THE UNIVERSITY OF MELBOURNE

Petascale
Campus Initiative

Shaping the future of data-intensive research

2019 – 2020

MELBOURNE DATA ANALYTICS PLATFORM

We acknowledge that we are on the unceded lands of the Wurundjeri, Boon Wurrung, Wathaurong, Taungurung, Yorta Yorta, and the Dja Dja Wurrung Peoples who have been custodians of the land on which we work, and pay our respects to their Elders past and present. We respect their heritage and ongoing relationship with land, sea, sky, and waterways.

Creative Producer: **Jo Condon**, PCI / MDAP
Designer: **Anthony Vanzella**, Strategy and Culture
Editors: **A/Prof Paul Gruba, Prof Andrew Turpin and Dr Andrew Siebel**, PCI / MDAP
Design consultants: **Kevin Calthorpe and Georgina Wilkinson**, Strategy and Culture
Contributors: **MDAP team and Collaborators**, acknowledged within
Approved by: **Prof Andrew Turpin,** MDAP Director

Published: July 2020

Contents

Foreword

The University of Melbourne is committed to ongoing investment in research infrastructure and services to enable endeavours across all disciplines. University platforms range from advanced microscopy to cultural informatics, metabolomics, health economics, and now, data analytics and stewardship in the Melbourne Data Analytics Platform (MDAP). The expertise of staff associated with these research infrastructure platforms, together with connecting precinct, national and global capabilities, critically enables our fundamental and translational research.

In response to the emerging importance of computing and large data sets across all research disciplines, we established MDAP in May 2019 as part of the Petascale Campus Initiative (PCI). Now a team of 20 staff, MDAP supports new research directions in all disciplines, and facilitates interdisciplinary research at The University. In addition to direct collaboration with researchers, MDAP works closely with the wider academy to direct researchers to other supporting services, platforms and infrastructure within The University.

For the long term sustainability of research at The University that depends on computing and data, it is essential that it can attract and retain excellent support staff. MDAP has established itself as an attractive career option for those who might otherwise command high salaries outside the academic research sector. Our MDAP staff have academic status and the associated career progression opportunities as Academic Specialists. This includes, but is not restricted to co-authorship on publications, supervision of students through the MDAP Intern program, and attainment of research funding.

MDAP is currently unique in Australia and the region because of the academic status of its staff, and also due to the mix of humanities, social science and computational experts who are passionate about driving research in new and exciting directions.

Prof Liz Sonenberg
Pro Vice-Chancellor
Research Systems

Prof Andrew Turpin
MDAP Director

Dr Andrew Siebel
MDAP Manager

“There is global recognition of the need for a new cadre of digitally skilled professionals in the research sector to advance research, that combine computational and data skills with research discipline expertise to advance research outcomes.

In Australian universities, MDAP is leading the way as a model for building career paths for such a workforce.”

Dr Michelle Barker
Chair, OECD Expert Group
on Digital Skills

About MDAP

We are a team of Research Data Specialists and Research Data Stewards working to enable data-intensive research across all disciplines. We collaborate with researchers throughout their research journey — from research design and methodology, to extensive data analytics and data preservation. We're passionate about enhancing research impact by applying and sharing data-intensive research methodologies.

RESEARCH CONSULTATION, DESIGN AND DATA MANAGEMENT

We work with researchers to plan and design data-intensive projects. Throughout the research life cycle, we guide projects and teams to work more effectively with data and new methods and techniques. Each collaboration places researchers at the forefront, and is dedicated to achieving productive outcomes, improving efficiencies and fostering digital capabilities.

DATA ANALYSIS AND COMPUTATION

Data analysis and computation, at times, demands highly specialised skills. Working within and across the team, we have a unique interdisciplinary capability to assist researchers to gain insights to their data, move it through sophisticated pipelines and create ways to combine digital and human expertise that are needed to work effectively with Big Data.

DATA PRESERVATION AND PUBLISHING

Funding agencies, governments and research professionals now expect research data to be professionally managed and disseminated. In collaboration with University research teams and other university research enablers and providers, we offer support throughout the final stages of the research life cycle to ensure compliance, safeguard assets, and meet standards. Our ultimate aim: to help usher in the new era of global research made possible through effective research data management and publication.

Our Team

Introducing our multidisciplinary team with diverse backgrounds led by Prof Andrew Turpin (MDAP Director) and Dr Andrew Siebel (MDAP Manager).

Jo Condon
Communications Specialist

Kim Doyle
Research Data Specialist

Dr Noel Faux
Senior Research Data Specialist

Dr Emily Fitzgerald
Research Data Specialist

Jonathan Garber
Research Data Specialist

Zaher Joukhadar
Research Data Specialist

Dr Aleks Michalewicz
Research Data Specialist

Dr Simon Mutch
Senior Research Data Specialist

Usha Nattala
Research Data Specialist

Priyanka Pillai
Research Data Specialist

Dr Mar Quiroga
Research Data Specialist

Dr Daniel Russo-Batterham
Research Data Specialist

Bobbie Shaban
Research Data Specialist

Ursula Soulsby
Business Analyst

Dr Kristal Spreadborough
Research Data Steward

Dr Edoardo Tescari
Senior Research Data Specialist

Karen Thompson
Senior Research Data Steward

Robert Turnbull
Research Data Specialist

Geordie Zhang
Research Data Specialist

Purpose and Principles

Our purpose is to increase The University's capability for innovative, impactful, data-intensive research through multidisciplinary collaboration. To achieve this:

WE ARE COLLABORATIVE

As Academic Specialists, we will collaborate with researchers across The University, and help connect them to other researchers, platforms and services. We value and respect the expertise of all collaborators, and base our work practices on partnerships.

WE ARE TEAM-ORIENTED

Our success and value as a team comes from our combined capabilities across a broad range of computational disciplines, experience with data, organisational skill sets and academic domains. We engage with researchers with a team-first approach, and support each other in solving challenging problems.

WE ARE MULTIDISCIPLINARY

We have a great breadth of discipline backgrounds in our team, and use our deep technical knowledge to stimulate research into new directions. Bringing together researchers from different disciplines along with our own experience and expertise results in impactful research outcomes from our collaborations.

Context

RESPONDING TO NATIONAL AND LOCAL NEEDS

Responding to, and indeed influencing, the demands of a data-rich world lies at the core of much of today's research and innovation. At MDAP, we strive to produce new knowledge across all University domains, from analysing speech data using natural language processing, to characterising parasite genomes using novel bioinformatics pipelines, and developing a data governance platform for the Indigenous Data Network. Our research collaborators know that they can rely on us to provide resources, tools and expertise to help create powerful results. The University's commitment to supporting data-intensive research is already yielding outcomes that will improve lives and the environment, now and for future generations.

RESPONDING TO GLOBAL CHALLENGES

More than ever, data lies at the heart of evidence-based decisions that affect lives across the world. The scale, speed and depth of such data is changing daily. From smartphones to traffic lights to artwork to climate change, large amounts of data now flows in and out of our work and home lives. Understanding data, and harnessing its insights, is the challenge of our time and the light for our collective future.

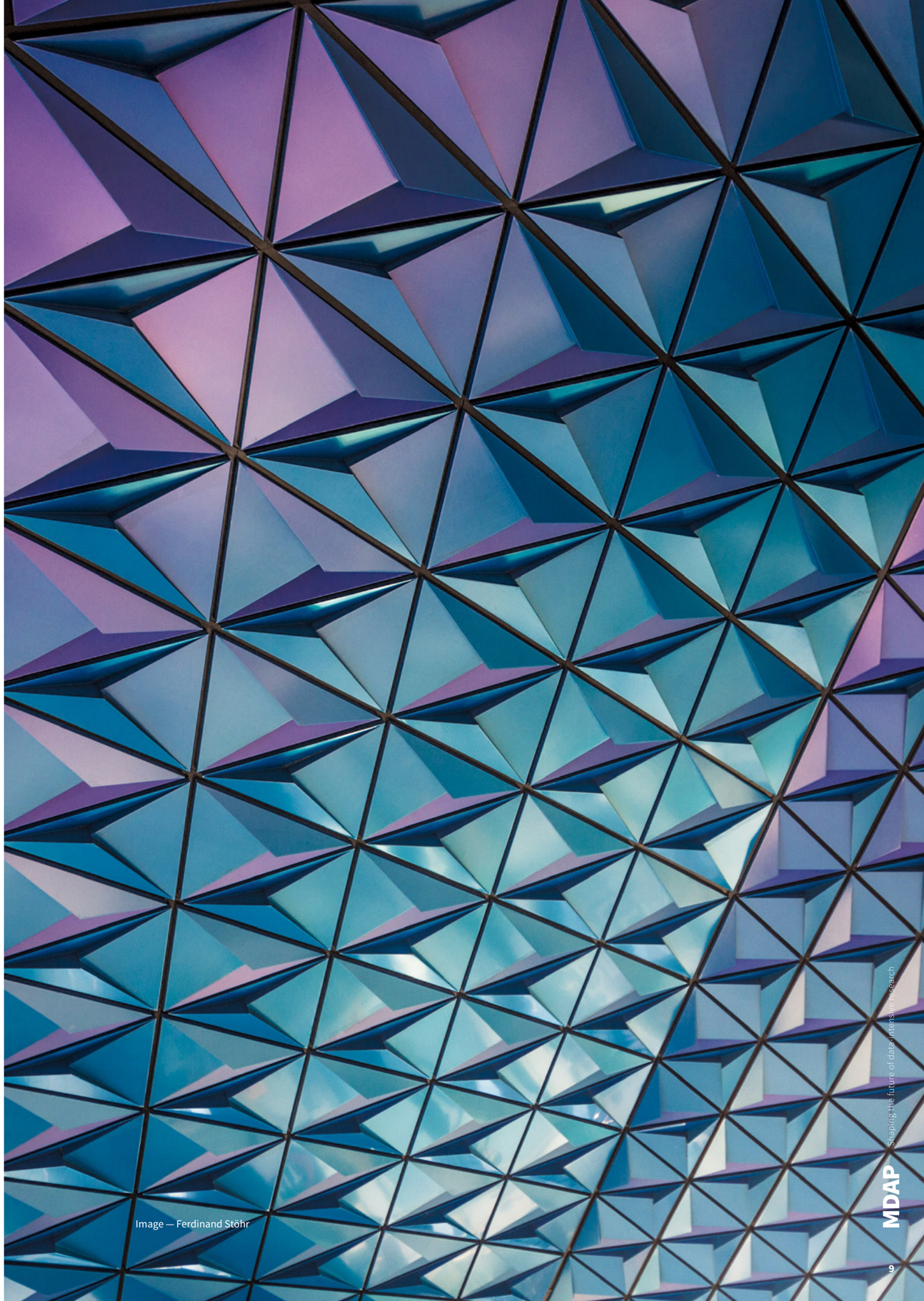
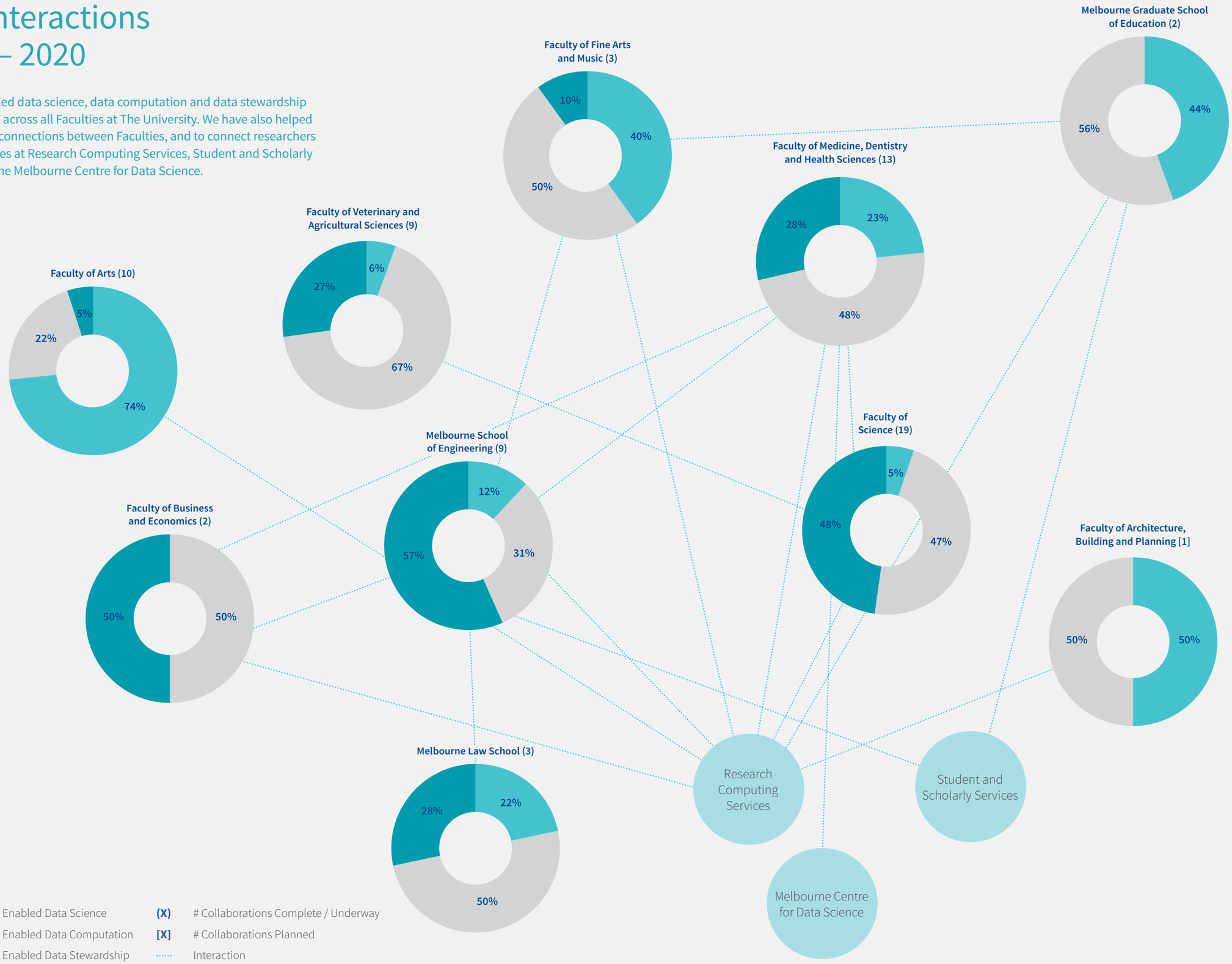


Image — Ferdinand Stöhr

Our Interactions 2019 – 2020

We have enabled data science, data computation and data stewardship collaborations across all Faculties at The University. We have also helped to create new connections between Faculties, and to connect researchers to other services at Research Computing Services, Student and Scholarly Services and the Melbourne Centre for Data Science.



LEGEND

- % MDAP Enabled Data Science
- % MDAP Enabled Data Computation
- % MDAP Enabled Data Stewardship
- **(X)** # Collaborations Complete / Underway
- **[X]** # Collaborations Planned
- ····· Interaction

“Data in both basic and applied medical research is becoming increasingly large and complex. MDAP has provided Melbourne with extra capability in bioinformatics, machine learning and advanced data processing.”

Prof Mike McGuckin
Associate Dean of Research,
Faculty of Medicine, Dentistry and Health Sciences,
The University of Melbourne

Querying six billion Tweets to explore gendered hate speech

The microblogging site, Twitter, has become a global repository of contemporary culture and information. In an investigation of the drivers of gendered hate speech, researchers from the Melbourne School Of Psychological Sciences collaborated with MDAP to create techniques and workflows to query a massive collection of six billion Tweets from around the world. Their output has already begun to generate new insights and lay the foundations for ongoing research in the area.

LOCATING DATA

The first challenge for the collaborative team was to identify, amongst the billions available, subsets of Tweets that could be analysed to yield results. At the start of the study, no formal database of geolocated Tweets existed and each query involved the laborious process of transferring terabytes of data from one system to another, and then repeating that for each new query.

BUILDING A DATA PROCESSING AND VISUALISATION PIPELINE

Working closely with the researchers, MDAP developed a set of processes that automated many aspects of searching, analysing and visualising the Twitter data. The solution required joining a range of loosely-coupled components including constructing a dedicated database, a space for performing analysis and queries, and specific visualisation tools. It can now be used to extract datasets that can provide clear evidence on global patterns of gendered hate speech — a clear demonstrator of the possibilities of transdisciplinary research.

IN MERE SECONDS

As a result of the collaborative effort, it is now possible to query billions of geolocated Tweets in mere seconds with phenomenal efficiency and power. With this tool, researchers can investigate a new range of questions, for example: What factors drive domestic violence and mental illness globally? How does economic inequality change the way people think and behave online? Can Twitter chatter be analysed in ways that anticipate violent activities? Such new tools allow new research possibilities.

ENABLING FUTURE POSSIBILITIES

Chief Investigator, Dr Khandis Blake says that MDAP has helped open a wide range of future research possibilities:

> *"This would have been immensely difficult without MDAP's expertise. In our first published research using these data, we've already found a way to use misogynistic Twitter chatter to predict domestic violence outbreaks — a year in advance. We are immensely grateful to MDAP for helping us generate this important knowledge and benefit policy-makers worldwide."*

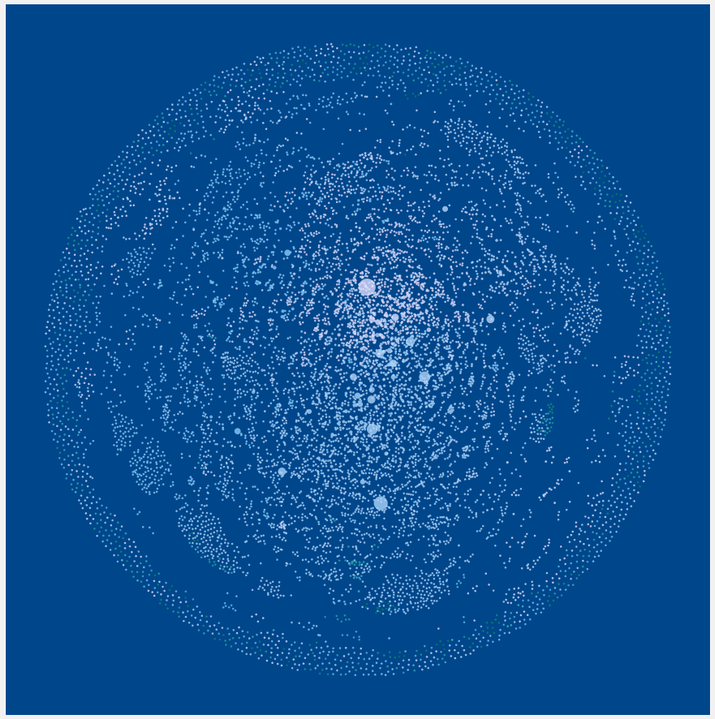
A PATHWAY FOR INTERVENTION

How people behave and present themselves online is a socially important but historically novel aspect of the human social repertoire. Yet the state of knowledge about how and why people behave as they do online remains embryonic. In particular, the uncivil, sexist, and often abusive behaviours people engage in on social media have profound consequences but are inadequately explained.

The workflow MDAP developed will deliver novel and rigorous tests of theory concerning the drivers of these behaviours. We expect the research to lead on to both a sounder understanding of where bad behaviour online comes from and to policy and other interventions to ameliorate such behaviour.

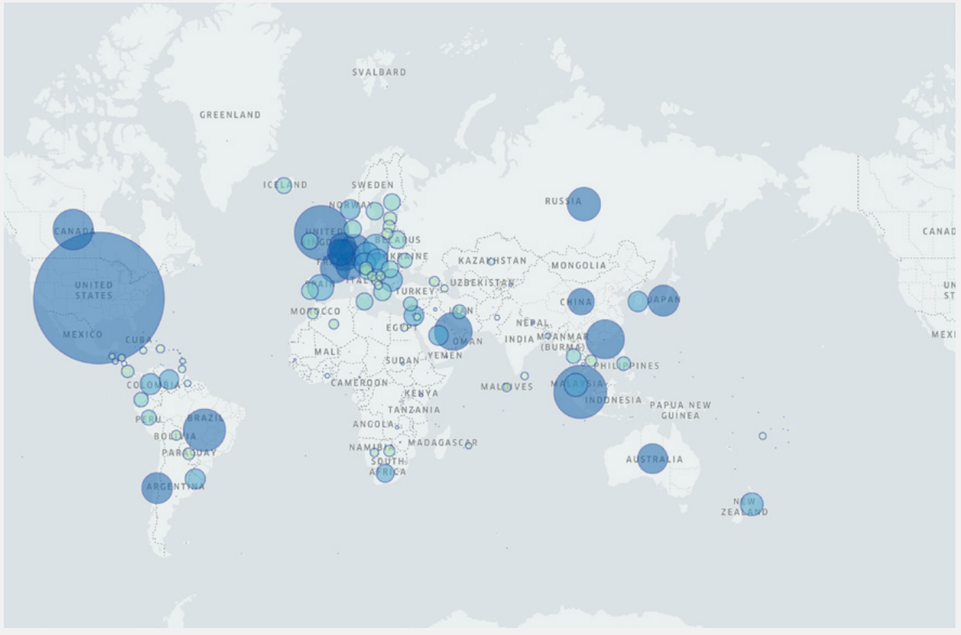
Collaborators —

Dr Khandis Blake and Dr Tim Schofield from the **Faculty of Medicine, Dentistry and Health Sciences**, and
Dr Daniel Russo-Batterham, Kim Doyle and Saket Khandelwal from the **Melbourne Data Analytics Platform**



Gephi visualisation of 50,000 Tweets themed by 'masculinity' including keywords and hashtags like #pecs, #bicepsworkout, #alphamale, #strongman, #spartan, #musclemen, and #trainharder. Each point here represents a hashtag. Hashtags that frequently appear in similar contexts are clustered together. Colours help to separate key groups of clusters, while more common hashtags are depicted by larger nodes. We can use these features to identify the many ways that masculinity is conveyed across the collection.

Image —
Dr Daniel Russo-Batterham



A map of the world showing the number of Tweets containing sexy selfie hashtags across 113 nations, adjusting for English language Tweet posting frequency, population size, and human development. Results show that sexy selfies are driven by economic conditions in which people worry about their social standing.

Image — Dr Khandis Blake

Supporting Indigenous governance of Indigenous data

Aboriginal self-determination means that Aboriginal people take charge of their own future and strategies to achieve parity in terms of health, economic, social, and justice outcomes. A prerequisite is that Aboriginal people determine the narrative about them: Setting their own measures of parity and prosperity, tracking progress towards them, and influencing policies that affect them.

In the Goulburn Murray Region, Kaiela Institute and its precursor organisations have pursued the goal of self-determination for more than three decades. Working together, MDAP, Scholarly Services, the Indigenous Data Network (IDN) and Kaiela Institute have taken a big step towards the creation of an Aboriginal data governance framework and associated policies. Along the way, each member of the team has learned a lot about the power of collaboration.

OWNING NARRATIVE

A large and increasing number of public and private agencies collect data about Aboriginal and Torres Strait Islander people, which is then used to guide the design and implementation of national, state, regional, and local policies, as well as their evaluation. However, Aboriginal and Torres Strait Islander people have little or no control over the data collected about them, nor over the construction and application of these policies and measures drawn from it. This can, and has, led to both inefficiencies in resource allocation and negative consequences such as the now deeply entrenched framing of Aboriginal and Torres Strait Islander people through a lens of deficit and disadvantage.

The achievement of parity is reliant upon the development of a community-owned definition of prosperity, community-set measures of progress towards this vision, community-led data collection, and community-governed access to, analysis of, and use of these data.

COMBINING EXPERTISE

The IDN and Kaiela Institute formed a relationship to advance Aboriginal community control over their data. Based in northern Victoria, Kaiela Institute is an Aboriginal policy think tank and a backbone for the Goulburn-Murray Empowered Communities initiative. It works closely with Aboriginal-led community organisations throughout the region.

Enabled by the Petascale Campus Initiative (PCI) and the IDN, a writing team with representatives from MDAP, Scholarly Services, and Kaiela Institute came together to design a Data Governance Framework that addressed academic, regulatory, and cultural dimensions.

The MDAP Research Data Stewards brought extensive experience in developing frameworks for data governance. Together, we undertook extensive research in finding the right set of guiding principles and culturally appropriate frameworks to develop resources to support a "two-way governance" model that balances Kaiela Institute's need to address community cultural protocol as well as corporate and data governance protocols.

Collaborators —

Dr James Rose, A/Prof Douglas Boyle, Karyn Ferguson and Raelene Nixon from the **Faculty of Medicine, Dentistry and Health Sciences**; Jasmine Graham and Sönke Tremper from the **Kaiela Institute**; Priyanka Pillai, Dr Kristal Spreadborough and Dr Aleks Michalewicz from the **Melbourne Data Analytics Platform**; and Gene Melzack from The University's **Student and Scholarly Services**

BUILDING CAPACITY FOR SELF-GOVERNANCE

MDAP Research Data Stewards worked closely with the team at Kaiela Institute to assist in the development of a data strategy, including data governance framework and data unit terms of reference for the Algabonyah Data & Research Unit, a self-governing unit established by the Kaiela Institute.

The self-governing data strategy will provide the mechanisms for Indigenous communities in the Goulburn Murray Region to manage, control, protect and honour their information both externally and internally. The capability of managing and controlling data assets will facilitate evidence-based decision making. The resources developed by the team will be submitted to the Victorian Department of Premier and Cabinet.

Kaiela Institute investigator, Sönke Tremper reflects on this project on behalf of collaborators:

> *"Working with the MDAP team really highlighted two things for me: The power of expertise, and the power of collaboration with the intent to improve the status quo and without ego. This is rare, and it showcased that wonderful things can happen when those with power understand that their biggest possible contribution is to share it."*

KAIELA INSTITUTE

Indigenous Data Network:
PUTTING INDIGENOUS DATA BACK IN COMMUNITY HANDS

Data relating to Aboriginal and Torres Strait Islander people, together with decisions and capabilities made using it must be governed by Aboriginal and Torres Strait Islander people. Learn more about how they are empowering Aboriginal communities to build a bright and sustainable future for those communities: **go.unimelb.edu.au/8z6j | www.kaielainstitute.org.au**

Images — The University of Melbourne, The Indigenous Data Network and Kaiela Institute

Applying machine learning to speech data to reduce neurodegenerative diseases

Neurodegenerative diseases affect millions of people around the world and place an enormous burden on global healthcare systems. In many cases, diagnoses are made when symptoms are advanced; if diseases are detected early however, valuable interventions can be made sooner and lead to better long-term outcomes. Working in close collaboration with expert researchers, MDAP applied machine learning techniques to speech data that has resulted in the ability to quickly classify Friedreich ataxia and multiple sclerosis with an impressive 85% accuracy.

A NEED TO IDENTIFY SYMPTOMS EARLY AND QUICKLY

Speech impairments often precede other symptoms of neurodegenerative disease. Various acoustic features of speech offer information that can identify different pathological conditions before more serious symptoms appear. Therefore, sophisticated speech biometrics can provide a preclinical diagnostic tool for neurodegenerative diseases. Early identification of disease onset can assist medical teams in providing timely symptomatic treatment for patients.

APPLYING MACHINE LEARNING TECHNIQUES TO SPEECH DATA

Can machine learning approaches determine which acoustic features can accurately identify neurodegenerative disease? Working as a cohesive team, the researchers used signal processing to select relevant acoustic features of speech data from healthy people and those with Friedreich ataxia, multiple sclerosis, and Huntington's disease. Learning from each other through experimentation and iterative refinement of techniques, the team were able to identify early symptoms of neurodegenerative disease.

ACHIEVING HIGH ACCURACY

We found that the best sets of acoustic features that could distinguish groups were related to variability of speech rate and duration of vowel sounds. These preliminary results show that acoustic features of speech data are strong indicators of neurodegenerative disease.

A PLATFORM FOR FUTURE STUDIES

Chief Investigator, Dr Benjamin Schultz reflects:

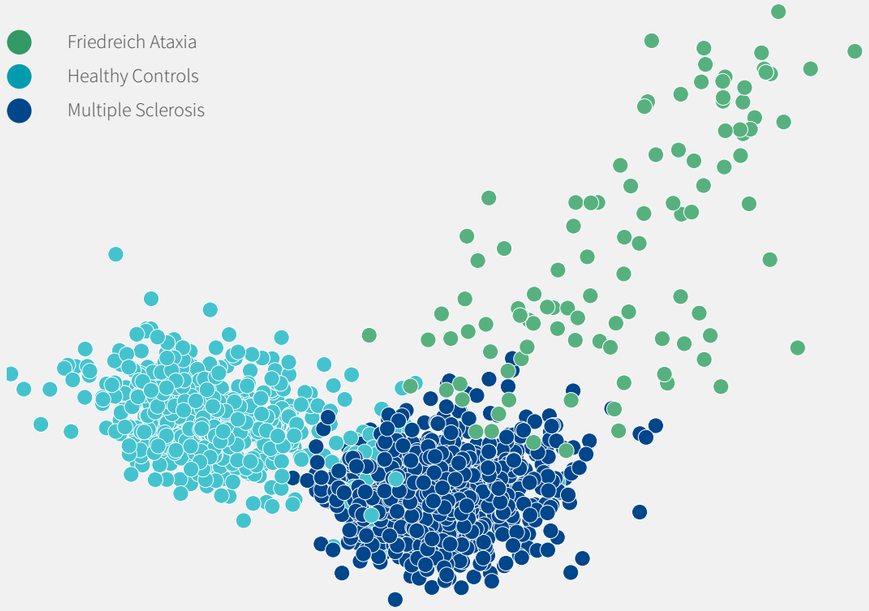
> *"MDAP has provided powerful machine learning infrastructure that the team can use for future work. As we collect more data, we will be able to better identify a wider range of neurodegenerative diseases with increased accuracy."*

PAVING THE WAY FOR IMPACT, ESPECIALLY IN REMOTE HEALTHCARE

This research is an important step in identifying neurodegenerative diseases using speech data. As speech data can be collected remotely, this research has strong implications for telehealth and remote healthcare. Tools that monitor and identify neurodegenerative diseases without invasive methods can reduce risks associated with visiting medical clinics and the financial burden on the healthcare system, as well as patients and their carers. Remote healthcare is particularly important in the wake of the recent COVID-19 pandemic as people with neurodegenerative disease are an at-risk group.

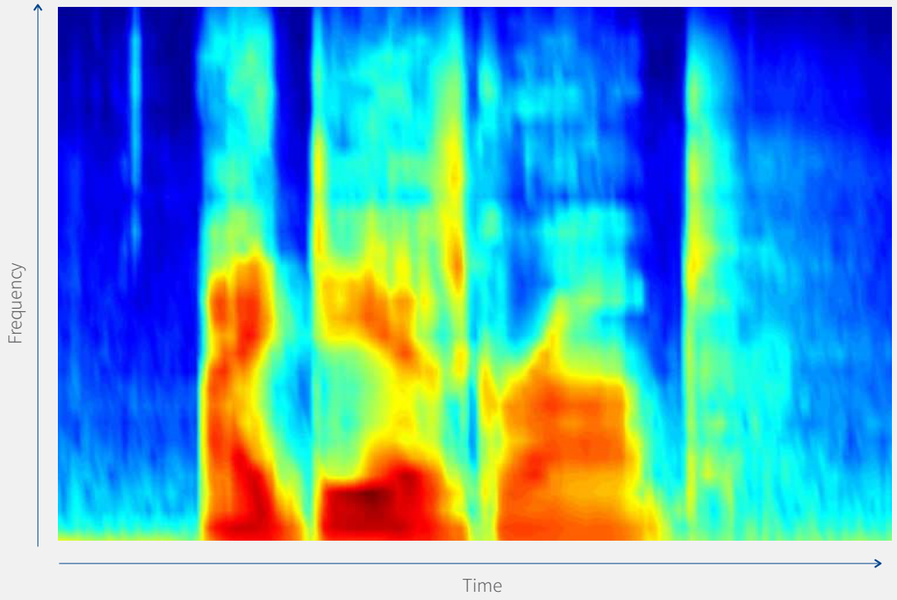
Collaborators —

A collaboration between Dr Benjamin Schultz and Prof Adam Vogel from the **Faculty of Medicine, Dentistry and Health Sciences**; and Zaher Joukhadar, Usha Nattala and Dr Maria del Mar Quiroga from the **Melbourne Data Analytics Platform**



Visualising speech data from people with neurodegenerative diseases using linear discriminant analysis to reveal definitive clusters for each of the diseases in question.

Image — Dr Maria del Mar Quiroga



Visualisation of speech data showing voice pitch (Mel-frequency; vertical axis) varying with time. Data is coloured by 'power' at different 'pitches', an important marker of speech timing drawn on in this study. Hotter colours (red) represent more power and cooler colours (blue) represent less power.

Image — Dr Benjamin Schultz

Image — Pierre Châtel-Innocenti

“The humanities and social sciences are increasingly exploring and applying data intensive methods to support their research. MDAP has provided a platform for our researchers to explore new techniques, build capability and improve data publishing mechanisms and community engagement.”

A/Prof Andrea Rizzi
Associate Dean of Research,
Faculty of Arts,
The University of Melbourne

Unleashing the power of information extraction in the legal domain

Extracting important legal principles from court judgements and applying them to new cases is a fundamental process of the common law system. Doing this quickly and accurately is a challenge for legal researchers and practitioners in Australia, as well as in all other common law countries. In this project, we start to develop facets of 'LawTech' — in particular, applying machine learning to extract information from court judgements — in a groundbreaking research for The University.

CALLING FOR AN INFORMATION EXTRACTION SUPERPOWER

In Australia, much of the law resides not in the statute books but in the hundreds of thousands of court judgments that collectively form common law. Rules of law develop incrementally — and sometimes rapidly — with new judgments and rulings constantly refining, extending, and establishing legal principles. With busy practices and faced with a sheer volume of judgments, lawyers and legal researchers can often only read and digest a small fraction of the potentially relevant judgments to a particular legal issue, possibly missing out on pertinent points of law and the overarching development of legal trends. Providing an ability to read through thousands of legal documents and extract relevant portions of text would be truly transformative throughout the domain of law.

A MILESTONE PROTOTYPE

As a step in this direction, this collaboration explores the application of machine learning to the classification of segments of texts from court judgements, into pre-chosen categories which are based on how lawyers use court judgements to prepare for new cases. A machine learning pipeline able to perform such a classification could markup thousands of court judgements in a relatively short amount of time, thus helping lawyers and legal researchers to extract the parts of court judgements most relevant to their work much more quickly.

The collaborative research team first reviewed existing machine learning methods used for legal text classification, and then built a prototype pipeline forthe classification of court judgement texts. The prototype used a set of judgements from the High Court as its dataset, and the Melbourne Law School provided the annotation of the judgement texts (i.e. the human legal-expert classification of the texts), which was used in the training, validation and testing of the pipeline.

As a part of the exploration, the prototype pipeline benchmarked and compared the performance of multiple types of machine learning text classification models in the context of the collaboration's research problem.

The team then reached out to experts in natural language processing (NLP) from the School of Computing and Information Systems, who have subsequently come onboard as collaborators. Work is now underway to build a fully developed pipeline, with further testing to be done with a larger set of Australian court judgments.

VISION, ENTHUSIASM AND COMMITMENT

Investigator, Hui Chia reflects on the collaboration:

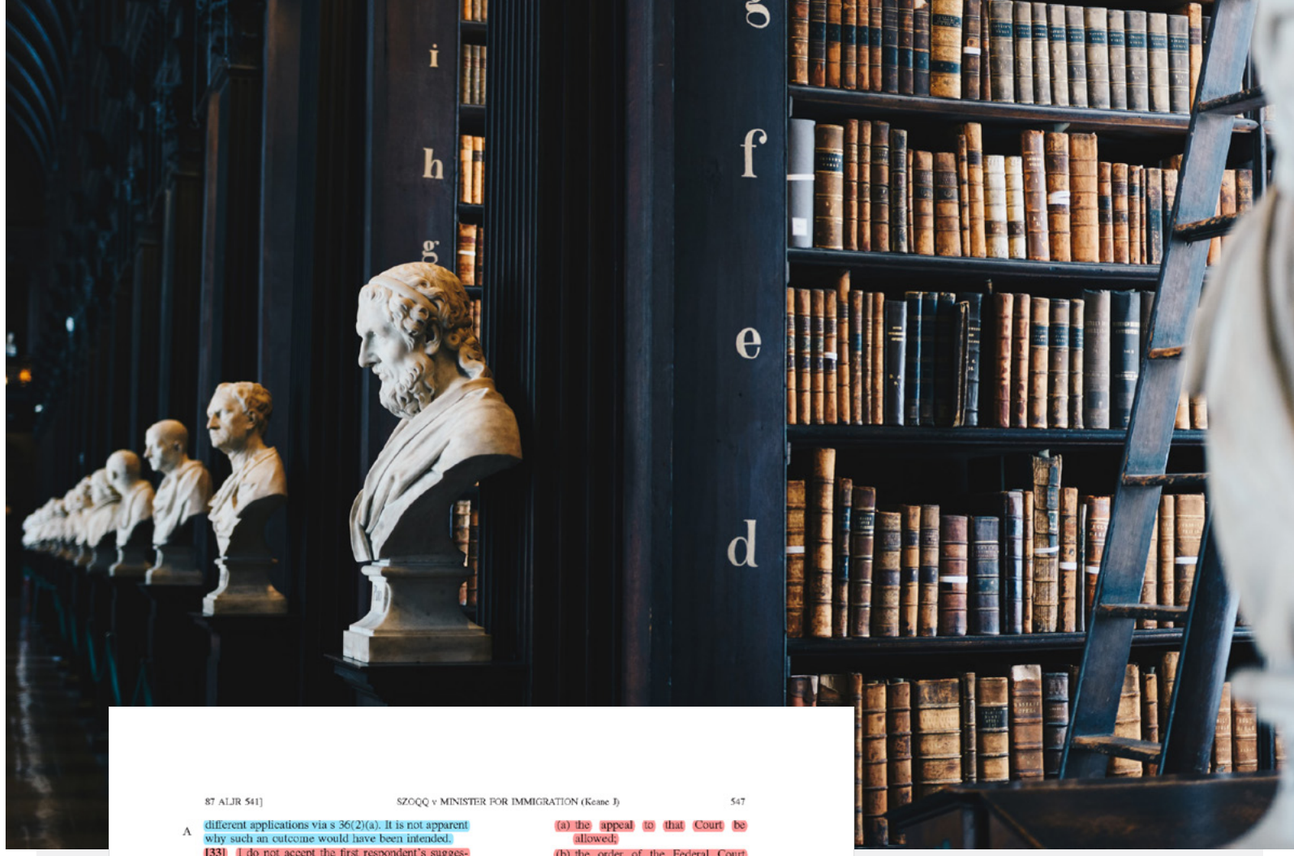
"I have been exceedingly impressed by the MDAP project team: firstly, the expertise and efficiency by which they have progressed with the project, and secondly, the enthusiasm and vision that they have demonstrated for the high-level goal of the project. Their commitment has given the project real momentum. I must also mention the outstanding dedication and professionalism shown by interns, Rohit Kumar Gupta and Saket Khandelwal. I am thrilled to be working with MDAP on this project and look forward to our continuing collaboration."

FORGING A NEW PATHWAY

MDAP is excited to be leading the way in building machine learning and data analytics capability for the legal domain, and helping to establish The University as a leading innovator in the field of legal NLP. Further research and publications will be produced with the Melbourne Law School team. As the research continues into its next phase, with collaborators from the School of Computing and Information Systems (Melbourne School of Engineering) coming onboard, this work is also forging new cross-disciplinary and cross-Faculty research partnerships within The University.

Collaborators —

Prof Jeannie Paterson and Hui Chia from **Melbourne Law School**; and Geordie Zhang, Priyanka Pillai, Dr Daniel Russo-Batterham, Kim Doyle, Dr Emily Fitzgerald, Rohit Gupta and Saket Khandelwal from the **Melbourne Data Analytics Platform**



87 ALJR 541] SZOQQ v MINISTER FOR IMMIGRATION (Keane J) 547

different applications via s 36(2)(a). It is not apparent why such an outcome would have been intended.

[33] I do not accept the first respondent's suggestion that, unless s 91U is understood as working an alteration of s 36(2)(a), it would be left with no work to do. The appellant argues that it might be considered for the purposes of s 55(1)(a)(iii) and s 501 of the Act. The first respondent disputes that s 91U is relevant for the purposes of s 501 because that section does not refer in terms to "particularly serious crime". However that may be, s 91U can readily be seen to be apt to aid the operation of s 499. And, bearing in mind the terms of s 499(2), it could also aid the making of regulations under s 31(3) of the Act.

[34] In addition, pursuant to s 31(3) a regulation might prescribe, as an additional criterion for the grant of a protection visa of the class referred to in s 36 of the Act, that a person not be a person to whom the "particularly serious crime" limb of Art 33(2) of the Refugees Convention applies.

[35] As to the first respondent's second point, the Minister's Second Reading Speech sheds no light on the problem of present concern. That is hardly surprising, given that it predates the decision in *NAGV*. And as I have noted, in this Court the first respondent did not seek to challenge the correctness of *NAGV*.

Conclusion and orders

[36] The appellant's submission that the proceedings below miscarried must be accepted. The appeal must be allowed to enable the appellant's application for a protection visa to be considered according to law.

[37] The appellant seeks his costs of the proceedings in the courts below and in this Court. Having regard to the failure of the parties to identify the point on which the appeal to this Court turns, and the arid exercise in which the parties have involved the courts before which this matter has come, I have given consideration to the possibility that there should be no order as to costs. I have, however, come to the conclusion that, bearing in mind that the parties were equally at fault in this regard, this consideration is not sufficient to displace the usual rule that costs should follow the event.

[38] I would make the following orders:

1. Appeal allowed.
2. Set aside the orders of the Full Court of the Federal Court of Australia dated 23 March 2012 and, in their place, order that:
   (a) the appeal to that Court be allowed;
   (b) the order of the Federal Court dated 4 November 2011 be set aside and, in its place, order that:
      (i) a writ of certiorari issue directed to the second respondent quashing its decision dated 2 September 2010;
      (ii) a writ of mandamus issue directed to the second respondent requiring it to review, according to law, the decision made by a delegate of the first respondent on 26 May 2009 to refuse the appellant a Protection (Class XA) visa; and
      (iii) the first respondent pay the appellant's costs in the Federal Court; and
   (c) the first respondent pay the appellant's costs in the Full Court of the Federal Court.
3. First respondent to pay the appellant's costs in this Court.

1. Appeal allowed.

2. Set aside the orders of the Full Court of the Federal Court of Australia dated 23 March 2012 and, in their place, order that:

(a) the appeal to that Court be allowed;

(b) the order of the Federal Court dated 4 November 2011 be set aside and, in its place, order that:

(i) a writ of certiorari issue directed to the second respondent quashing its decision dated 2 September 2010;

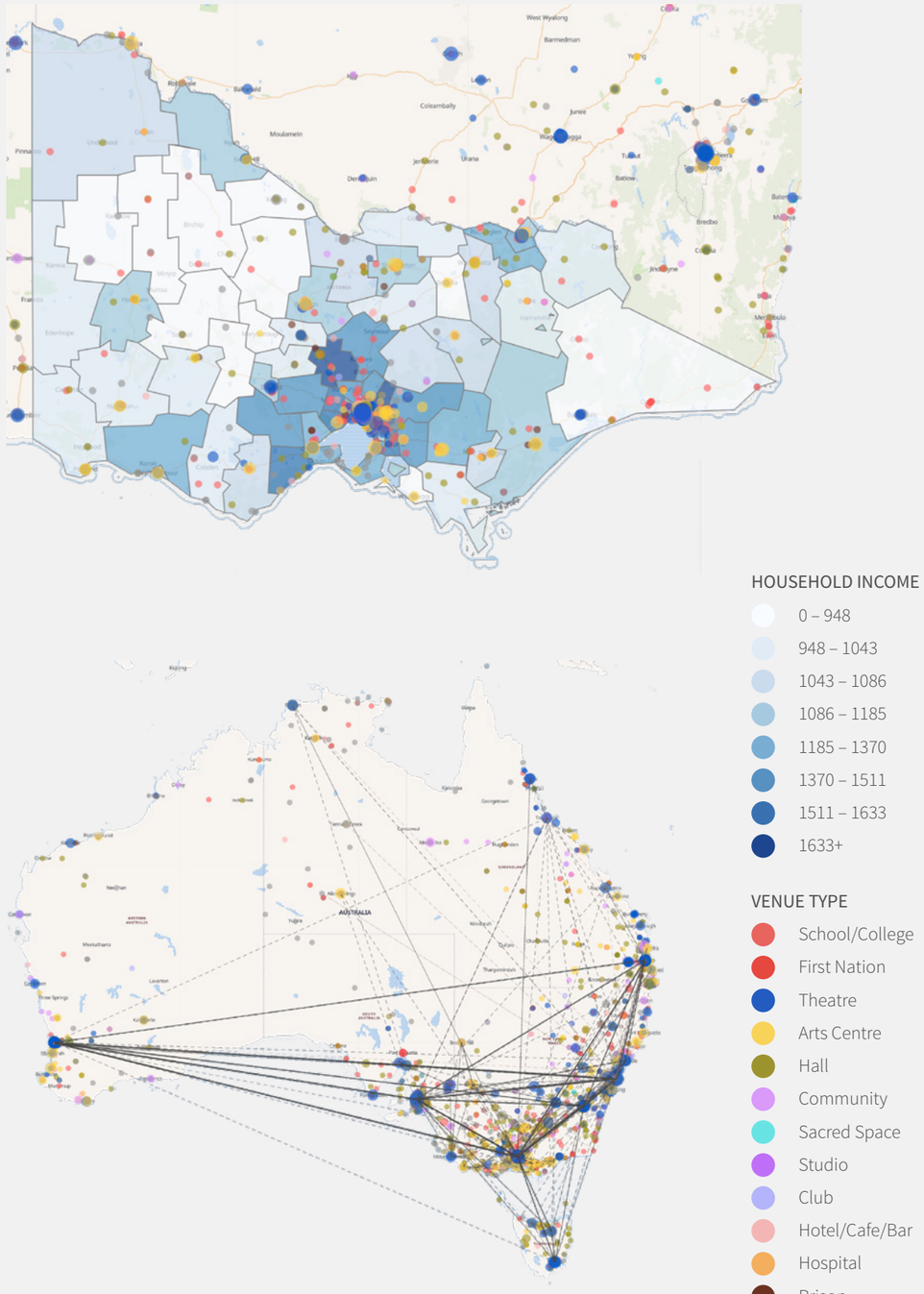
(ii) a writ of mandamus issue directed to the second respondent requiring it to review, according to law, the decision made by a delegate of the first respondent on 26 May 2009 to refuse the appellant a Protection (Class XA) visa; and

(iii) the first respondent pay the appellant's costs in the Federal Court; and

Image — Giammarco Boscaro

Researchers from Law, MDAP and now Engineering at The University are developing machine learning models to quickly read, annotate and extract important elements of court judgements that can inform a new case at hand. Something of a superpower for legal practitioners. Pictured: human annotated data that is used to train, test, and validate these machine learning models. Facts of the case are highlighted in green, judicial reasoning in blue, and the conclusion in red

Image — Extract from SZOQQ v Minister for Immigration and Border Protection [2013], HCA12, 541; from the High Court of Australia public database.



Visualising regional theatre across Victoria (top) and nationally (bottom) including touring paths (lines), venue types (dots) and household income (blue block colours).

Images — Usha Nattala

Visualising regional theatre to enhance access for young people

Young people growing up in rural areas can miss out on important personal and career development opportunities such as those made possible through exposure to live theatre. Working with researchers from the Digital Studio and Creative Convergence, the MDAP team helped to revitalise the visualisation tool CIRCUIT, which highlights gaps in regional theatre shows.

ADDRESSING ACCESS TO THEATRE EDUCATION

CIRCUIT was developed as part of an ARC Linkage Project. At the request of Bell Shakespeare Company, Creative Convergence began looking at the impact of theatre upon the lives of young people in regional Victoria. To understand where theatre was happening across the state of Victoria, they particularly wanted to know more about the difference between touring theatre and theatre that was developed in the regions, as well as the distance that students travelled to regularly see theatre.

FINDING DATA FOR MEANINGFUL VISUALISATION

Based on a preliminary data map from AusStage, the research team collected more detailed information from theatre companies about their programs over the last decade. The collaborators met to discuss and design the mapping tool to create visualisations of touring circuits by company, venue, location and genre, and then added essential socio-economic data from the Australian Bureau of Statistics. The many layers of the map provide information that assists theatre companies, local communities and researchers to situate theatre in a particular social or political context.

ENABLING CONVERGENCE

Now an interactive and accessible mapping tool, CIRCUIT is used by theatre scholars, the creative industry, government and other researchers to visualise theatre touring across Victoria in detail, as well as Australia-wide from 1965 to the present. Importantly, CIRCUIT helps theatre industry partners and the researchers to examine how theatre organisations might 'converge' or better collaborate across the sector and across regional areas. It also shows the State, Territory and Federal governments which theatre companies or venues are delivering what kind of product, and information about areas are not well serviced; and how this has changed over time.

Chief investigator, Prof Rachel Fensham reflects:

> *"Usha has been amazingly inventive in adding her data science expertise to this project that was a simple but not fully realisable objective; to deliver a mapping tool that would be flexible enough for a range of users but also precise and meaningful for research. She managed the MDAP interns with enthusiasm and made them feel they could really contribute to our understanding about the impact of theatre on young people generated by the Creative Convergence team of arts and humanities scholars."*

ENABLING FUTURE PLANNING

CIRCUIT is now seen as a valuable resource that enables the theatre industry to gain an accurate picture of touring schedules across Australia. In an industry devastated by recent bushfires, reduced funding and the effects of COVID-19 restrictions on live performance, CIRCUIT benefits the theatre industry through the development of an accurate and interactive data visualisation tool built for future planning.

Collaborators —

Prof Rachel Fensham, Dr Trent Ryan and Septi Rito Tombe from The University's **Digital Studio;** Dr Lynne Kent from the **ARC Linkage Creative Convergence Project**; and Usha Nattala and Navaneeth Rajagopalan from the **Melbourne Data Analytics Platform**

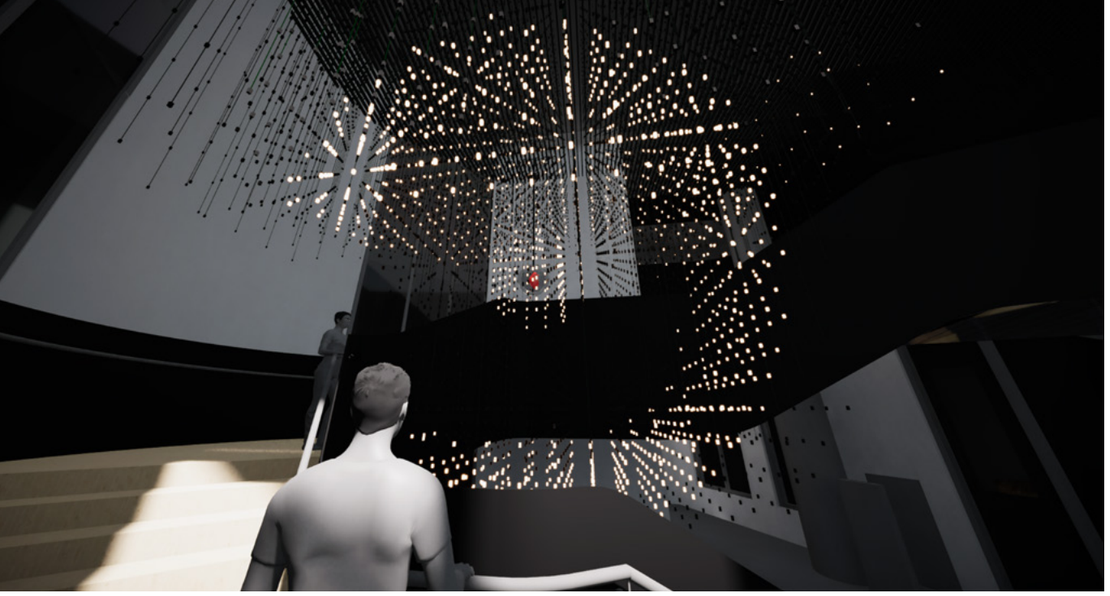


Image — EXP Productions

Designing a building with Heart

The construction of new spaces invites new possibilities. Flowing through discussions in the design of the innovative Melbourne Connect building, artists, architects and data scientists agreed that inspiration must lie at the heart of their work. Eventually, the idea of having a heart — one brimming with sensitivity and wired for input — took shape and will soon become realised through this collaborative project.

HOW DO YOU GIVE A BUILDING A HEART?

Melbourne Connect brims with hidden sensors that provide a torrent of live data about the building and the community it supports. In their vision for the artwork called 'The Heart' the collaborative team sought to empathise with the building and somehow feel what it feels as it learns and ages long into the future.

TRANSLATING DATA INTO A HEART 'PULSE'

How can data become a pulse? Using artificial intelligence and machine learning algorithms, The Heart compares the present moment with its previous experiences and seeks to link its experiences to contextualise the present. The Heart will dream during the night — deciding what to remember and forget.

CREATING THE INSTALLATION

With the full creative team including lighting designers, Unity developers, architects, builders, creative technologists and glass sculptors, the artist and MDAP have come to realise The Heart. The 'heart node' will be made from glass recycled from the old Royal Women's Hospital operating theatre lights. This will be surrounded by a 500 cubic-metre array of 16,540 custom-built lights that can each be individually addressed.

REFLECTING ON THE PROJECT SO FAR

Chief investigator, Robert Walton explains the integral part that MDAP played:

> *"We can only do this project that combines data, machine learning, artificial intelligence in a really cool, and unprecedented way here and now because of MDAP's expertise. MDAP allows us to think about big, complicated projects that might have an impact on people and the world. We need MDAP and Zaher!"*

INSPIRING A FUTURE OF SMART AND SENSITIVE ECOSYSTEMS

We are creating a permanent artwork that will soon be built into the physical infrastructure of a flagship development. It will be integral to The University: constantly prompting us to reflect on ourselves, our emotions and our environment such as to enhance empathy, gratefulness, and thoughtful actions in the world. A building constantly offering us its heart — a risky gesture of trust. But it also provokes us to work with heart, and dare to bring our whole selves to work at The University.

Collaborators —

Dr Robert Walton from the **Faculty of Fine Arts and Music** and Zaher Joukhadar from the **Melbourne Data Analytics Platform**

"As the amount of data collected and analysed for research purposes rapidly increases, the need for a workforce of computational experts with the communication skills to work across all disciplines has never been more urgent.

MDAP has realised Melbourne's goal of building such a workforce, ensuring that these cutting edge skills remain within The University's research capability."

Prof Rachel Webster AO FAA

School of Physics, Faculty of Science
The University of Melbourne

Processing Big Data to enhance global and regional climate projections

What will the weather be like, say, five or ten over even 20 years from today? Based on the Sixth Coupled Model Intercomparison Project (CMIP6), climate scientists and MDAP worked together to enhance climate projections and inform decision making. Using combined skills, the collaborative team made sets of Big Data accessible, built a probabilistic climate projections tool and discovered new insights.

AN INITIAL CHALLENGE

One of the first challenges the team faced was that projections were only available in a specialised binary data format. Such a format rendered projections nearly impossible to read, let alone use or understand.

To make the data accessible required the team to make deft combinations of specialist disciplinary knowledge and advanced technical training.

COMBINING EXPERTISE

To undertake the complex work, the researchers combined expertise in physics with deep understanding of statistics, programming and data visualisation. The climate scientists were able to provide a physical understanding of the data, and MDAP built the computing and statistical tools that were needed. By meeting regularly, the interdisciplinary collaborative team was able to draw on combined strengths to solve problems. The teams produced output quickly and, through data visualisation, was able to share data and results with an audience well beyond the immediate research community.

NEWLY REFINED DATA, A NEW PROJECTION TOOL, AND MULTIPLE PAPERS

The collaboration produced three key outputs. First, a new combined dataset, derived from an original set of CMIP6 data, which is now available as open access and is hosted at The University. Though still in the final stages of refinement, the team created a probabilistic climate projection tool for use around the world.

The team is also now set to publish multiple papers, with the MDAP team as co-authors, that document the novel techniques, data and models produced by the collaboration. Such work continues to bolster the reputation of The University as a world leader in probabilistic climate modelling.

Chief investigator, A/Prof Malte Meinshausen says data does not get much bigger than climate model output:

> *"Thanks to MDAP's help, we were able to tackle the 15 petabytes of data (that's 15 billion gigabytes or ~60 million macbooks worth of data) included in the CMIP6 archive. Without their help, we could not have comprehensively accessed, processed and analysed the archive within the tight timelines required."*

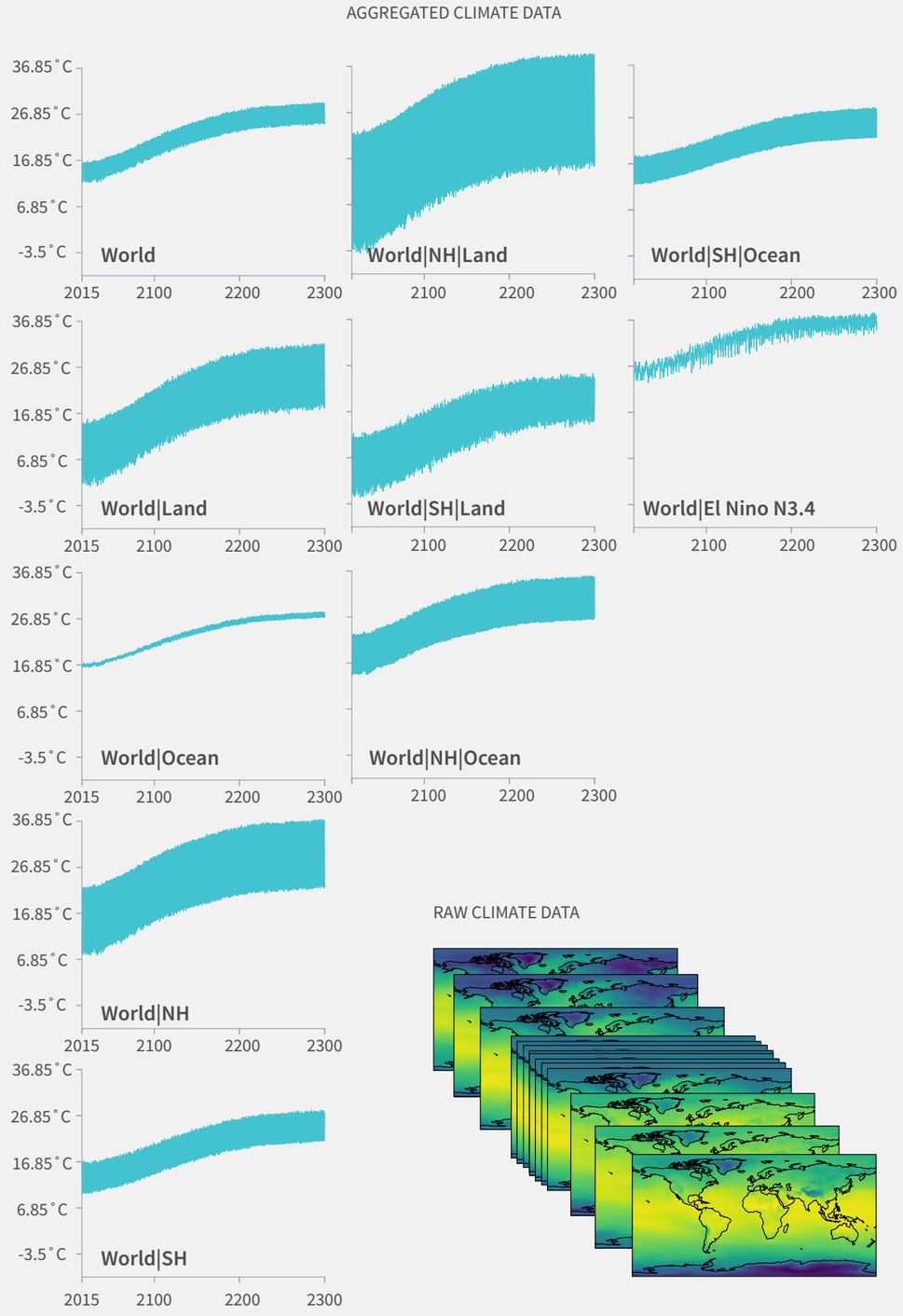
OPENLY ACCESSIBLE, GLOBALLY IMPACTFUL

At its core, the project provides a community service to researchers who work on climate change issues. With the data now openly accessible online (via website: cmip6.science.unimelb.edu.au), such work benefits both The University scholars in Geography, Earth Science and Engineering, as well those around the globe.

Working closely as a team, the researchers have added an extra layer of sophistication in data processing and visualisation that is well above and beyond what is currently available.

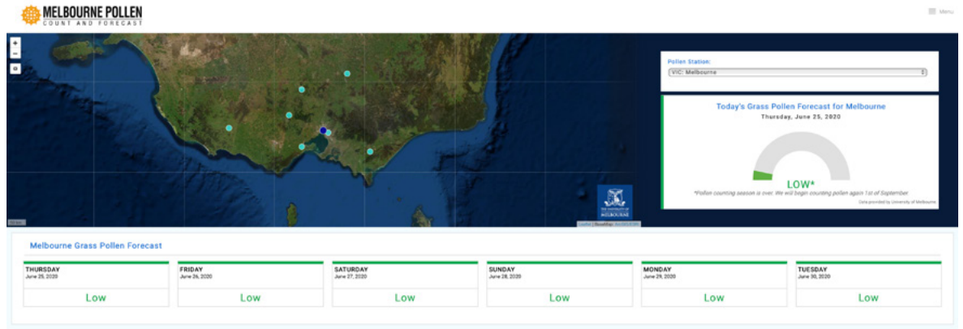
Collaborators —

A/Prof Malte Meinshausen, Mr Zebedee Nicholls and Mr Jared Lewis from the **Faculty of Science**; and Dr Simon Mutch, Dr Edoardo Tescari, Usha Nattala and Geordie Zhang from the **Melbourne Data Analytics Platform**



Example of regionally aggregated projections of the world's climate derived from output from the Sixth Coupled Model Intercomparison Project by University researchers. Each panel shows monthly, regional-mean projections for a different region of the world. A sample of the raw data used to derive these means is shown as a series of maps in the lower-right hand corner.

Image — A/Prof Malte Meinshausen and Zebedee Nicholls



Melbourne Pollen provides a daily pollen count and forecast of grass pollen levels in the air in Victoria around Melbourne for the next 6 days. This free service benefits members of the general public who suffer from the allergy conditions of hay fever and seasonal asthma.

Image — Melbourne Pollen

Image — Thomas Kinto

Improving Victoria's pollen forecasts

Based in the School of BioSciences, Melbourne Pollen has a decades-long record of Melbourne's grass pollen levels. MDAP have joined forces with Melbourne Pollen to apply powerful machine learning algorithms that more accurately forecast grass pollen levels.

SEEKING BETTER FORECASTS

Given the health risks pollen allergies pose, it is critical that forecasts convey meaning and are accurate. In cities like Melbourne with a history of thunderstorm asthma, poor quality forecasting can have potentially catastrophic consequences. Previously, forecasting of daily grass pollen levels across Victoria relied on a small set of factors such as wind speed, temperature and the day of the week that were combined with pollen data to update predictions. But powerful machine learning methods now provide an alternative approach to grass pollen forecasting and have the ability to learn with each new observation.

APPLYING MACHINE LEARNING TO POLLEN DATA

Set on improving their grass pollen forecast model, the research team sought to find ways to better analyse extensive pollen databases, survey responses and other inputs such as satellite measurements, weather, land use and vegetation data. Collaborative work with MDAP made this possible. Through advanced algorithms and machine learning techniques, forecasting processes are now capable of identifying and extracting patterns in the data without explicit instructions or continual human intervention.

REFLECTING ON THIS PROJECT

Though still a work in progress, Melbourne Pollen is now a fully-fledged collaboration with MDAP. The team's pilot version of the machine learning forecasts have seen some very positive outcomes, and possible extensions of this project have been presented as part of the Victorian Thunderstorm Asthma Review at the Bureau of Meteorology. Melbourne Pollen's coordinator, A/Prof Ed Newbigin from the School of BioSciences, reflects on the collaborative project with MDAP:

"This is something quite unique. From simple rules of thumb a few years ago, the skill of our forecasts has constantly improved and thanks to MDAP and Usha this trajectory is set to continue into the future. That's good for our users who rely on our daily forecasts and should make our thunderstorm asthma predictions better as well. This work helps keep Victorians healthy and safe."

OPPORTUNITY FOR NATIONAL IMPACT

Combined with an enormous level of community engagement and a range of digital resources, Melbourne Pollen has scope for a broad range of collaborative research opportunities.

Producing a national grass pollen model, for example, could greatly benefit the large numbers of pollen-allergic Australians who are unable to access current forecasts currently; indeed, developing a multi-model forecasting ensemble for grass pollen in Victoria and a national model would be a first for Australia.

Collaborators —

A/Prof Ed Newbigin, Dr Edwin Lampugnani and Dr Jeremy Silver from the **Faculty of Science**; and Usha Nattala and Lingyu Tang from the **Melbourne Data Analytics Platform**

Transforming the pipeline for effective genomics research

Diseases resulting from parasitic helminths (worms) can cause major socioeconomic impacts worldwide and current treatment methods are inadequate. To improve the understanding of helminths, and to devise new treatments, researchers from the Faculty of Veterinary and Agricultural Sciences have been investigating helminth genomes in studies that require sophisticated computational methods and capacity. Now, MDAP have helped to significantly develop their population genomics 'pipeline' to accelerate The University's capacity for speedy genomics research of helminths and other organisms.

DATA PROCESSING AT SNAIL PACE

Population genomics seeks to understand the genetic relationships within and between species populations as well as the processes that lead to these relationships. Exploring genetic structures in disease-causing pathogen populations can have important implications, for instance, for understanding the response(s) of a population to selection pressures, such as environmental changes and drug treatment, and can be central to elucidating pathogen ecology and epidemiology. The use of advanced informatics is enabling population genetic analyses of DNA sequence data sets produced using advanced DNA sequencing technologies. However, data processing has been slow, and there are no highly efficient, automated pipelines for the rapid analysis of such data sets.

DEFINING PROJECT OUTCOMES

After an initial scoping period and a series of regular meetings, the collaborative team outlined key outcomes to achieve, keeping each member of the team up to date with all developments (even during the COVID-19 pandemic!). Identified challenges were dealt with quickly and efficiently, and the establishment of environments as communication channels helped the project progress easily and allowed identified issues to be addressed quickly.

DELIVERING REMARKABLE EFFICIENCY

In six months, the collaboration has delivered a fully-functional processing pipeline. Using the completed workflow, experts are now able to analyse vast amounts of population genomic data, providing deep insights into the genetic compositions of parasite populations — sometimes in just weeks, rather than months.

In addition to increased efficiency, the analysis can be applied to a wide range of pathogens, parasites and other organisms. The collaboration has demonstrated how merging complementary skills and expertise can lead to synergies, and very rapid, beneficial and significant outcomes. The work will be shared in workshops, documentation and interactions for years to come.

Chief Investigator, Prof Robin Gasser reflects on this project:

> *"Most of our collaborations are international. In terms of the collaborative projects within our institution, this has been the best and most active, short-term collaboration that we have had — the engagement between groups was ideal, as it was synergistic and accelerated a key component of our genomics program."*

GLOBAL IMPLICATIONS

The improved analysis pipeline is easy to use, and resolves the critical and computationally-demanding, specific calibration step for single nucleotide polymorphism (SNP)-calling. The latter functionality is not found in any other population genomics pipelines, which are typically limited to model organisms, such as humans and mice.

Therefore, this pipeline will be highly utilised and should have a major positive impact on population genomic studies of parasites and other non-model organisms, conducted in national and international research communities. The new pipeline enables profound phylogenomic and population genetic research of parasites and evaluations of the emergence of drug resistance in such populations, and could assist in guiding new strategies to control major parasitic diseases worldwide.

Collaborators —

Prof Robin Gasser, Dr Neil Young, Dr Pasi Korhonen and Dr Liina Kinkar from the **Faculty of Veterinary and Agricultural Sciences**; and Bobbie Shaban and Dr Noel Faux from the **Melbourne Data Analytics Platform**

A high resolution image of the human helminth parasite Trichinella spiralis. In this image cysts of the parasite have been digested and a single worm isolated and mounted. The image was created by stacking and stitching together 6 individual images at X25 magnification.

Image — David Linstead

Interning at MDAP

MDAP is keen to encourage a new generation of research data specialists and research data stewards in our quarterly internship program. We have been impressed by interns' professionalism and intellectual contributions to our research collaborations, as have our collaborators!

"My experience at MDAP was both exciting and challenging. I learned to collaborate with researchers which has greatly increased my communication and stakeholder management skills."

Maria Viji Rashmi
Master of Data Science Student

"I can compare my experience to being an intern with the Avengers! It was an intensive and expeditious hands-on training on data science."

Rohit Kumar Gupta
Master of Data Science Student

"We learned that a good data scientist spends a lot of time crunching raw data before machine learning models can be applied to gain meaningful insights."

Pratibha Pratibha
Master of Data Science Student

"I also learnt to maintain a 'beginner's mindset'; most real-world problems are not structured, so curiosity, communication and collaborative skills are crucial to tackling them."

Saket Khandelwal
Master of Data Science Student

Connecting for impact

BUILDING THE THIRD SPACE

At MDAP, we understand how reaching out and strengthening our University's community of dedicated research support staff can foster productivity, signal value and lead to greater recognition. This is the "Third Space" between professional and academic worlds.

A strong Third Space will lead to the retention of excellent staff within The University, and sustains research impact. Through a series of discussions and workshops, consultations and advice, each member of the MDAP team of Academic Specialists is charged with building this dynamic and important community.

SPARKING NEW POSSIBILITIES

What does a law professor need from a data visualization specialist? And what would an expert in the prevention of domestic violence see in billions of Tweets? At MDAP, we work with the curious, the insightful and the innovative to see how the potential can become the possible.

We work to see where data can be gathered, used, reused, and preserved in ways that spark the imagination, illuminate the edge and power the floodlights across new fields of research.

SITUATED IN A LARGER COMMUNITY

Australia fosters innovation. Up and down the latitudes and back and forth across time zones, Australians have come to see how data-intensive research will drive policy, investment and positive change throughout the nation and in our region. At MDAP, we are leading efforts to develop the Third Space across Australian institutions to help progress the nation.

Optic cable can now send data around the world in a millisecond. Massive servers can hold lakes of data and handheld devices and those on a wrist or in an operating theatre can help teach us, guide us and entertain us. Or keep us on track. At MDAP, we engage with international universities who share our vision for the Third Space as a driver of data-intensive research that harnesses digital infrastructure and builds capability.

Future

As we reflect on our first year, we know that what we have commenced at MDAP and across The University is unique in Australia, productive for researchers and powerful for society. We have glimpsed an understanding of the power of data-intensive research to transform our understanding of the world, and build new research directions across and between disciplines.

NEW, EXPANDED COLLABORATIONS

Already we have launched a new series of collaborative projects and can see the trajectory of our work reach into distant fields (Page 39). Our excitement and passion is spreading. Our expertise and skills are deepening. Working together, we will help shape future research.

OVERCOMING SILOS

As the world and The University continue to evolve and face complex challenges, it is more important than ever to sustain data-intensive and multidisciplinary research such as that enabled by MDAP and its collaborators. Together we are at the crossroads of an exciting and rewarding journey.

AN EMPIRICAL STUDY OF HOW THE US TERMINATION RIGHT OPERATES (PART OF THE AUTHOR'S INTEREST PROJECT)

Dr Rebecca Giblin, Faculty of Law

NATURAL LANGUAGE HISTORICAL MAPPING

Dr Mitchell Harrop, Faculty of Arts

MELBOURNE POLLEN RESEARCH PLATFORM

A/Prof Edward Newbigin, Faculty of Science

THE MEGALITHIC JAR SITES OF LAOS

Dr Louise Shewan, Faculty of Science

EVALUATION OF AUTOMATED DEIDENTIFICATION OF HOSPITAL AND GENERAL PRACTICE HEALTH RECORDS

Prof Karin Verspoor, Melbourne School of Engineering and A/Prof Douglas Boyle, Faculty of Medicine, Dentistry and Health Sciences

UNLOCKING PUBLISHED METAGENOMES AS A SOURCE OF INFORMATION FOR MICROBIAL EUKARYOTES

A/Prof Heroen Verbruggen, Faculty of Science

SUSTAINABLE ENERGY ADVOCACY COALITIONS

Dr Alfonso Martinez Arranz, Melbourne School of Engineering

AUTOMATING IMAGE SEGMENTATION AND MORPHOMETRIC ANALYSIS FOR ACCELERATING NEUROSCIENCE AT THE NANOSCALE

Dr Calvin Eiber, Faculty of Medicine, Dentistry and Health Science

USING NOVEL ANALYTIC METHODS TO PREDICT AND PREVENT FALLS AND FALL-RELATED INJURIES IN OLDER PERSONS WITH DEMENTIA

Prof Gustavo Duque, Faculty of Medicine, Dentistry and Health Science

CAUSALITY IN COMPLEX DYNAMICAL SYSTEMS: IMPLEMENTING CORE ALGORITHMS IN A DISTRIBUTED COMPUTING ENVIRONMENT

A/Prof Michael Zyphur, Faculty of Business and Economics

THE CREATIVE INDUSTRIES IN DAYS OF ISOLATION: A FAST PACE SHIFT TO MAKING, LEARNING AND LIVING IN A CRISIS

Dr Kathryn Coleman, Melbourne Graduate School of Education

INTRODUCING NEW ANALYSIS MODALITIES TO THE STEMFORMATICS STEM CELL ATLAS

Prof Christine Wells and Dr Jarny Choi, Faculty of Medicine, Dentistry and Health Science

DEVELOPMENT OF ARC-LP FUNDED POLYMUSE (POLYMERS IN MUSEUM COLLECTIONS) DATABASE TO FACILITATE DATA MANAGEMENT, CURATION, MIGRATION, ANALYSIS, VISUALISATION AND SHARING

Dr Petronella Nel, Faculty of Arts

EARLY MODERN WOMEN TRANSLATORS IN EUROPE (15TH-18TH C.)

Prof Véronique Duché, Faculty of Arts

MAPPING OUR EMOTIONAL LIVES: BUILDING A SEARCHABLE DATABASE OF EXPERIENCE SAMPLING DATA ON EMOTIONAL PROCESSES

Dr Elise Kalokerinos, Faculty of Medicine, Dentistry and Health Science

RURAL LAND USE CLASSIFICATION ALGORITHMS FOR EMERGENCY ANIMAL DISEASE RESPONSES

Dr Simon Firestone, Faculty of Veterinary and Agricultural Sciences

MACHINE LEARNING AND PATTERNS OF PRIMARY CARE UTILISATION IN CANCER DIAGNOSIS AND OUTCOMES

Prof Jon Emery, Faculty of Medicine, Dentistry and Health Science

Learn about these collaborations: **http://go.unimelb.edu.au/u9wr**

THE UNIVERSITY OF
MELBOURNE

Petascale
Campus Initiative

mdap.unimelb.edu.au