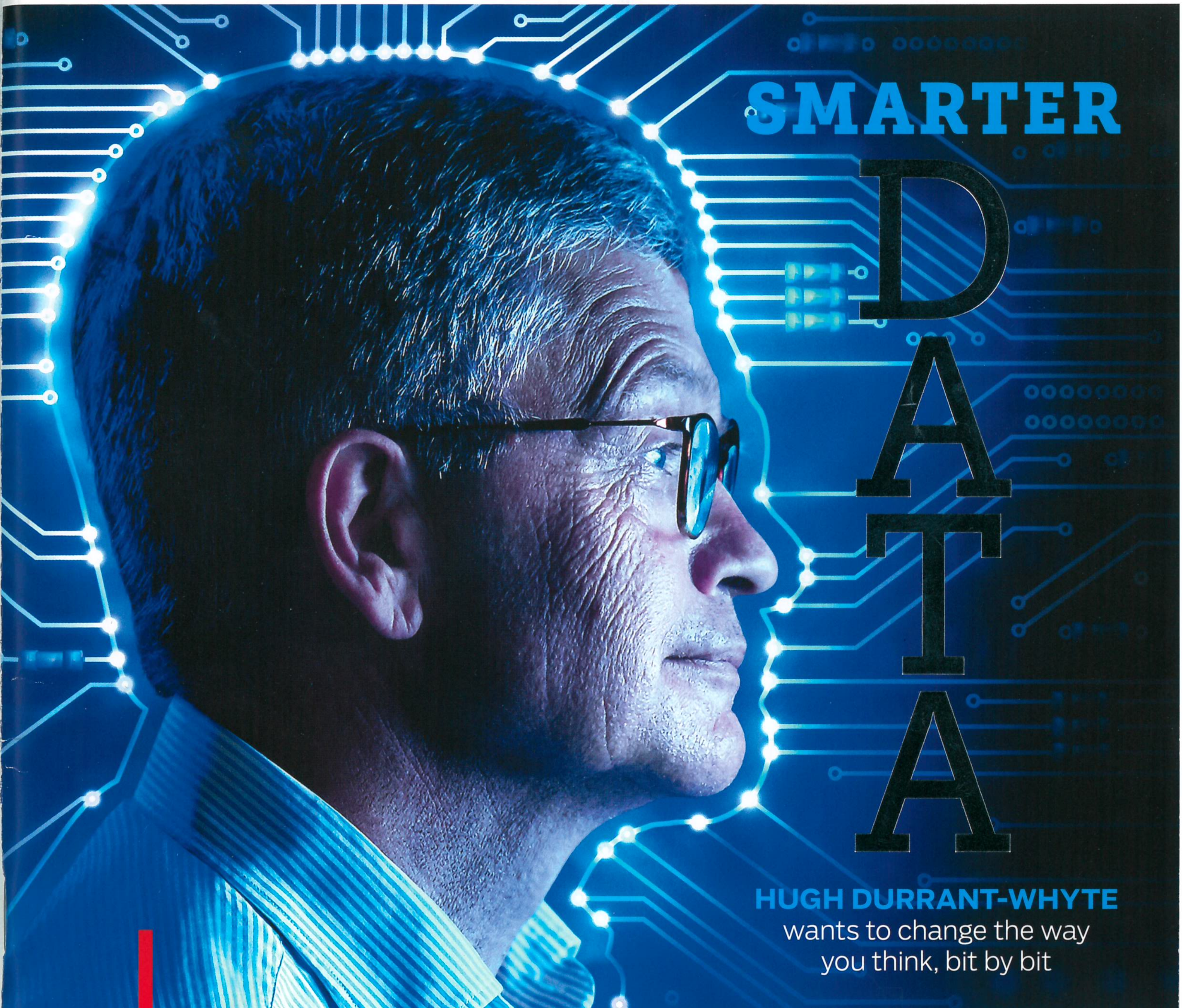
VOL. 3 NO. 3 APRIL 2017

create

ENGINEERS AUSTRALIA

ENGINEERING IDEAS INTO REALITY



Crop circles
The path from fuzzy logic to precision agriculture

Flying solo
How drones are taking the danger out of rail jobs

Beyond 3D
Young innovator turns the formwork world on its head

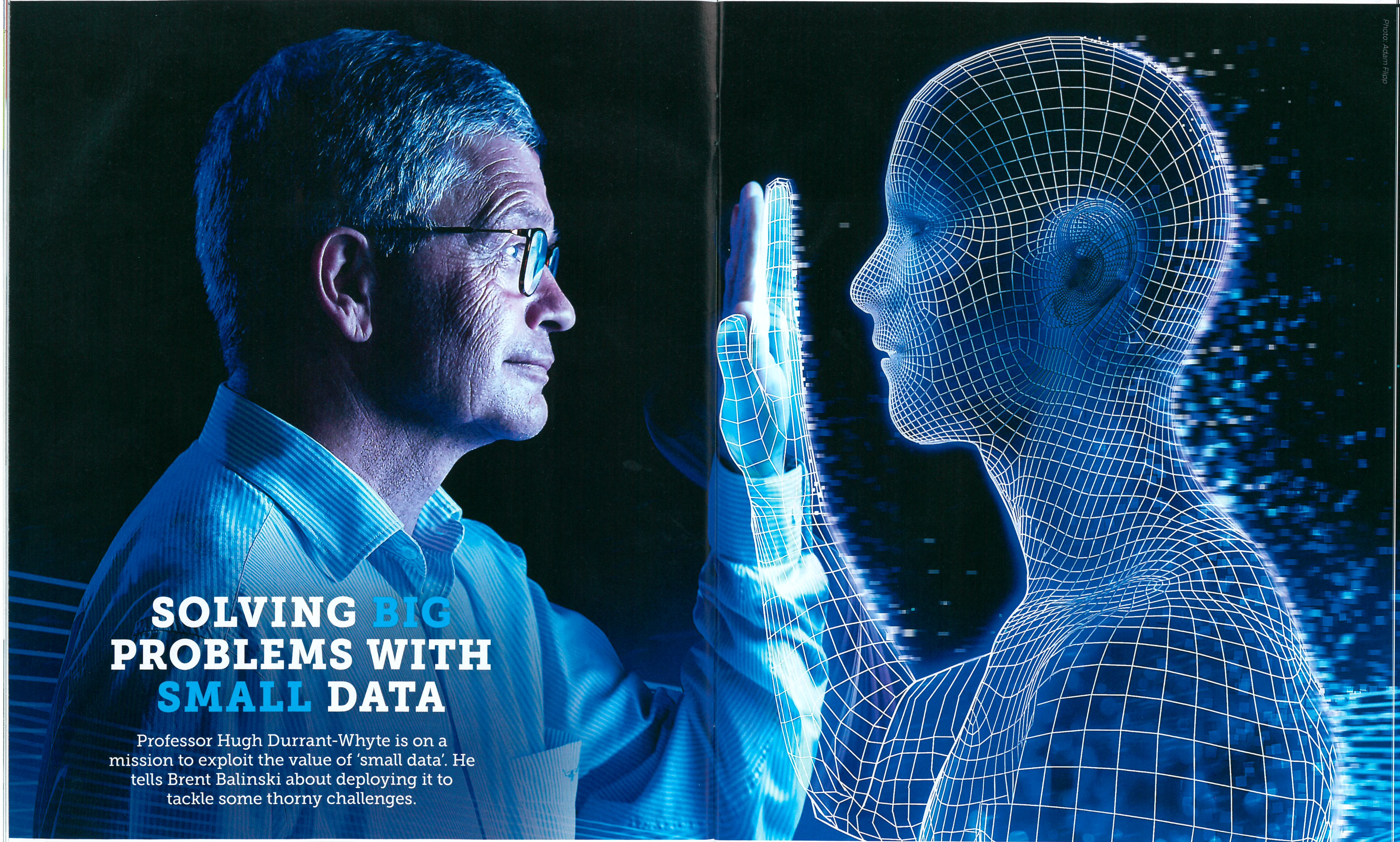


Photo: Adam Filipp

SOLVING BIG PROBLEMS WITH SMALL DATA

Professor Hugh Durrant-Whyte is on a mission to exploit the value of 'small data'. He tells Brent Balinski about deploying it to tackle some thorny challenges.

The explosion of digital data in recent times has been dramatic; according to IBM, 90 per cent of it was created in the last two years. It's the currency of the digital age, some say, and you see the results of data science all the time, such as in Netflix recommendations, targeted ads, and credit card transactions flagged as suspicious.

Companies are seemingly crying out for data scientists to make the best of the "big data" trend. For example, McKinsey Global Institute claimed in 2013 that the US suffered a significant "shortage of the analytical and managerial talent" needed to make the most of it.

The field has grown and evolved significantly, according to Professor Hugh Durrant-Whyte, Director of University of Sydney's Centre for Translational Data Science (CTDS).

"The whole data science field – which has been sort of maturing in the last 20 or 30 years – has got to the point where it can be applied a lot more broadly than just in areas like advertising, where you hear a lot of what's going on," he tells *create*.

Durrant-Whyte's centre (launched last October) is trying to find the "knowledge needles in the data haystack" in a wide array of areas, and involves staff from each of the university's faculties.

Starting with a problem and using machine learning to fossick through massively different data sets is a promising new way to find solutions in situations in a range of areas from the lactation of cows to the way buildings stay up.

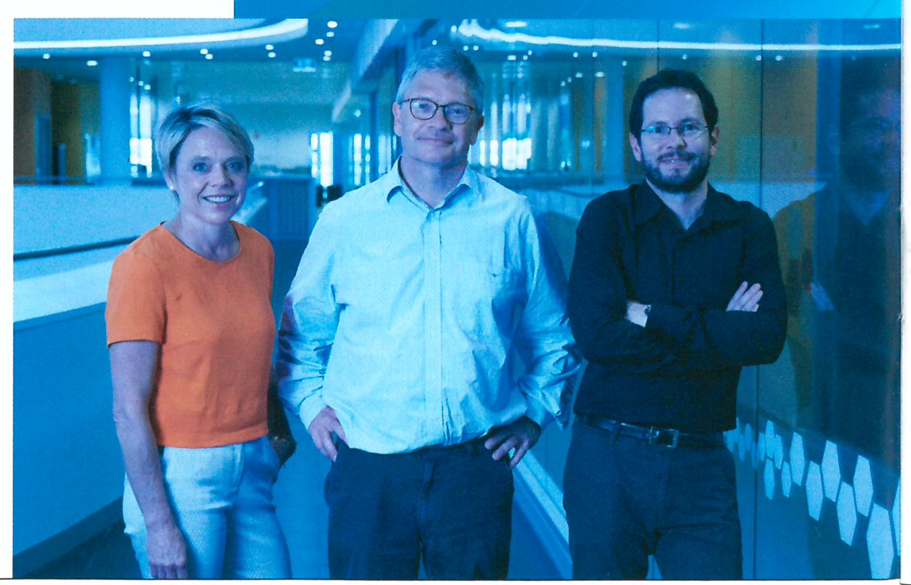
"What we're really trying to do now is use data to fundamentally change the way that people think about a whole range of areas," he adds.

Found in translation

The types of problems Durrant-Whyte is interested in are extraordinarily complex, real world dilemmas, and unable to be explicitly answered.

Things are problem-led in this approach, rather than a typical solution looking for an application.

"It really is just going out, looking for really interesting problems that are quite hard to address with conventional mathematics, which have a lot of uncertainty associated with them, and therefore are quite amenable to the data-science kind of problem space, and using and developing new techniques to address those problems," he explains.



Right: Hugh Durrant-Whyte in the Centre for Translational Data Science. Below: With colleagues Sally Cripps and Fabio Ramos.

"They're not simple like, 'Do I sell you something?' They're quite complex, like, 'What are the manifold ways that someone might develop a mental health problem?' The amount of data you have usually isn't big enough. Therefore, you need to tackle it with completely different techniques."

The above situation is a new one to be tackled in this way, which aims to actually explain why things happen rather than just establish simple correlations.

Instead of applying a model to a situation – for which there could be many possibilities – the approach is to address it differently. Rather than working backwards from a solution, the idea is to go about building a probabilistic solution for a problem with lots of uncertainties built into it.

> "What we're really trying to do now is use data to fundamentally change the way that people think about a whole range of areas."

For mental health, for example, there might be approximately 1,000 different factors impacting the development of abnormal issues, therefore two to the power of 1,000 possible models (a space larger in number than the number of atoms in the universe).

"In the past, what's happened is that people have picked a model and because the space is so big, that model is almost certainly correct," he says.

So what they've chosen to do is turn that process on its head.

"We said actually, of all the possible models, these are the ones that actually best explain the data for this person, and this is why," Durrant-Whyte says.

"That means that you actually have a model that allows you to intervene, that says, 'At school, at age 13, if this person was dealt with in a different way, he would have a much lower propensity to suffer mental health conditions in later age."

The discovery models for such problems could then be applied elsewhere.

The centre also makes a distinction between 'big data' uses of data science, and labels its approach to problems of an uncertain and complex nature as 'small data'.

There might be a lot of information to process, and from many different data sets, but little data to say which model is the correct one.

For these diverse, real-life situations with many variables and little certainty, the results are probabilistic in nature.

"If I was to try and summarise what we do succinctly, it's really to use data and data science methods to build really new discovery models that span right across all areas of human endeavour that really understand the kind of uncertainty and model complexity that you get in real environments."

Buried answers

Two other small data problems Durrant-Whyte has been involved with in recent years involve what lies beneath.

The water pipe failure prediction tool, which he was involved in when he headed National ICT Australia (NICTA, which Durrant-Whyte left in 2014, and which is now part of CSIRO's Data61) helps prioritise the maintenance of water utilities.

The project combined data from 27 utilities worldwide, 9 million pipes (spanning 525,000 kilometres), and 700,000 failure incidences.

According to CSIRO, the tool could save $700 million annually by focussing efforts on preventative work (the estimated current spend on fixing failures is $1.4 billion).

"You basically take a lot of data about how that pipe was built, how it was manufactured, and additionally external information like weather conditions, traffic conditions, things like that, to make a much better model of why and how ⧉

"It really is just going out, looking for really interesting problems which are quite hard to address with conventional mathematics."

NO TEXTING, NO DRIVING

There are two surprising things – given his stature as a technologist – Professor Durrant-Whyte does not own. These are a driver's license and a mobile phone.

A driver's license has been unnecessary in inner Sydney, he says, since it's easy to get around without driving a car.

"I've got four children, they all drive," he says. "My wife drives. Everyone I know drives, and I just cycle and that's fine."

As for the lack of a phone, it helps plans to be stuck to and focus to be kept. It's hard to disagree that people spend a lot of time unnecessarily glued to to their phones nowadays. And not having one can be an enabler and a delegation aid.

"If you look at the way leadership works, people also spend too much of their time micro-directing everyone," Durrant-Whyte says.

"In the past, we used to plan. We said, 'Okay, we're going to do the following things. If plan A doesn't work, we'll go to plan B, and so on... It meant that you could walk away and you've delegated it effectively."

DOCK STARS

AutoStrads are the brainchild of Hugh Durrant-Whyte, who led much of their development at the Australian Centre for Field Robotics at Sydney University. The eight-wheel robots straddle shipping containers and move them around the terminal autonomously, using radar and high-precision GPS. They place containers into position with better than 2 cm accuracy. The AutoStrad technology was later acquired by logistics company Kalmar Global.

CSC 450 SPECS

Capacity:	50 t
Service weight:	70 t
Wheel load:	15 t (fully loaded)
Lifting height:	12,000 mm (front cabin)
Pick-up height:	500 mm
Turning radius (inside):	3550 mm
Turning radius (outside):	9470 mm
Speed:	25 km/hour
Wheels:	eight

things like water mains fail," he explains. "If you think about it, there's no good mathematical model for that because it depends on so many different things, and it's a relatively rare event as well, which makes it even more problematic."

Another project with lots of unknowns to factor in is mapping geothermal resources. According to one estimate by Geoscience Australia, if properly harnessed, one per cent of these resources could supply Australia's energy needs – greenhouse emissions-free – for 26,000 years. The untapped potential is vast, but exploiting these "hot rocks" around 4 kilometres below the surface is a delicate operation (polluting an aquifer is a danger) as well as expensive and difficult.

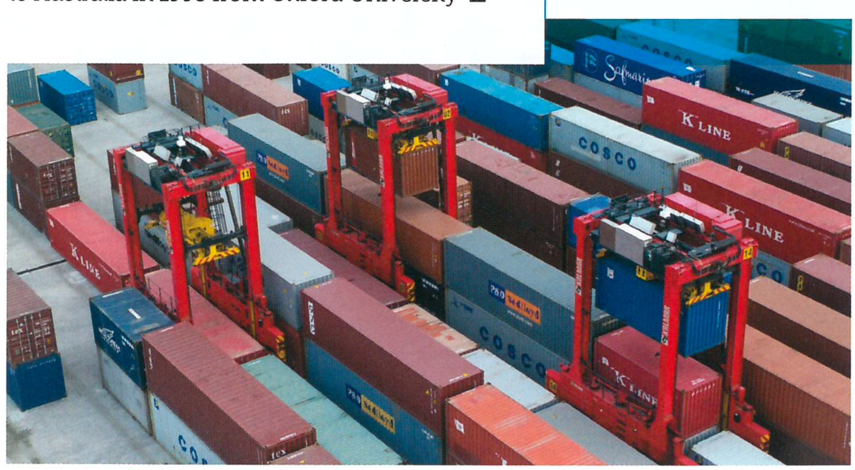
Fusing vast amounts of different data - including seismic, remote sensing imagery and magnetotellurics – and using non-parametric Bayesian algorithms, an idea of what is below the surface can be provided. There's a lot of information, and it could be interpreted in "almost infinite" ways.

"[It] is also an interesting example because the uncertainty in geological information is very high, and as you can imagine, that depth underneath an area the size of Australia, the possible geologies are extraordinarily complex," Durrant-Whyte says.

"Yet, even though we have a lot of data, it still doesn't tell you with any high degree of certainty what the geology actually looks like. So coming up with uncertainty-qualified models for geology really changes the whole nature of, for example, exploration, particularly in areas like geothermal energy, oil, and gas."

Reformed roboticist

Well before these projects, Durrant-Whyte came to Australia in 1995 from Oxford University ⊠



to take a mechatronics professorship at Sydney, lured by the prospect of continuing his work in field robotics (a sub-discipline he is credited with pioneering) in a more favourable climate. Australia was also, he believed, the ideal place for a field roboticist to make his mark – wide open, sparsely-populated, with many opportunities to automate.

His work as a roboticist is renowned, and has earned him awards including the 2010 NSW Scientist of The Year. His first industry project here resulted in the robotisation of Brisbane and then Port Botany (featuring giant AutoStrad machines transporting shipping containers).

This was followed by helping Rio Tinto automate pit to port at its Pilbara operations, including a fleet of self-driving trains and trucks.

"I often pointed out to the Americans that we have more autonomous vehicles in Australia working and operating than they do in California," he offers.

"It's just that they're in mines and on container terminals, and so on."

After leaving University of Sydney's Australian Centre for Field Robotics (which he founded in 1999) in 2010, he vowed to give up robotics, explaining it in terms of a mission accomplished.

"The first thing is, I think you always have to get to a point and decide to move on and do something else," he says. "Do I regret it? Yes, the nice thing about robotics is when you're finished and done, you can video it and it moves and makes nice pictures."

He will return to the UK this year as Chief Scientific Advisor for their Ministry of Defence.

The jobs of the future

Given his work impacting logistics and mining, it is no surprise that Durrant-Whyte is regularly asked to comment on how well-placed the country's workforce is for technological disruption.

A CEDA paper he was lead author on in 2015 used machine learning to determine the susceptibility of Australian jobs to automation in 10 to 15 years. It found 40 per cent of current jobs had a high probability of being lost to computerisation and automation in the period.

The modelling attracted significant media coverage and the paper was one of the most popular Durrant-Whyte has written.

Where the media attention was directed has been unfortunate, he believes. He predicts the lost jobs will be replaced, but at the higher and lower socioeconomic ends and not in between.

"The problem is that there won't be good jobs. That's the worry. Okay? Those middle ones will have gone," he says. "What's happening is, the jobs are being polarised, and that is very worrying. That is very worrying indeed."

There is also the concern that students are still signing up at universities for those disappearing jobs. The topic of the future returns us to the scholar's post-robotics passion, and an area he believes will become essential for just about every profession. Nowadays it should be a requirement for every student to study data science as part of their degree, he contends.

"If you think of every endeavour that this university does, from medicine to law to arts to engineering, essentially all of those areas are now data areas," he says. "If you're in the arts there are even machine-learning algorithms for understanding paintings and composition, and things like that. Every area now is in some sense underpinned by this notion of data, and this notion that data and evidence show you things."

"The nice thing about robotics is when you're finished and done, you can video it and it moves and makes nice pictures."

ARTEMIS

The Artemis supercomputer was launched in June 2015 and upgraded last year with its power tripled. It is part of the Sydney Informatics Hub capability.

It is used for disciplines including computational chemistry, environmental science, fluid dynamics, materials science, and mechanical and structural engineering. A lot of the machine learning done at the CTDS doesn't require a supercomputer.

Problems are often divided into "thousands, if not millions of parts" and processed in parallel. Artemis can do this.

Durrant-Whyte adds that a lot of the algorithms developed at his centre can be uploaded to a public cloud like Amazon or Azure and processed there.

"It's interesting that a lot of these new techniques are all being developed with that kind of publicly available, massively parallel cloud computing in mind," he says.

"I think increasingly it will head that way, although Artemis is necessary because sometimes you need to be very close to the data."

Compute, storage and management nodes connected by high-performance, low-latency interconnect based on Mellanox InfiniBand (IB), which is a proprietary networking nfrastructure.

There is also a 10Gbps Ethernet management network interconnecting the compute, login and management nodes to facilitate installation of the compute nodes and PBSpro batch job management.

ARTEMIS SPECS

Cores:
4,264 in total

Standard compute nodes (Haswell):
56 (24 cores per node)

Standard compute nodes (Broadwell):
80 (32 cores per node)

High memory compute nodes (512 GB):
2 (24 cores per node)

Very high memory compute nodes (6 TB):
3 (64 cores per node)

GPU compute nodes:
5 (24 cores per node)

Images courtesy of the University of Sydney and Digital Realty

Asked about the most impactful development in AI in the last five to ten years, his response is again heavy on the rise of data, and the use of this to build predictive models that can make decisions automatically. It follows a paradigm shift over a longer period away from model-based towards data-based AI.

"And a big one that we've all learned in robotics, as well as AI, is data is the way you learn; I think that's been critical," he adds. "What will be the biggest impact in the next five or 10 years? That's interesting. I have a feeling that the biggest impacts will probably be in areas like social science and health, and possibly even engineering systems, rather than in anymore Google, or Facebook, or Twitter type of AI."

He predicts that data and machine learning will create disruptions in this field that haven't yet been properly thought about.

"I think in engineering, particularly, engineers have not really quite understood that yet. By that I mean, the traditional mechanical and civil engineers - a lot of the data they generate is also subject to machine learning and an ability to learn new models from data. I think that will be a profound shift in the engineering profession, as a whole." ●