Data driven approaches to customer service: What we need to consider

November 2019

Dr. Ian Oppermann

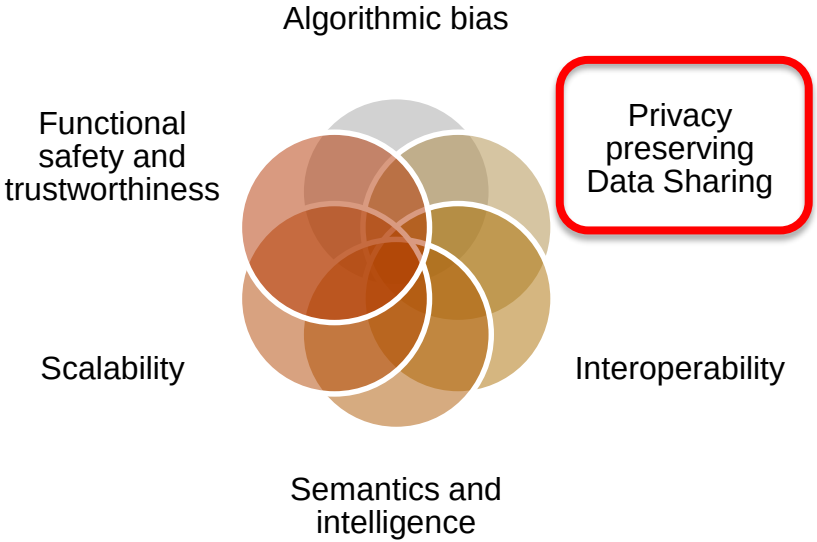
NSW Government Chief Data Scientist,
CEO, NSW Data Analytics Centre

Australia 2050

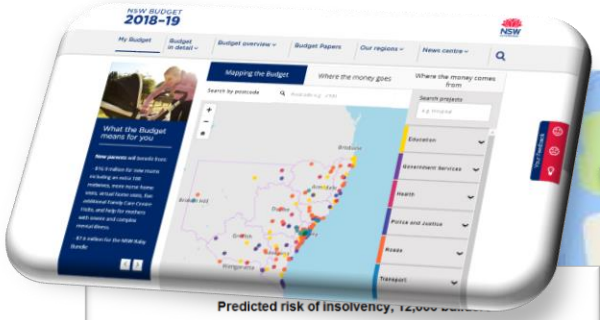
Population : XXm?
Average Age : increase
Retirees to Working age : increase
Urbanisation : increase
Unemployment : mixed
Megacities : increase
Dominant Industries : Services
Climate: Warmer and Changed

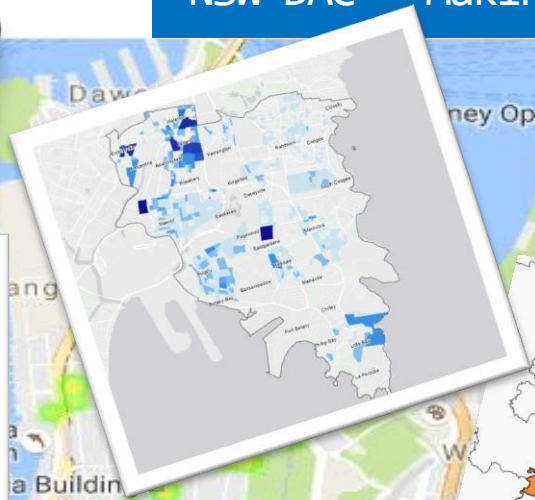
*.. **artificial intelligence** will become a core technology across many different industries and one of the driving forces of the coming fourth industrial revolution, the standardization community will play a critical role in shaping its future.*

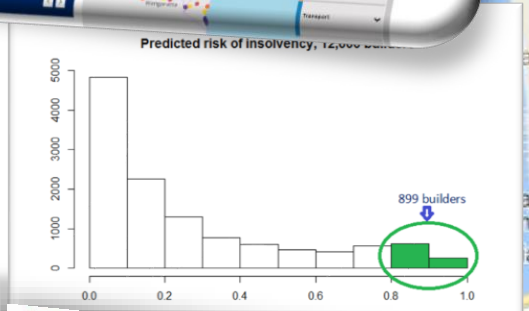
https://basecamp.iec.ch/downloads/white-paper/

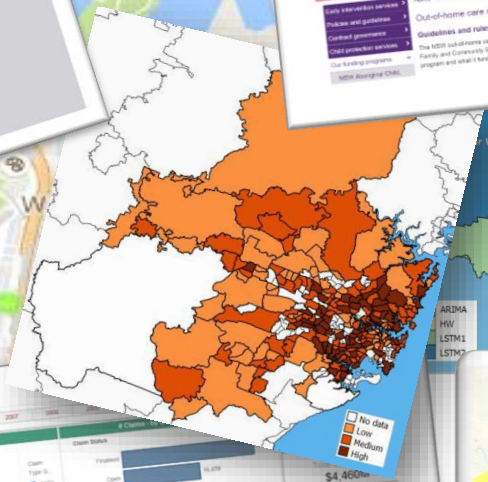


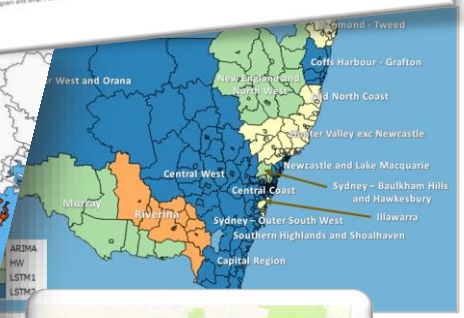
NSW DAC - Making a Difference in the Real World

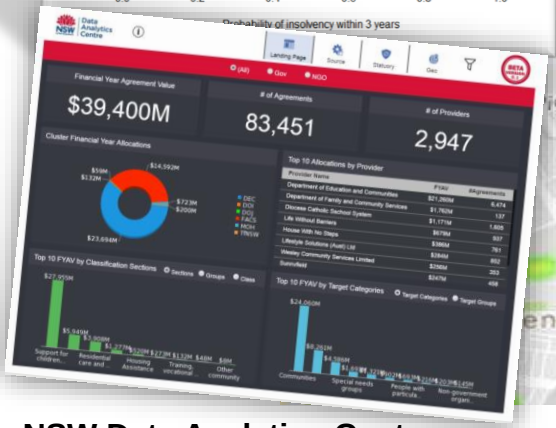








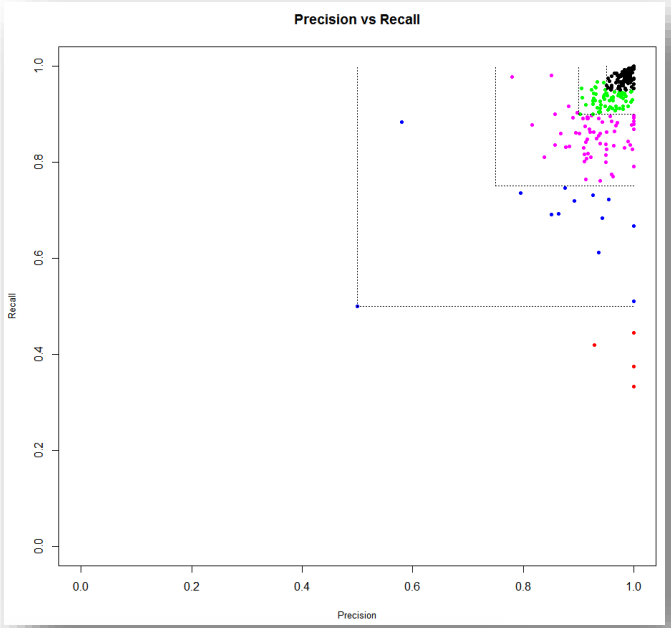




NSW Data Analytics Centre

Procurement Spend Categorisation

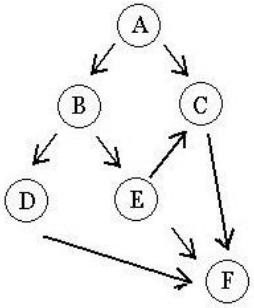
- NSW spends approximately $30bn per annum through procurement.
- Categorisation is currently performed using a rules based system “Spend Cube” with more than 1million rules
- DAC Natural Language Processing Machine Learning tool was trained and then categorised 10,000,000 transactions from the test set into 274 categories.
- Overall accuracy of >96% and runs in hours (c.f. weeks for spend cube)
- Experiments were run without the “Other” category forcing categorisation of almost 300,000 items
- Microsoft announced as commercialisation partner for tool



The following table shows random sample of 200 rows (out of total 200,524 rows) where **VENDOR_BASED_MAPPING is “Other” but Predicted Category is not “Other”:**

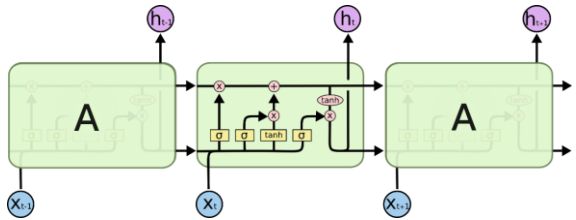
SUBCUBE_NAME	TRANSACTION_ID	VENDOR_NAME	GL_ACCOUNT_NAME	DESCRIPTION	VENDOR_BASED_MAPPING	Predicted.Category.1
JUSTICE	9762835	Wine Country Motor Inn	Accommodation and Meals Actual	4 x Accommodation = $140,	Other	Accommodation
JUSTICE	9780362	Hotel Ibis Thornleigh	Accommodation and Meals Actual	Acc/Meals - 20/11-25/11	Other	Accommodation
JUSTICE	9785418	Riverview Boutique Mot	Accommodation and Meals Actual	accommodation	Other	Accommodation
JUSTICE	9748584	Bathurst Heritage Motor Inn	Accommodation and Meals Actual	accommodation = $135,	Other	Accommodation
JUSTICE	9777357	Bells Motel	Accommodation and Meals Actual	Accommodation for Edan Mumford at	Other	Accommodation

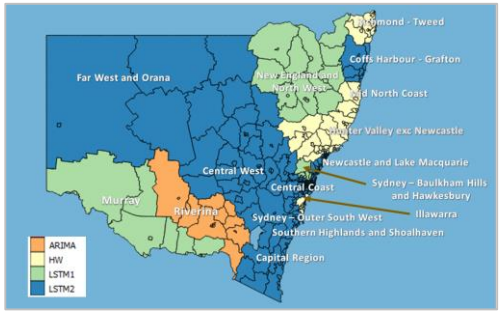
Predicting school age enrolment



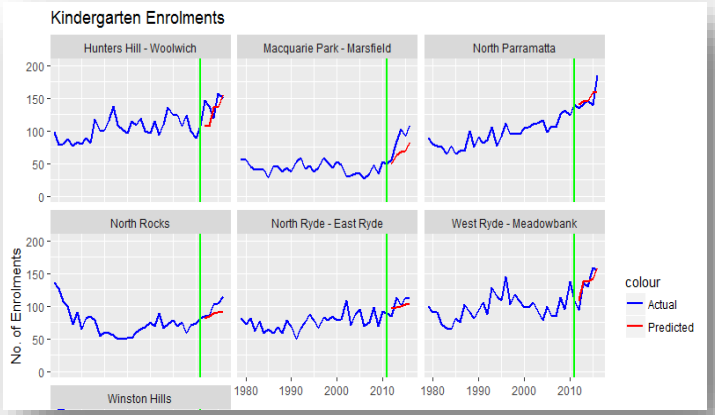
Data Exploration
Feature Engineering

Model Build
And Validation





Model
Deployment

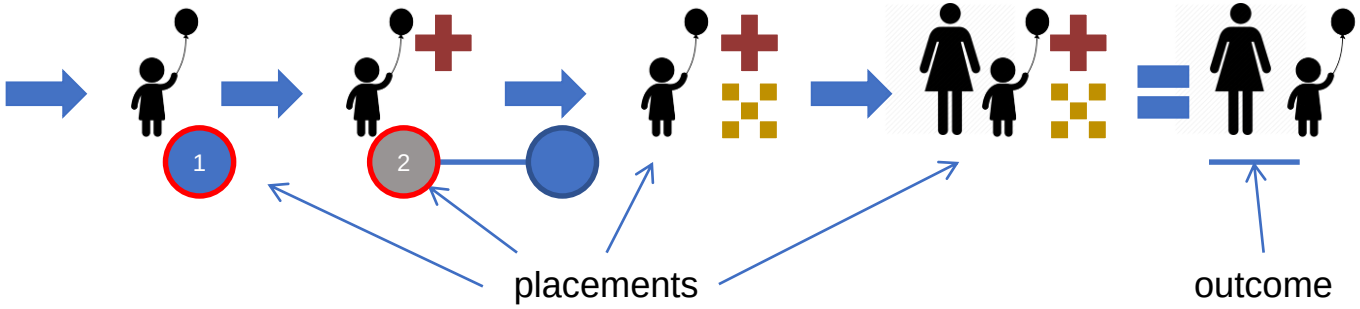


Role of the DAC

Multi-agency Data integration
Spatial and temporal visualisations
Deep Learning model evaluations

Supporting OOHC Reform

- Create “pathways” of all children in OOHC and identify cohorts of particular interest
- Represent OOHC as a sequence of placement events
- Each sequence has a final placement or exit
- At each placement the child accumulates service history
- Millions of records
- Health, FACS, Justice, Education, Industry, Transport



Data Sharing Strategy

… supporting automated data sharing across government

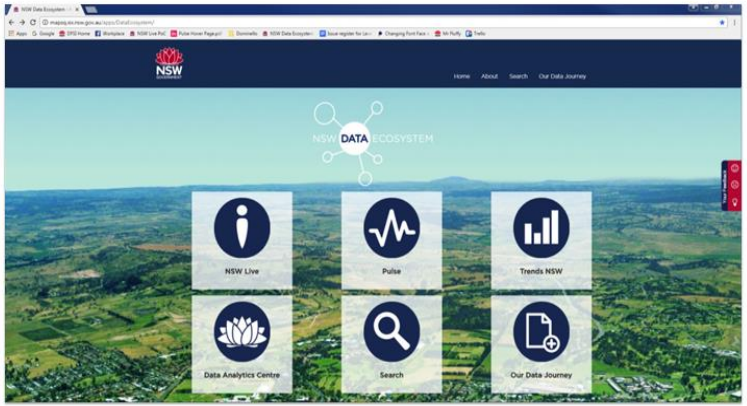
NSW Data Ecosystem

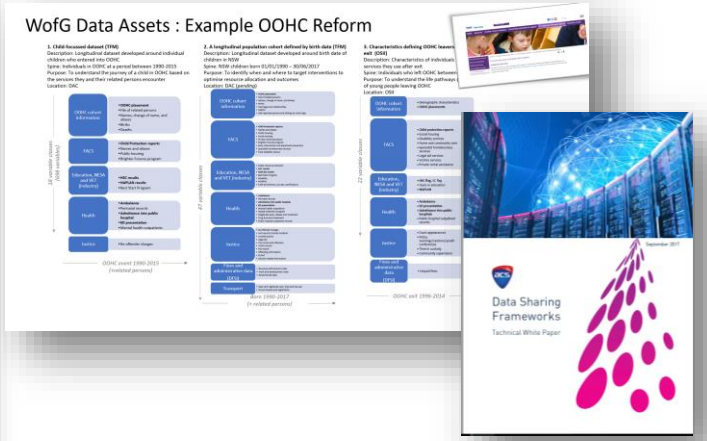
digital nsw
DESIGNING OUR DIGITAL FUTURE

Key part of our digital and data agenda is the development of the NSW Data Ecosystem

It will provide government and citizens with the capability to unleash the full potential of our data

Make it easier and faster for people to turn data into insight and impact





Role of the DAC

Working with specialist organizations (including ACS, CSIRO), state and federal governments, industry and privacy advocates to develop frameworks, trial technical solutions, inform policy and help scale automated data sharing

NSW Government Agencies

NGO's

Companies

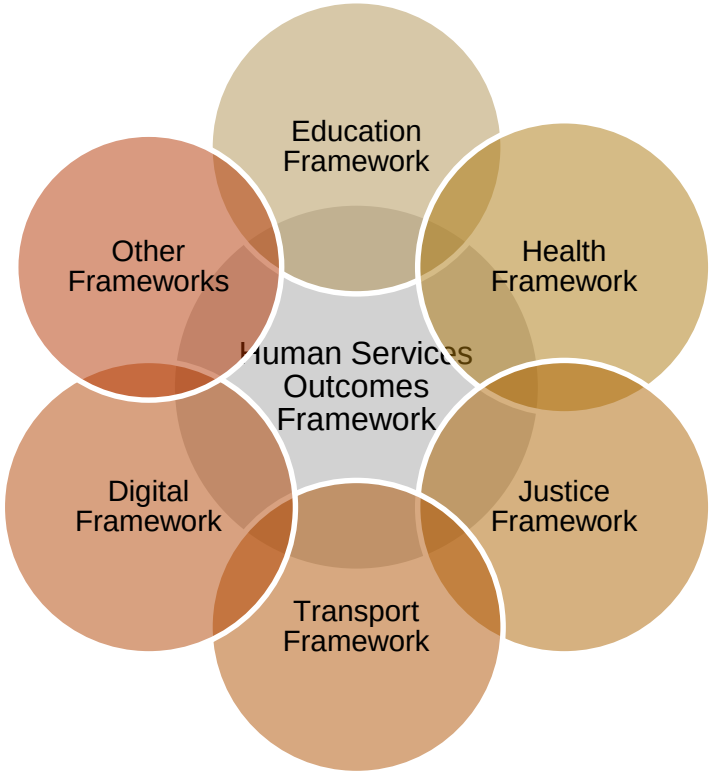
Other Government Agencies

General Public

NSW Data Analytics Centre

Data Analytics Strategy for Outcomes Based Budgeting

… Supporting quantification of Outcomes Frameworks

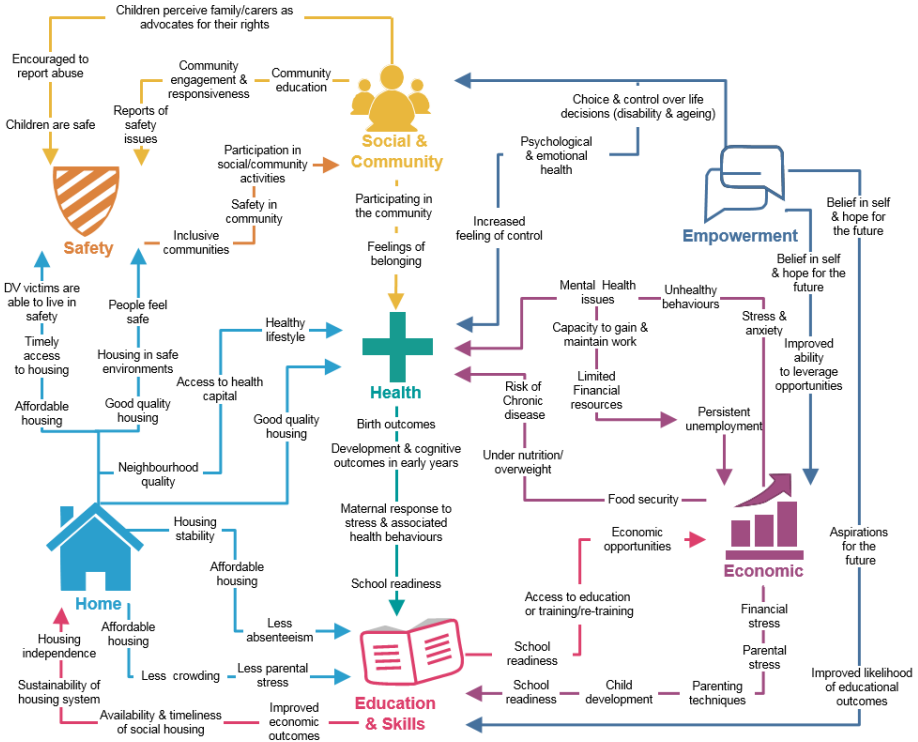


Role of the DAC

Build or enrich the datasets to support quantification of Outcomes Frameworks.
Help agencies identify cohorts and analyse current outcomes.

Data Analytics Strategy for Outcomes Based Budgeting

… Supporting mapping in data of evidence pathways



Outcomes Framework

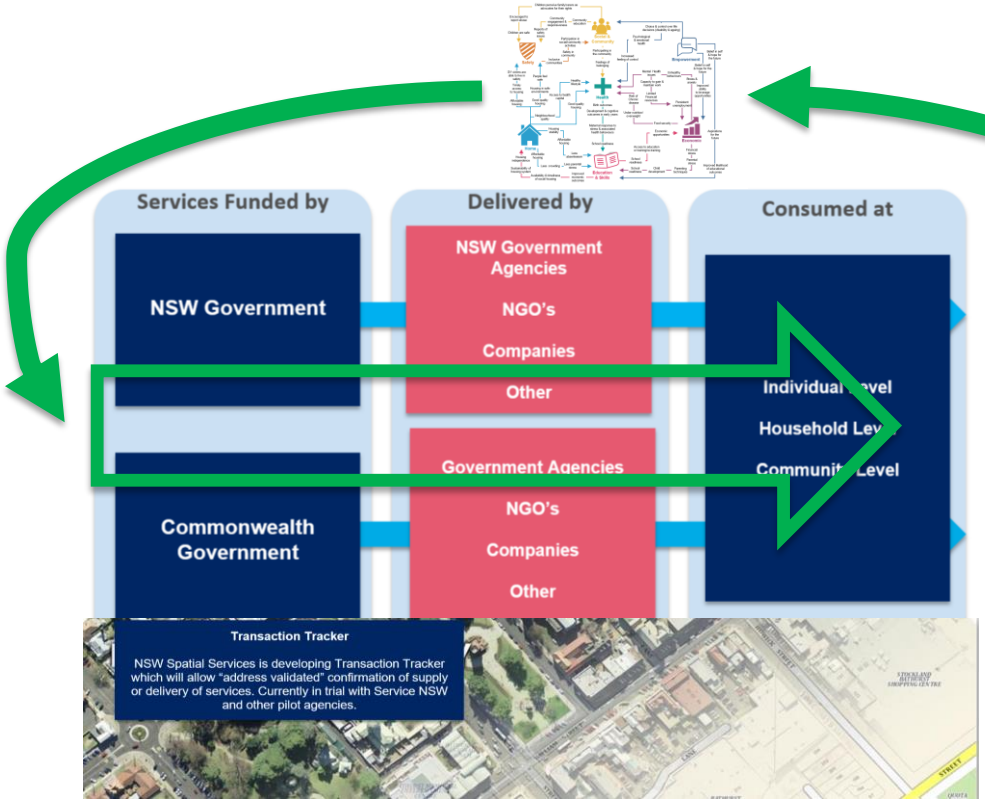
Role of the DAC

Build the dataset to support development or enrichment of evidence pathways such as the work done for out-of-home care (OOHC) services.

Help agencies identify cohorts and analyse pathways.

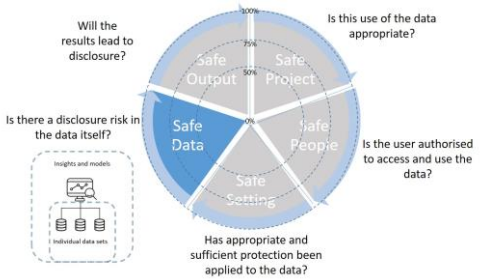
Data Analytics Strategy for Outcomes

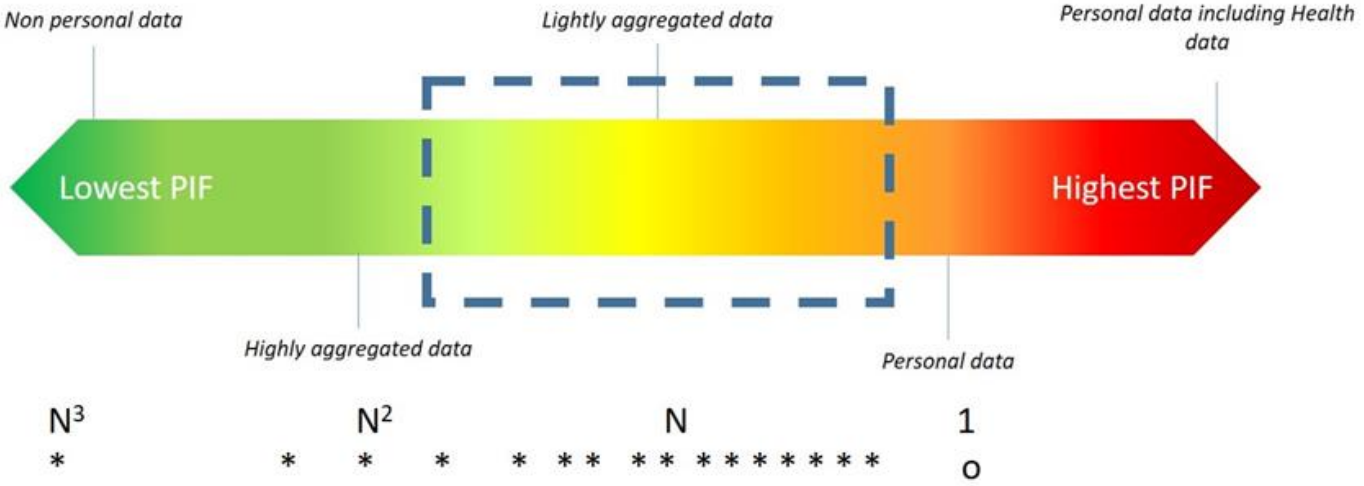
services mapped to place, contributing to outcomes frameworks, analysed against evidence pathways to inform policy



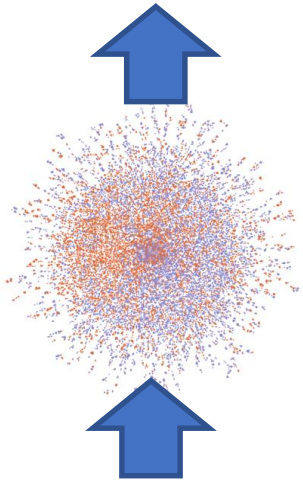
Outcomes Framework

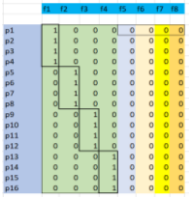
Safe Data: Personal Information Factor

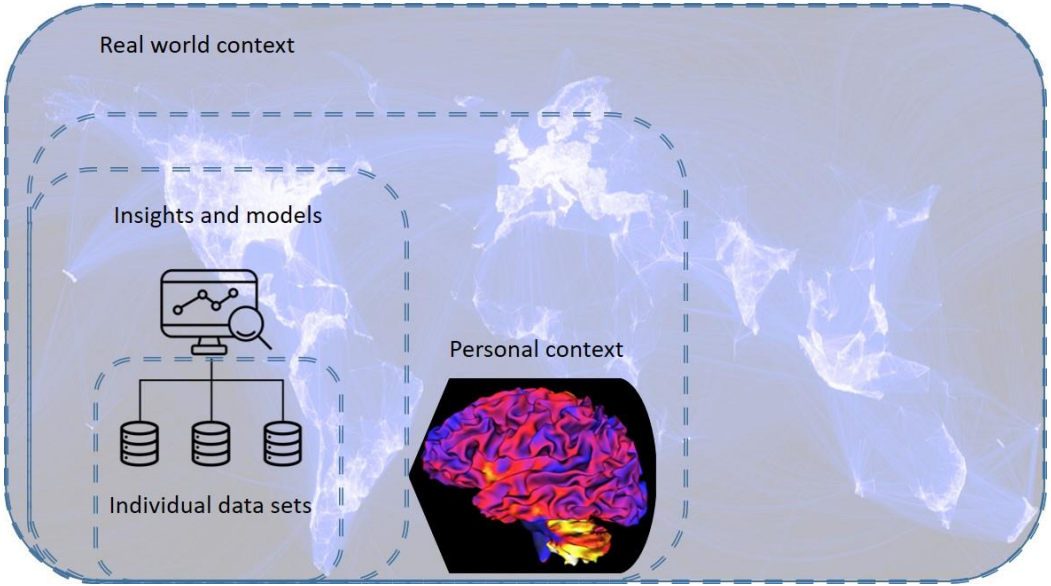




Context : Different environments for use of data

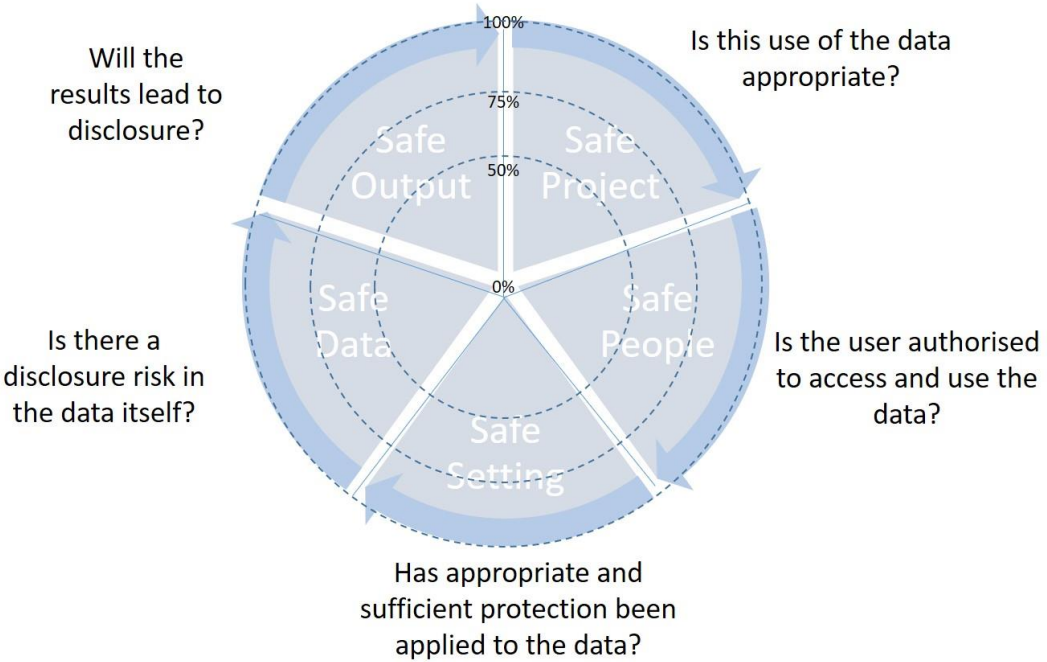






Privacy Preserving Data Sharing

Quantified “Fives Safes”



Other Risks

Safe Organisation

- refers to the systems and processes employed by an organisation to ensure the Safes Framework is applied throughout the project and with the long-term management of data and outputs. Safe organisations may include those which adhere to data protection, quality standards and cyber security standards. Safe Organisation may consider

Safe Lifecycle

- refers to the time sensitivity of a dataset or output. Data may be highly sensitive for a specific period and then may be not sensitive at all. For example, a city plan that might involve the mandated acquisition of an individual's home to enable the construction of a new road may be very sensitive until the home is demolished. At this time there is no remaining value in protecting the data or output. Considering the complete lifecycle of a dataset may add additional insight and tools to help effectively anonymise and protect privacy rights. Safe Lifecycle may consider

Safe Outcomes

- refers to the ultimate uses of the project outputs. A variety of "Outcomes Frameworks" have been developed which can be informed by the outputs of individual data linkage and analysis projects. Safe Outcomes may consider

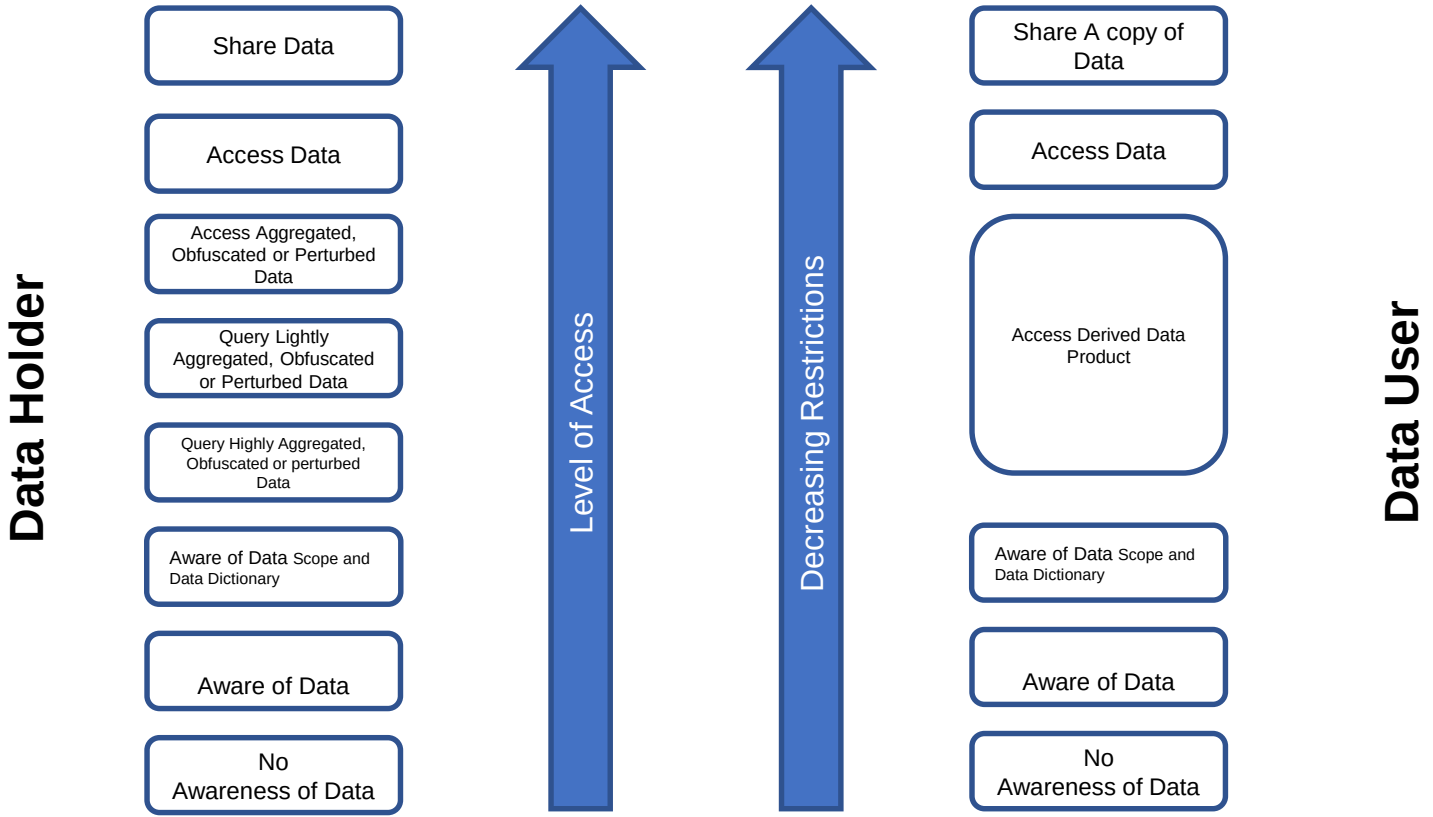
Safe Use

- refers to the use of the outputs within the Outcomes framework specifically, how much interpretation or context is required to appropriately use the outputs, including the degree to which a decision or action can be informed, or automated based on this output. Safe Use may consider

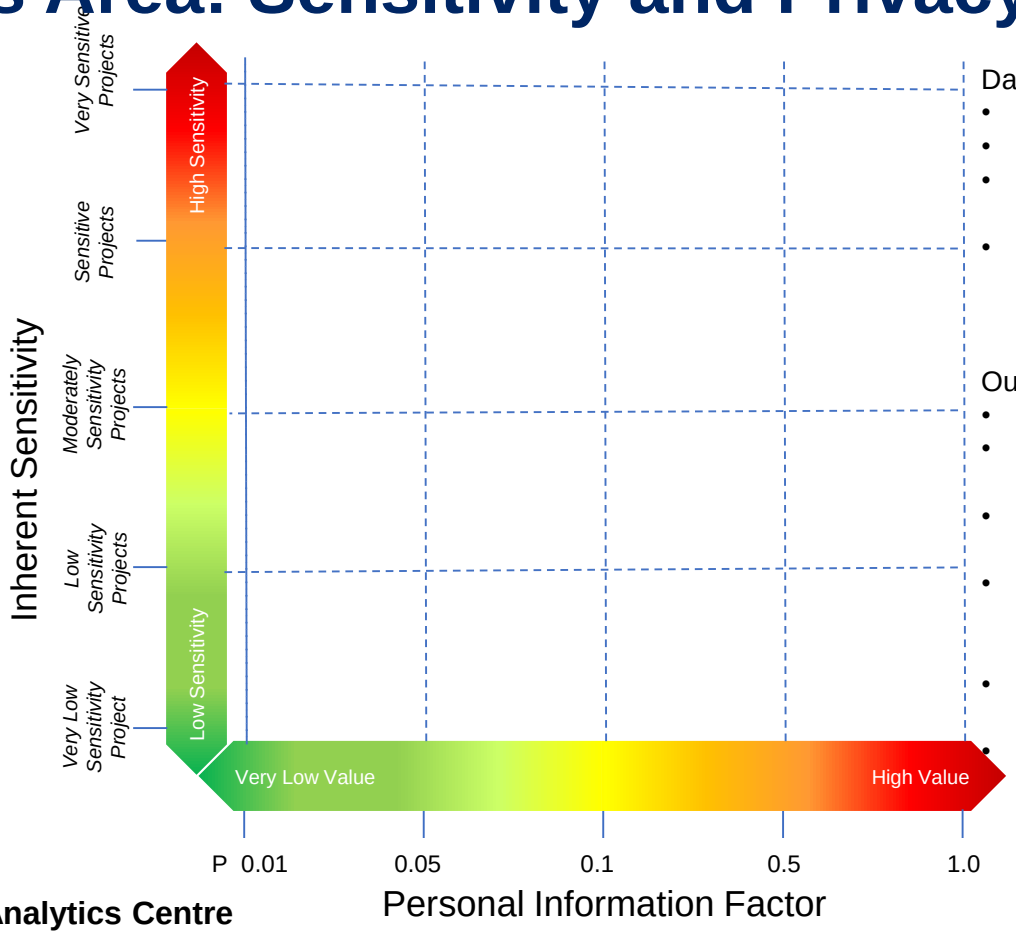
Safe Response

- refers to the systems and process which need to be in place to address adverse consequences of sharing of data or sharing or actions taken based on outputs. Safe Response may consider

Focus Area : Data Sharing and Use Frameworks



Focus Area: Sensitivity and Privacy



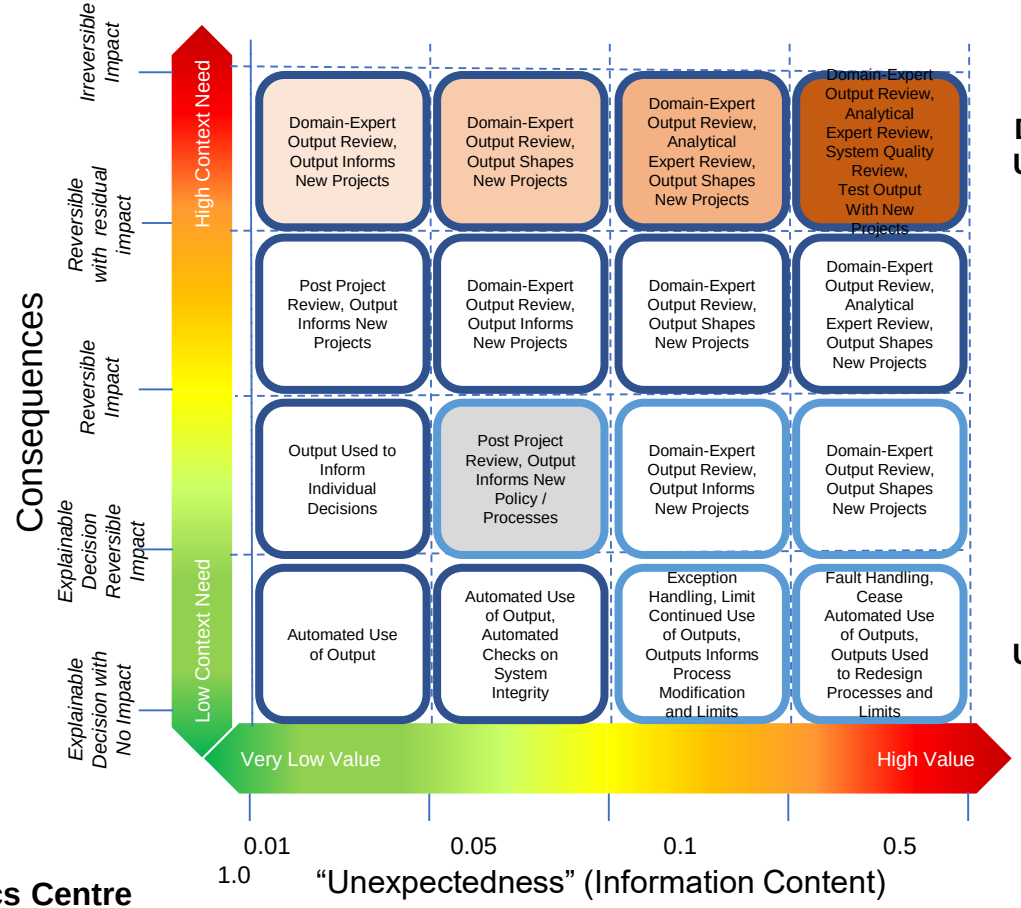
Data driven sensitivities:

- Vulnerable groups (e.g. at risk populations)
- Sensitive subjects (e.g health data)
- Linked data (or linkage is occurring as part of the project)
- Information relating to ethnic or racial origin, political opinions, religious or philosophical beliefs, trade union membership, health or sexual life.

Output driven sensitivities:

- The audience and the use cases for the outputs
- Outputs require high levels of context so as not to be misinterpreted
- Outputs to be used in decision making when the quality of the data/analytics may not support this
- Potential to have unintended consequences (e.g. contradicts government policy, social implications, causes harm to individuals)
- Impacts for individuals, groups or organisations (consider who is impacted and to what extent?)
- Produce unexpected results/insights (e.g. The findings differ from project expectations or other research in the area)

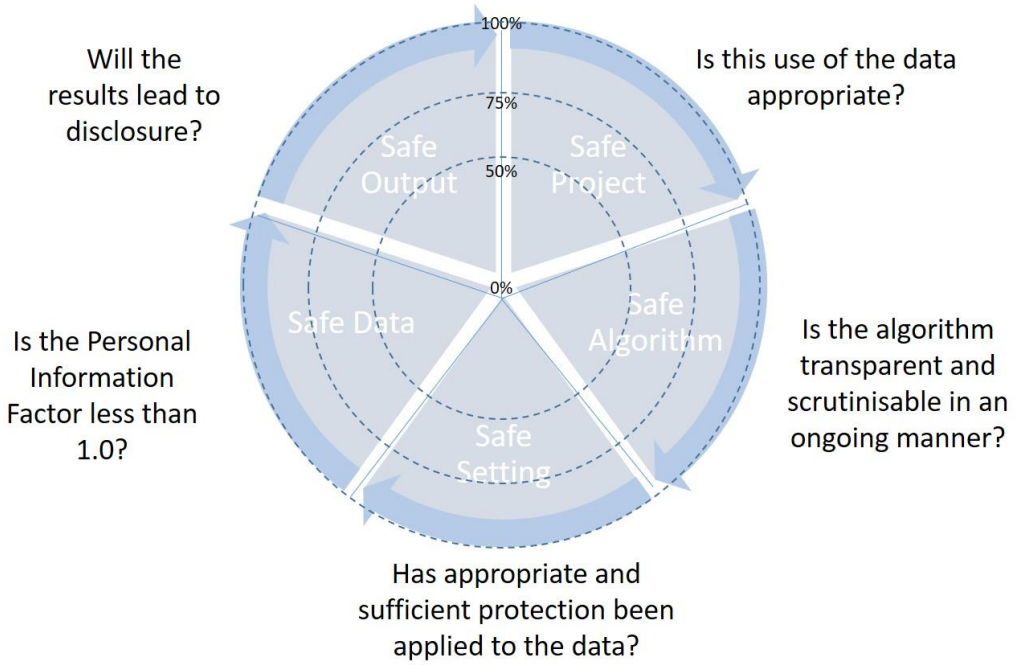
Focus : Safe Use of Outputs (Consequences)

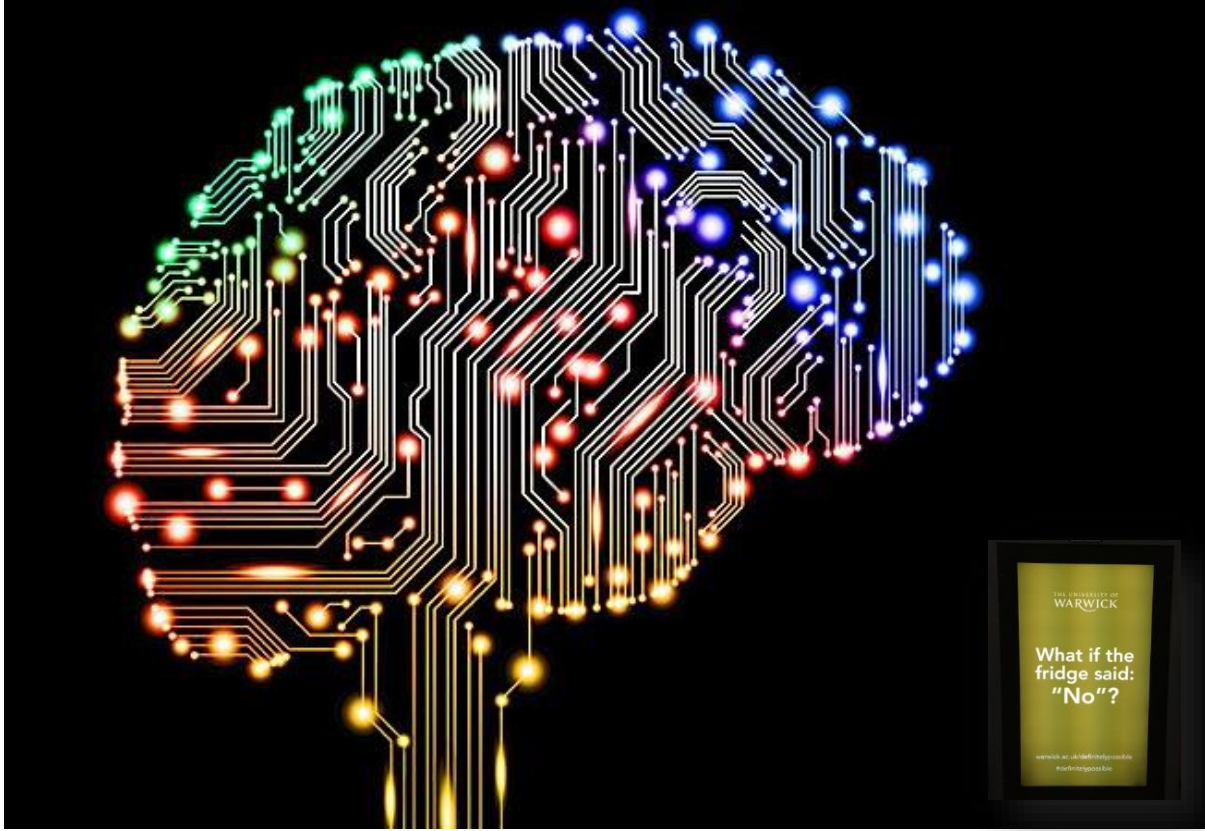


Domain Expert Use of Outputs

Autonomous Use of Outputs

What happens when People are Algorithms?





Standards Needed

Dr. Ian Oppermann
CEO and Chief Data Scientist, NSW Data Analytics Centre

Data Analytics Centre |Department of Customer Service
e ian.oppermann@customerservice.nsw.gov.au | www.dac.nsw.gov.au
2 – 24 Rawson Place, Sydney NSW 2000

NSW Data Analytics Centre