

DATA LEADERS WHO'S WHO

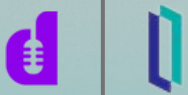
DRIVING INNOVATION WITH DATA

Featured in this week's interview

Ian Oppermann

Chief Data Scientist

NSW Department of Customer Service



TECHNOLOGY

Editors note

Data has become the core of how we operate as a society, and is central to business success. Data-driven businesses are finding new ways to compete by letting their data uncover unique points of differentiation. The value of data can't be overstated; it has been referred to as the new currency. We are proud to bring you the latest Data Leader's Who's Who and support the increasingly critical role that these emerging and established data leaders play both now and into the future.

The insights these leaders provide will be pivotal in strengthening the position of data in the business community and we thank them for sharing their valuable experiences.

Andrew Aho, Regional Director Data Platforms, InterSystems | Felipe Flores, Founder, Data Futurology

PRINCIPLES BASED DATA ENGAGEMENT



Ian Oppermann

Chief Data Scientist

NSW Department of Customer Service

Dr. Ian Oppermann is a renowned expert in the digital economy and has extensive knowledge and experience in big data, broadband services, and technology's impact on society. In this article, he leverages his over 30 years of expertise to outline the key components of a successful data strategy. He believes that data can be difficult to harness because of its value and the lack of standardised methods for sharing and using it. A strong data strategy should embrace the limitless potential of data while establishing clear guidelines for data sharing and usage through a limited set of data handling protocols. Oppermann also shares on his contributions to two exciting government projects; the Out of Home Care reform and the creation of the NSW AI Assurance Framework.

STRATEGY & LEADERSHIP

How do you devise a data strategy? What sets apart the good from the bad?

Data is an elusive commodity. Partly because of the utility of data and partly because we do not really (yet) have good general data sharing and use frameworks. Some really significant complications exist when we think about the general use of data:

- Every data set is unique – it has its own history, its own chain of custody, its own provenance, its own data quality.
- Each data set can potentially be used in an infinite number of ways. Whether a dataset is fit for the purpose it is about to be used for, is currently a very subjective consideration. We lack general frameworks for data quality and for general evaluation of “fitness for purpose”.
- Every product created from data can also potentially be used in an infinite number of ways. We lack general frameworks with recommendations or restrictions on how data products should be appropriately used.

- Context matters: context changes how a dataset or a data product can be safely and appropriately used or governed.
- Every dataset of data product can have a very long and complex life. Datasets or data products can be combined with other data sets / data products. This could change the level of sensitivity of a dataset / data product or the level of personal information in the dataset/ data product. Of course, these recombined outputs can themselves be recombined on ‘ad infinitum’. So, we are challenged by the “next” use of data or the “next next” when thinking about how to appropriately share and use data. This is particularly true if data “escapes” from its intended governance framework and gets into the wild. This leads a lot of data custodians to simply lock it down and not allow future, secondary uses of data and data products.



A good data strategy tries to acknowledge these multiple areas of “infinite” possibilities, “infinite” complexity, or “infinite” uniqueness and bring a finite number of data handling recipes into a data environment to ensure appropriate controls are placed on data sharing and use. These controls need to cover technical, domain and governance capability of the people involved, roles and responsibilities, and guidance / restrictions on use at each stage of the data lifecycle. The levels of control could range from “very high control” to intermediate levels of control, even to “no control” or open data environments.

A good data strategy also identifies the metadata required to record the journey of data and its use, support discoverability, support data quality assessment and support future uses of data.

A good data strategy also speaks to what happens when things go wrong, when data or products escape from their intended environments of control, and the plans for redress.

Finally, a good data strategy is based on standards. There are many important Australian and international standards and the very worst thing anyone can do is reinvent the wheel. If Australian or international standards do not exist, reach out across your organisation and seek consensus on data sharing frameworks, consistent minimum levels of metadata.



"A GOOD DATA STRATEGY ALSO SPEAKS TO WHAT HAPPENS WHEN THINGS GO WRONG, WHEN DATA OR PRODUCTS ESCAPE FROM THEIR INTENDED ENVIRONMENTS OF CONTROL"

What are the essential qualities of a data leader?

I believe the role of a “data leader” has changed over time. Not so long ago, data leaders were engineering, or science types as reflected in titles like Data Scientist or Data Engineer. While the domain discipline is still needed, what is increasingly needed is to be a storyteller, with an “outcomes” focus, being able to paint a vivid picture for non-specialists, helping people acknowledge and cope with the complexity of wicked challenges to be addressed by data, as well as being able to dive into the detail and the complexity if needed. Just being skilled in the handling of data is far from sufficient.

How have you found success in raising data literacy in your organisation? How do you get involved in educating peers and execs?

People engage with principles far more readily than technical frameworks, so it is always best to start any conversation, framework development, or skill development with principles. Principles by themselves however are not sufficient for data sharing and use. It is essential to ultimately connect “the principles to the bits”. This means finding contextually relevant ways to connect the “agreed-to-in-principle” guidance, to formal policies with more specific requirements, and then connect these to frameworks which cover all possible envisages scenarios and contexts, and then finally to the bits in the datasets themselves with their inherent data quality, privacy and sensitivities characteristics. This is hard given all the particularities of any given dataset, the ranges of context of data use, the sensitivities of the data and associated data products, and how well controlled the data use environment is.

Without a principles-based engagement however, it is near impossible to engage people sufficiently to the point where they are willing to make the investment of time and effort to get into specific frameworks for data use in the context it operates in.

Specifically back to data literacy, it is important for people to understand data sharing and use that principles level, but helping people on the journey of mastering the major elements of the data lifecycle, appropriate controls, considerations for use, and guidance / restrictions / prohibitions on use of data products generated, are all stages of mapping the "principles to the bits".

What work are you most proud of?

It has certainly been an interesting ride over the past few years.

The most significant work done whilst I was leading the Data Analytics Centre (2015-2019) was the work to help with the reform of Out of Home Care (OOHC) which is a service for children identified as being at risk of significant harm. The underpinning data set, known as the Human Services Data Set (HSDS) brings together data collected by individual government agencies to take a unique and powerful view of service usage and effectiveness to improve outcomes for children. The separate data are de-identified and linked by a specialist data linkage centre, meaning all records are anonymous. The data set is unprecedented in scale in NSW, bringing together 27 years of data, over seven million records, from over 60 frontline data sets in 11 government agencies.

The NSW AI Assurance Framework is probably the second most impactful piece of work. This framework really is an attempt to connect the principles of the NSW AI Strategy and the NSW AI Ethics policy to the specific data and algorithms for literally any NSW government use of AI

The framework was developed during 2021, tested against many real-world use cases, and finally endorsed by NSW Cabinet in late 2021. It is mandatory to apply this Assurance Framework in NSW and the NSW AI Review Committee has been working with NSW government agencies providing guidance and feedback on existing and planned AI projects.

Finally, the work with the Australian Computer Society on privacy preserving data sharing frameworks, and with international standards on data usage standards has been slow going.

However it has produced some very useful resources for every one involved in data sharing and use including for the purposes of AI.

What do you wish senior leadership knew or understood?

This is the definition of a career limiting question!

The one thing I think it is important for senior leadership to understand is that data sharing and use, in the general sense, is not a solved problem. Every single use case has specific requirements, every dataset is unique, every change in context as to how data is used changes the controls required, and every data product can live a life of its own. Understanding these complexities, means that we can develop frameworks of control which allow us to appropriately use and share data. Ignoring these complexities leads to either "closed" data mentality, or potential risks from not having the right controls in place.



"IGNORING THESE COMPLEXITIES LEADS TO EITHER "CLOSED" DATA MENTALITY, OR POTENTIAL RISKS FROM NOT HAVING THE RIGHT CONTROLS IN PLACE. "



AI PROJECTS, PRIORITISATION AND PRODUCTIONISING AI

What have been major AI watershed moments in your career – or alternatively in the industry?

AI has a long history. I recall taking a special AI course at university back in the late 80's. I also recall the frustration at being beaten by an algorithm playing chess during that same period. It has come a long way since then. A couple of major call outs of course include IBM's Watson winning at the game Jeopardy in 2011 and then Google's AI beating a grand master at Go in 2016. What has really impressed me recently was in 2021 when Beethoven's unfinished Tenth Symphony was completed by AI, and then in 2022 when AI accurately predicted gene expressions in a type of yeast. That last neural network predicted the effectiveness and evolution of gene promoter sequences, something which until then was an amazingly complex computational task.

AI capability continues to accelerate, and we are now being forced to seriously consider the implications of what AI can be made to do. Moral and ethical challenges have been debated for some time with regard to the use of AI, but now we also need

to consider such issues such as if AI can "own" an invention or a patent, if the style of a human creation should be protected, and even if use of AI should be prevented in certain domains.

What are some of the lessons learned you've encountered when getting AI products into production?

Many lessons learned have been captured in the principles of the NSW AI Assurance Framework which is now mandatory to apply for all AI projects within the NSW government. The most important is the need to consider possible harms during the design phase of an AI project. Harms which relate to the intended recipient of the AI service are the more obvious areas of consideration, but we also ask Agencies to consider harms associated with others who are not directly engaging with the AI system, and harms associated with repeated application / engagement with the AI system. A second is the need to pilot and review before scaling any AI application. This review is increasingly independent if the potential harms are anything more than insignificant. A third principle is the need to re-assess the performance of the AI system at regular intervals.

A final one is to consider AI systems as a simply a "use" of data. If we are uncomfortable about what an AI system is doing

(or proposed to do), we assess the system without the AI piece, then without the data piece. If we are still uncomfortable, then it is what is being done rather than how it is being done. This not only enables us to separately consider the data and the AI issues and what we need to do to ensure appropriate safeguards, but also allows us to consider if the use of data and AI is really just going to amplify a flawed process.

These principles are all intended to avoid the challenges of AI being used to scale up existing processes, or to identify as early as possible any problems with the deployment of a new AI driven service.

THE FUTURE OF DATA & ANALYTICS

What new technology and innovations do you see as being the most critical to the industry over the next 18 months?

In my opinion, the most exciting technologies currently are AI and quantum.

AI has reached the stage where it is really becoming useful at a general level. Anyone who has experimented with ChatGPT will know just how far readily accessible AI has become. I was very tempted and asked ChatGPT to write a speech in the style of me. It "knew" who I was and wrote something that I could believe was written by me if I had not known better.

Whilst ChatGPT has still got a way to go, AI is increasingly useful and increasingly being embedded everywhere. The future of just about any complex system from energy to telecommunications, from logistics to warfare will see AI used in every aspect of those systems.

Quantum is also pretty exciting. The point of “quantum supremacy” where a quantum computer exceeds the capability of a conventional silicon computer is getting very close. Quantum enabled devices will change what we can compute, it will supercharge AI, but also make us rethink a number of fundamentals in the digital world including digital security.

How do you ensure you are leveraging new tech for innovation, rather than tech for tech's sake?

Many interesting projects go nowhere because the outcomes frameworks have not been defined. We used to ask the question “what does good look like?” however that is always subjective. We can do better and put in place “outcomes frameworks” which describe a number of interconnected real-world outcomes and identify quantified measures of indicators for each outcome. This then allows us to get much more specific as to what we mean by “good”. For example, the NSW Human Services Outcomes Framework (HSOF) considers many of the main goals of human services Agencies such as health, education and safety. Underpinning each outcome domain are a series of indicators and it is understood that these outcome domains will interact with each other as policies or services are redesigned to improve individual outcomes. We can then consider and debate much more openly what the appropriate level of those different outcomes are, and the

significance of the interactions with outcomes.

Developing the outcomes frameworks takes a great deal of care, consideration and consensus even to create high level outcomes. It is worth the effort however as it lifts the conversation as to what is trying to be achieved to a letter that genuinely speaks to higher level goals. It can also free people from self-imposed limitations of what they believe can be achieved based on what has been done in the past.

What legacy do you hope to leave behind you at your organisation?

The overarching goal of the last 7 or so years has been to help government think differently about data sharing and use, tackling the “unable”, “unwilling” or “not allowed” mentality that can prevail when seeking to use data. This has led to some really exciting work exploring the art-of-the-possible, as well as really fundamental work developing frameworks, measures, indicators and tools.

Just helping to identify that there is a world between “open” and “closed” data is a good start, but what I really want to enable is helping people understanding their data from a sensitivity and privacy perspective, and then using “recipes” or “playbooks” for appropriate data sharing and use. The NSW AI Assurance Framework, the work with the Australian Computer Society and the work with Standards are all elements of those “recipes” and “playbooks”.

***"THE FUTURE OF JUST ABOUT ANY
COMPLEX SYSTEM FROM ENERGY TO
TELECOMMUNICATIONS, FROM LOGISTICS
TO WARFARE WILL SEE AI USED IN EVERY
ASPECT OF THOSE SYSTEMS."***

