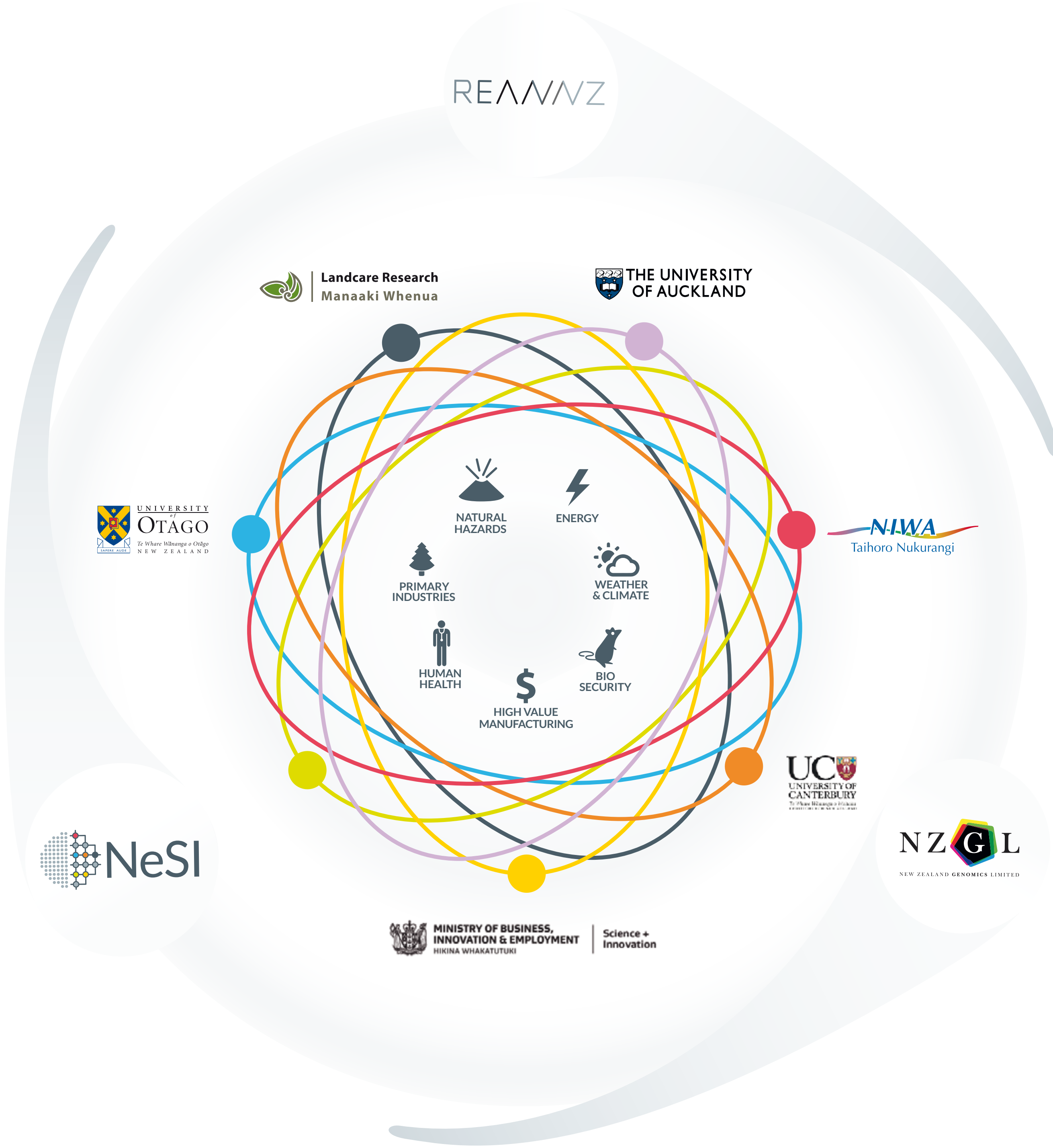
2015 provides a fresh start for NeSI; coming out of the 2013/2014 review of and reinvestment into NeSI by NeSI's collaborators and the Crown, the path ahead is now clear. NeSI's Board of Directors outlined a plan for the coming four years that aims to enhance our existing capabilities and further align NeSI with the new strategic directions being laid out for the research system as a whole.

NeSI's purpose lies in growing the computing capability of researchers to ensure New Zealand's future prosperity. It does this through delivering high performance computing and analytics services, and enhances research capabilities through consultancy and training. To achieve this requires NeSI to evolve in two key ways:

- Moving from local to national: NeSI's team and investments started from the individual capabilities of several institutions. The first three years have clarified how best these capabilities can be applied to support research. NeSI's team will come together as one national team aligned along lines of service and function. NeSI's investment framework prioritises joint planning and investment with research communities and institutions, aligning investments with national priorities first, thereby meeting the mix of institutional needs across the sector.
- Focus on growing capabilities for research: Building a national team that delivers at the leading edge of performance requires a culture of excellence and continuous improvement. NeSI applies this excellence to build research capabilities through training and consultancy on research projects, especially in software development that implements research methods.

Within this context NeSI has set the following goals for 2015:

1. Making it easier to start: we will remove barriers to empower researchers to make effective use of advanced computing capabilities;
2. Improving time to solution: we will enhance the capabilities available to researchers to speed up their research as they approach their most challenging questions;
3. Telling our stories: as we develop unique insights into the successes and enablers of success, we will share these insights and grow a greater community around their application.



During 2015 NeSI anticipates working across the sector to meet its goals, and looks forward to significantly progressing national infrastructure services and capabilities in partnership with the sector.

nesi.org.nz