

# The future of data science looks spectacular

July 21 2015, by Glenn Wightwick

---



The author (right) and Toby Grime, Artistic Director Animal Logic (left), inside UTS's new Data Arena. Credit: UTS

It wasn't that long ago that we lived in an entirely analogue world. From telephones to televisions and books to binders, digital technology was largely relegated to the laboratory.

But during the 1960s, computing had started to make its way into the back offices of larger organisations, performing functions like accounting, payroll and stock management. Yet, the vast majority of systems at that time (such as the [healthcare system](#), electricity grids or

transport networks) and the technology we interacted with were still analogue.

Roll forward a generation, and today our world is highly digital. Ones and zeroes pervade our lives. Computing has invaded almost every aspect of human endeavour, from [health care](#) and manufacturing, to telecommunications, sport, entertainment and the media.

Take smartphones, which have been around for less than a decade, and consider how many separate analogue things they have replaced: a street directory, cassette player, notebook, address book, newspaper, camera, video camera, postcards, compass, diary, dictaphone, pager, phone and even a spirit level!

Underpinning this, of course, has been the explosion of the internet. In addition to the use of the internet by humans, we are seeing an even more pervasive use for connecting all manner of devices, machines and systems together – the so-called Internet of Things (or the "Industrial Internet" or "Internet-of-Everything").

## **Complex systems**

We now live in an era where most systems have been instrumented and produce very large volumes of digital data. The analysis of this data can provide insights into these systems in ways that were never possible in an analogue world.

Data science is bringing together fields such as statistics, machine learning, analytics and visualisation to provide a rigorous foundation for this field. And it is doing this in the same way that computer science emerged in the 1950s to underpin computing.

In the past, we have successfully developed complex mathematical

models to explain and predict physical phenomena. For example, we can accurately predict the strength of a bridge, or the interaction of chemical molecules.

Then there's the weather, which is notoriously difficult to forecast. Yet, based on numerical [weather prediction models](#) and large volumes of observational data along with powerful computers, we have improved forecast accuracy to the point where a five-day forecast today is as reliable as a two-day forecast was 20 years ago.

But there are many problems where the underlying models are not easy to define. There isn't a set of mathematical equations that characterise the health care system or patterns of cybercrime.

What we do have, though, is increasing volumes of data collected from myriad sources. The challenge is that this data is often in many forms, from many sources, at different scales and contains errors and uncertainty.

So rather than trying to develop deterministic models, as we did for bridges or chemical interactions, we can develop data-driven models. These models integrate data from all the various sources and can take into account the errors and uncertainty in the data. We can test these models against specific hypothesis and refine them.

It is also critical that we look at these models and the data that underpins them.

## **360 degree data**

At my university, we have built a Data Arena to enable the exploration and visualisation of data. The facility leverages open-source software, high-performance computing and techniques from movie visual effects

to map streams of data into a fully immersive 3D stereo video system that projects 24 million pixels onto a four metre high and ten metre diameter cylindrical screen.

Standing in the middle of this facility and interacting with data in real-time is a powerful experience. Already we have built pipelines to ingest data from high-resolution optical microscopes and helped our researchers gain insight into how bacteria travel across surfaces.

We read 22 million points of data collected by a CSIRO Zebedee which had scanned the Wombeyan Caves, and ten minutes later we were flying through the cave in 3D and exploring underground.

No matter what sort of data we have been exploring, we have inevitably discovered something interesting.

In a couple of cases, it has been immediately obvious we have errors in the data. In an astronomical dataset, we discovered we had a massive number of duplicate data points. In other situations, we have observed patterns that hadn't been evident to domain experts who had been analysing the data.

This phenomenon is the classic "unknown unknown" (made famous in 2002 by US Secretary of Defence Donald Rumsfeld) and highlights the power of the human visual system to spot patterns or anomalies.

Today's world is drenched in data. It is opening up new possibilities and new avenues of research and understanding. But we need tools that can manage such staggering volumes of data if we're to put it all to good use. Our eyes are one such tool, but even they need help from spaces such as that provided by Data Arena.

*This story is published courtesy of [The Conversation](#) (under Creative*

*Commons-Attribution/No derivatives).*

Source: The Conversation

Citation: The future of data science looks spectacular (2015, July 21) retrieved 25 April 2024 from <https://phys.org/news/2015-07-future-science-spectacular.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.