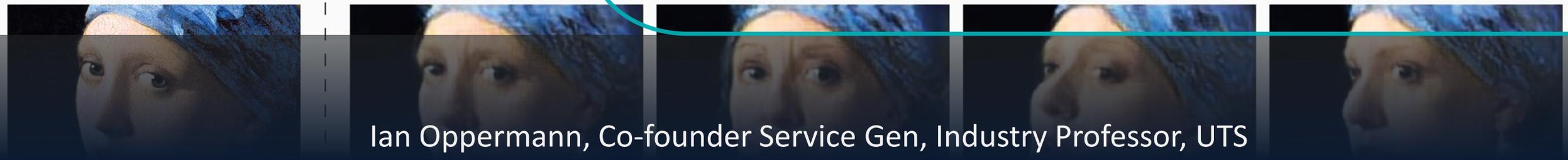The Importance of Data Governance and Privacy in a Rapidly Evolving Digital World

… welcome to the new normal…

Ian Oppermann, Co-founder Service Gen, Industry Professor, UTS

We want to use data for good purposes

Emergency First Responder

"I want a real time data asset which gives the mobile phone coverage of all operators during an emergency, and to I want to know what coverage is at risk from a propagating fire front"

Boosting the Circular Economy

"We want to build a data asset which shows the journey of every company dealing with waste in the Circular Economy. We will use this data asset to reform the system"

Transport

"We want to build a real-time bus passenger data asset to help schedule additional resources during peak times, and better connect with other transport services."

Defibrillator Data Asset

"I want to build, maintain and provide access to a data asset with all of the location, accessibility and condition of all AED's in my State. This will help responders, researchers and people planning where to put more"
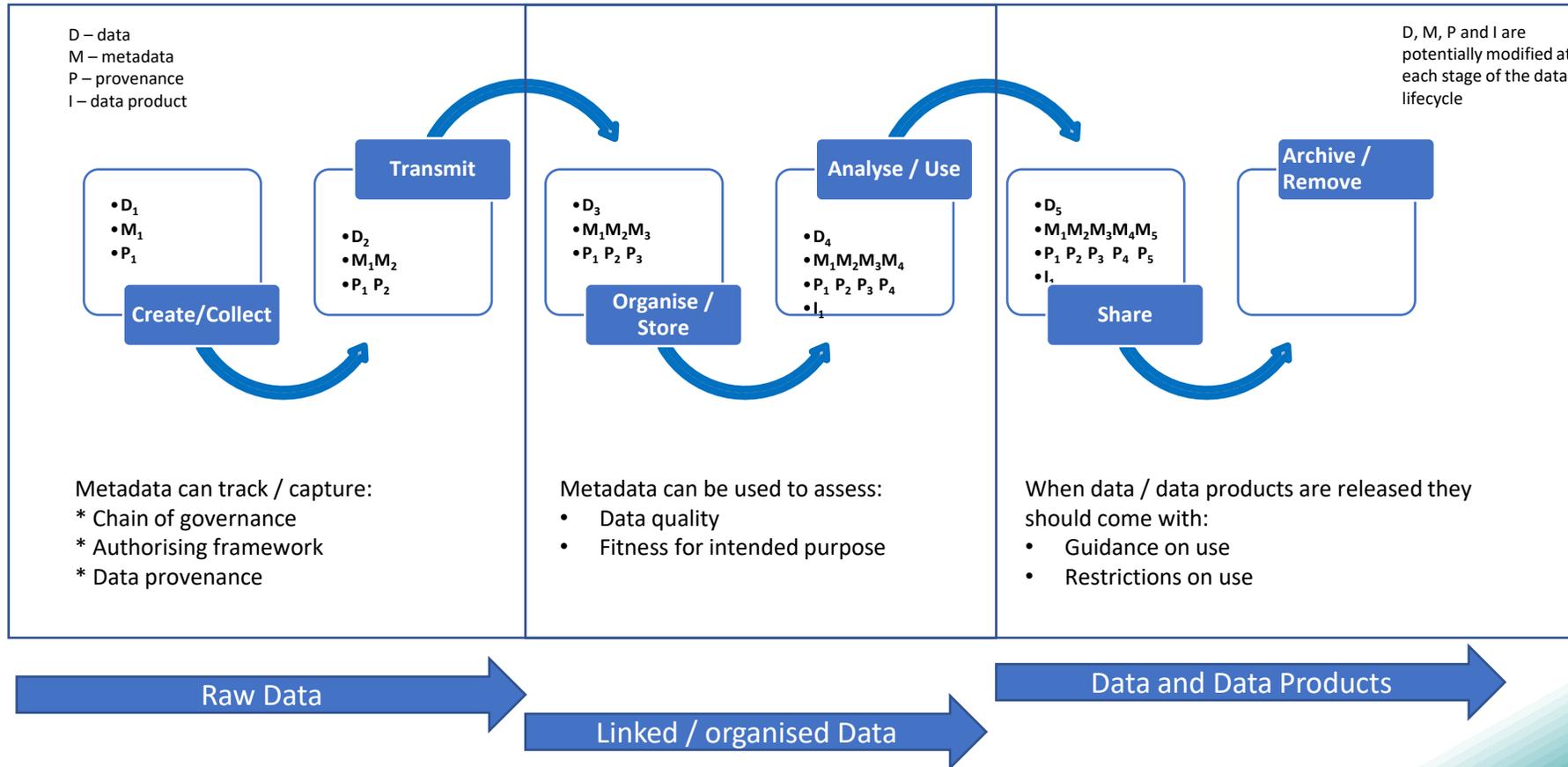
STEM Education

"I want to understand the journey of a child at school and understand the factors that lead to them firstly taking, and then completing STEM subjects".

Where do you start?

A simplified data lifecycle

Data lifecycles can be infinite. A simplified model allows you an entry point when you are considering "use"

D – data
M – metadata
P – provenance
I – data product

D, M, P and I are potentially modified at each stage of the data lifecycle

Transmit

Create/Collect

- D_1
- M_1
- P_1

- D_2
- $M_1 M_2$
- $P_1 P_2$

Analyse / Use

Organise / Store

- D_3
- $M_1 M_2 M_3$
- $P_1 P_2 P_3$

- D_4
- $M_1 M_2 M_3 M_4$
- $P_1 P_2 P_3 P_4$
- I_1

Archive / Remove

Share

- D_5
- $M_1 M_2 M_3 M_4 M_5$
- $P_1 P_2 P_3 P_4 P_5$
- $I.$

Metadata can track / capture:
* Chain of governance
* Authorising framework
* Data provenance

Metadata can be used to assess:
- Data quality
- Fitness for intended purpose

When data / data products are released they should come with:
- Guidance on use
- Restrictions on use

Raw Data

Linked / organised Data

Data and Data Products

Data (data products) can be both "personal" and "sensitive"

Some sensitivities are common. Some are very specific to your environment.

Sensitivities about data itself:
1. Concerns that data contains high levels of personal information
2. Concerns that data contains uniquely identifiable individuals
3. Concerns that sensitive subjects are captured in data (culturally subjective but often described e.g. religion)
4. Concerns about data quality (accuracy, timeliness, completeness, and consistency)
5. Concerns about fitness-for-purpose of data for analysis

Sensitivities about capability and governance:
6. Concerns that context is not captured with data (metadata, provenance, consent)
7. Concerns about authority to share data for analysis
8. Concerns about poor governance or accidental release of data or insights (outputs)
9. Concerns that expert knowledge / context is required to appropriately interpret data and results of analysis
10. Concerns about authority to release results of analysis

Sensitivities about use of insights:
11. Concerns about the level confidence in outputs (accuracy, precision, consistency, explainability, bias)
12. Concerns about unintended consequences from how outputs (insights / data driven decisions) will be used
13. concerns about whether human judgement will be applied before an insight becomes a decision
14. Concerns possible harms resulting from use of outputs (reversable, reversable with cost, irreversible)
15. Concerns that results from analysis may lead to negative surprises (especially for data not analysed before)
16. Concerns that commercial value may be degraded if insights are shared
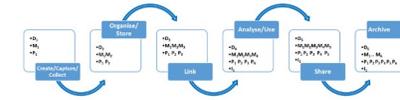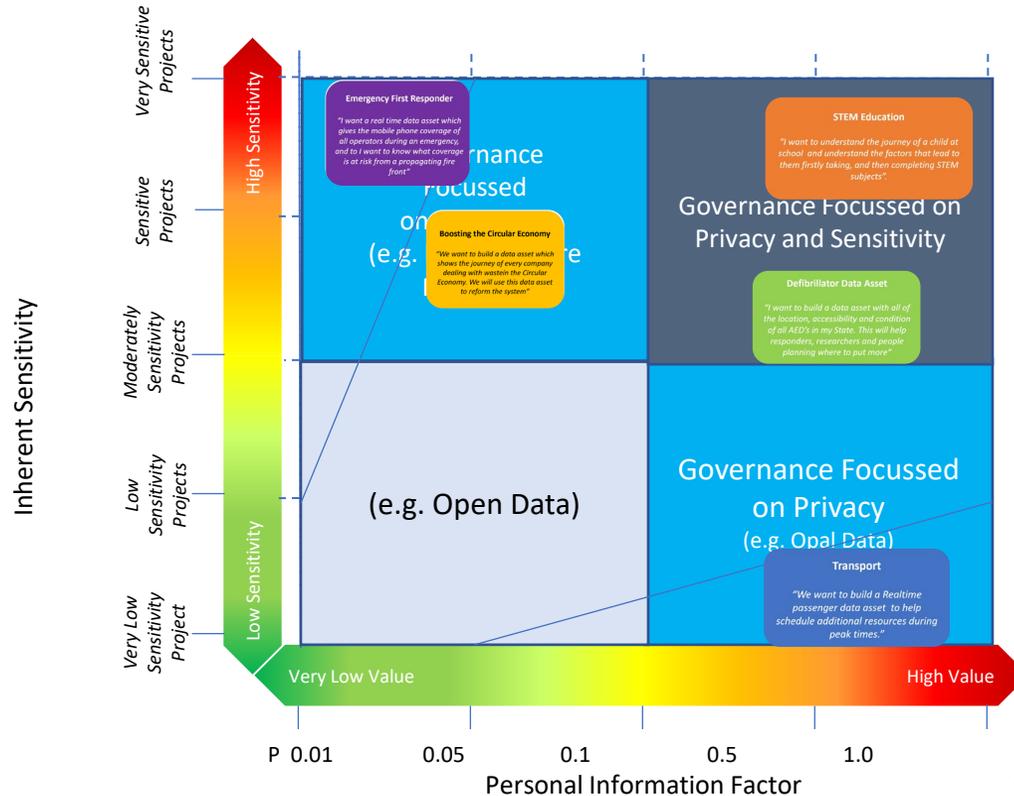
L   M   H
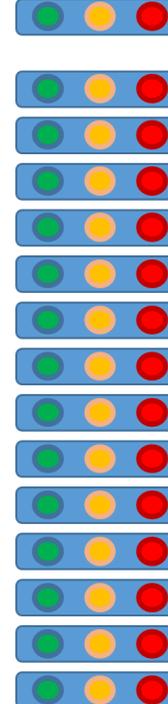
PIF: L   M   H

Inherent Sensitivity: L   M   H

Governance can be focused on "personal" and "sensitive"

These sensitivities can change as data is "used" and data products created.

This governance framework should be applied to think about where mitigations should be applied across the data lifecycle based on

- The level of personal information in a data set
- The inherent sensitivity of the data itself or the use of that data
- Levels of protection which must be applied to data products produced at the "Use" step of the data lifecycle.

A PIF tool demonstration video is available at https://www.youtube.com/watch?v=wrD6FI2U4Rs.

An open source PIF tool is available at https://github.com/PIFtools/piflib.

Frameworks of control can be designed for "personal" and "sensitive"

These sensitivities can change as data is "used" and data products created.
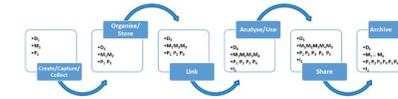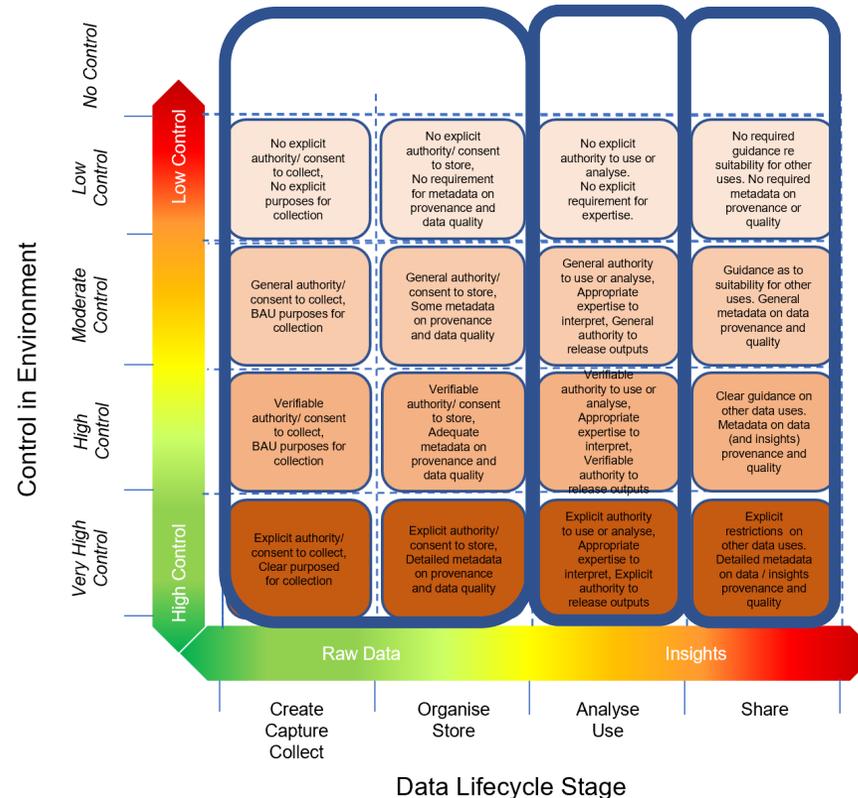

May have assumed authority to collect, use, and Use data. May have metadata on data provenance and quality. **Data - low PIF.**

Must have understanding of data quality and provenance, capable analysists and domain experts, adequate governance / security at each stage. **May have** broad authority to collect, use, and Use data. **Data - moderately sensitive / moderate PIF.**

Must have understanding of data quality and provenance, highly skilled analysists and domain experts, strong governance / security at each stage. **May have** general authority to collect, use, and Use data. **Data - high sensitivity / high PIF.**

Must have explicit purpose and authority, high quality data and metadata, expert analysts and domain experts, strong governance / security at each stage. Explicit restrictions on secondary use of data and insights. **Data - very high sensitivity and very high PIF**

Control in Environment

	No Control	Low Control
No Control / Low Control	No explicit authority/ consent to collect, No explicit purposes for collection	No explicit authority/ consent to store, No requirement for metadata on provenance and data quality	No explicit authority to use or analyse. No explicit requirement for expertise.	No required guidance re suitability for other uses. No required metadata on provenance or quality
Low Control / Moderate Control	General authority/ consent to collect, BAU purposes for collection	General authority/ consent to store, Some metadata on provenance and data quality	General authority to use or analyse, Appropriate expertise to interpret, General authority to release outputs	Guidance as to suitability for other uses. General metadata on data provenance and quality
High Control	Verifiable authority/ consent to collect, BAU purposes for collection	Verifiable authority/ consent to store, Adequate metadata on provenance and data quality	Verifiable authority to use or analyse, Appropriate expertise to interpret, Verifiable authority to release outputs	Clear guidance on other data uses. Metadata on data (and insights) provenance and quality
Very High Control / High Control	Explicit authority/ consent to collect, Clear purposed for collection	Explicit authority/ consent to store, Detailed metadata on provenance and data quality	Explicit authority to use or analyse, Appropriate expertise to interpret, Explicit authority to release outputs	Explicit restrictions on other data uses. Detailed metadata on data / insights provenance and quality

Raw Data → **Insights**

Create Capture Collect | Organise Store | Analyse Use | Share

Data Lifecycle Stage

- Control = (proven) capability * (assessable) governance * (verifiable) purpose

- Capability includes skill in all stages of Data Lifecycle - data analysis, data provenance, governance, security

- High Control = skilled people working in strong governance environment with clearly authorised purpose

- No Control environment = no assessments or no restriction on people accessing or utilising data

- Requires an objective, repeatable, standardised assessment of
  - capability,
  - governance,
  - purpose,
  - data quality and provenance
  - sensitivity of data
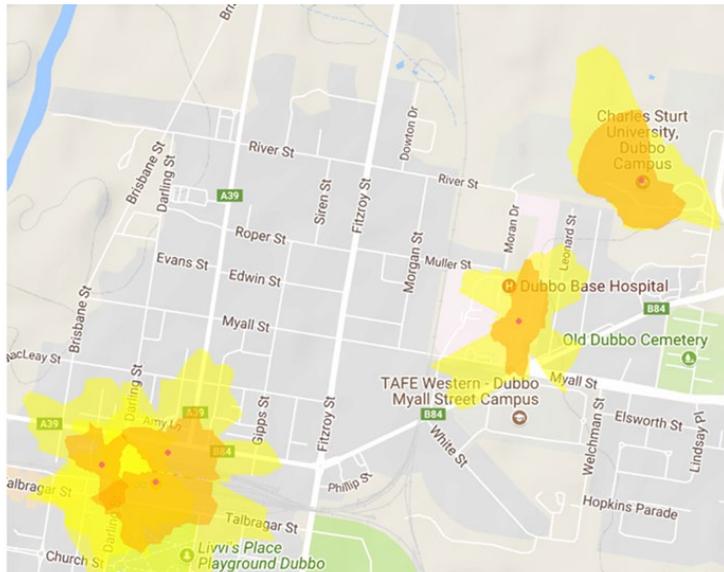 - degree of personal information contained in datasets

Real World Example – Defibrillator Data Asset

These sensitivities can change as data is "used" and data products created.
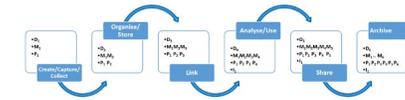
Where is the AED Near Me?
**Real-time access, current (time of day) AED availability**

Authorising Framework: You are a data custodian within an appropriately authorised State Government Authority. AED information may be supplied by AED Owners or crowdsourced from the public. You are producing a plan to describe the data asset, how to build it and maintain it, and how to provide access.

ymbols: 2.5 min from AEDs
mbols: 5 min from AEDs

Example Use case:
AED Coverage in the Community
**Non-real-time access, historic data**
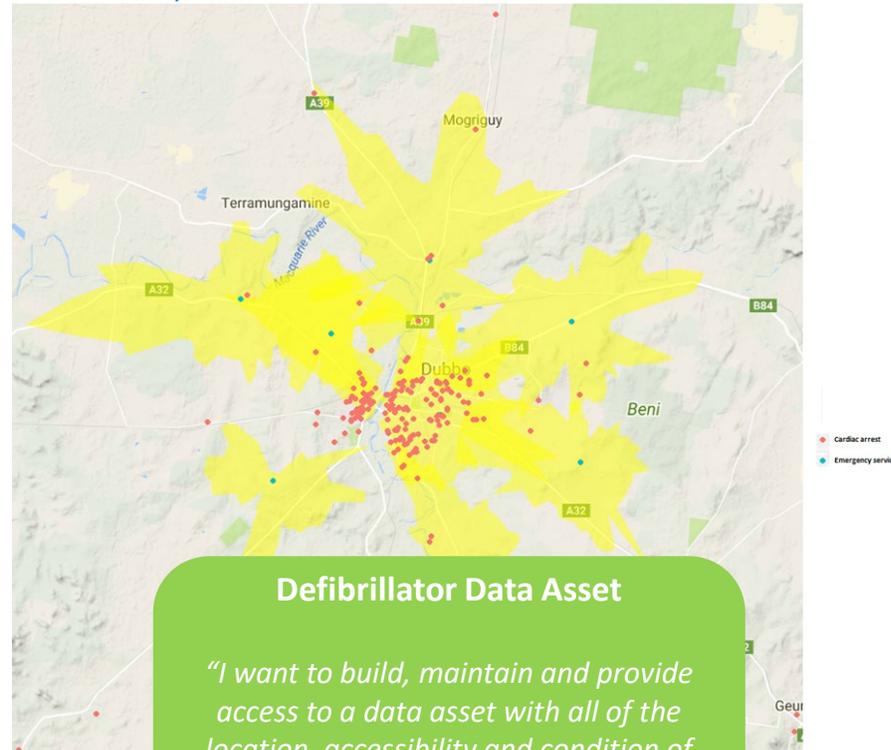
Defibrillator Data Asset

"I want to build, maintain and provide access to a data asset with all of the location, accessibility and condition of all AED's in my State. This will help responders, researchers and people planning where to put more"

Real World Example – Defibrillator Data Asset

Different data products for different levels of control.
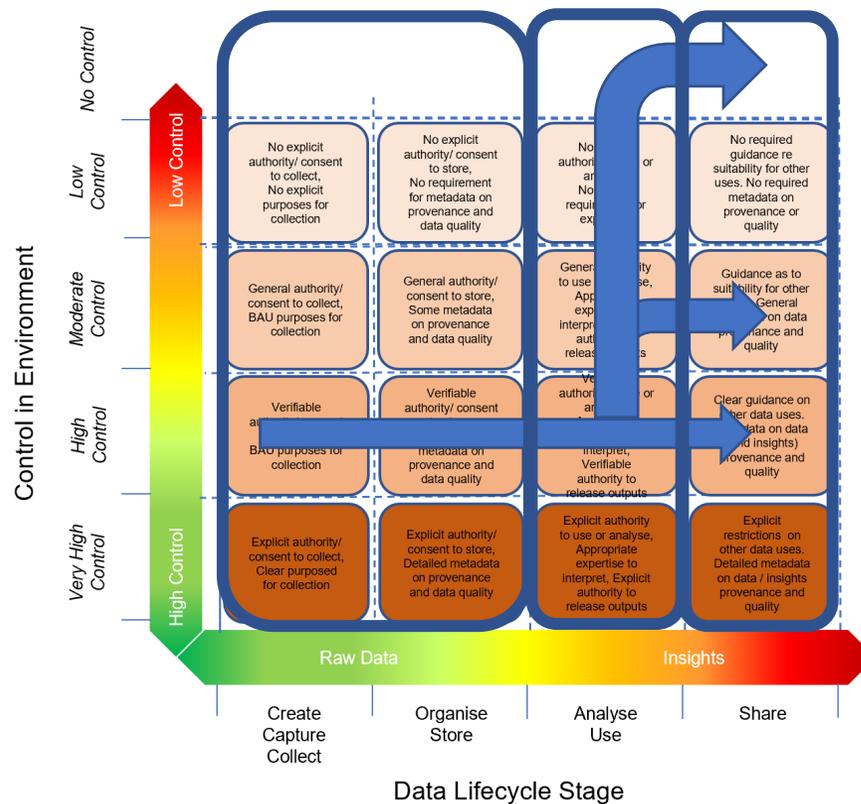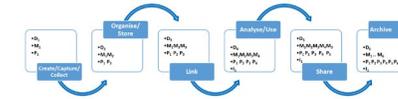
Three data products which could be created are:

- <u>High Control Environment</u>: Raw data with personal information and sensitive location information. Only accessed by data custodians.
- <u>Moderate Control Environment</u>: Raw data without personal information and with sensitive location information. Accessed by researchers and policy makers under conditions of confidentiality.
- <u>No Control Environment</u>: Raw data without personal information and without sensitive location information.

A Very High Control Environment is not required as data collection and use does not require a special legal instrument.

International Standards for Data Sharing and Data Quality

In April 2024, two international standards were published by ISO and IEC which built on the work undertaken with the ACS:

- ISO/IEC 5207:2024 (Terminology and use cases)
- ISO/IEC 5212:204 (Guidance for data use)

ISO/IEC 5207:2024 - Information technology — Data usage — Terminology and use cases
ISO/IEC 5212:2024 - Information technology — Data usage — Guidance for data usage
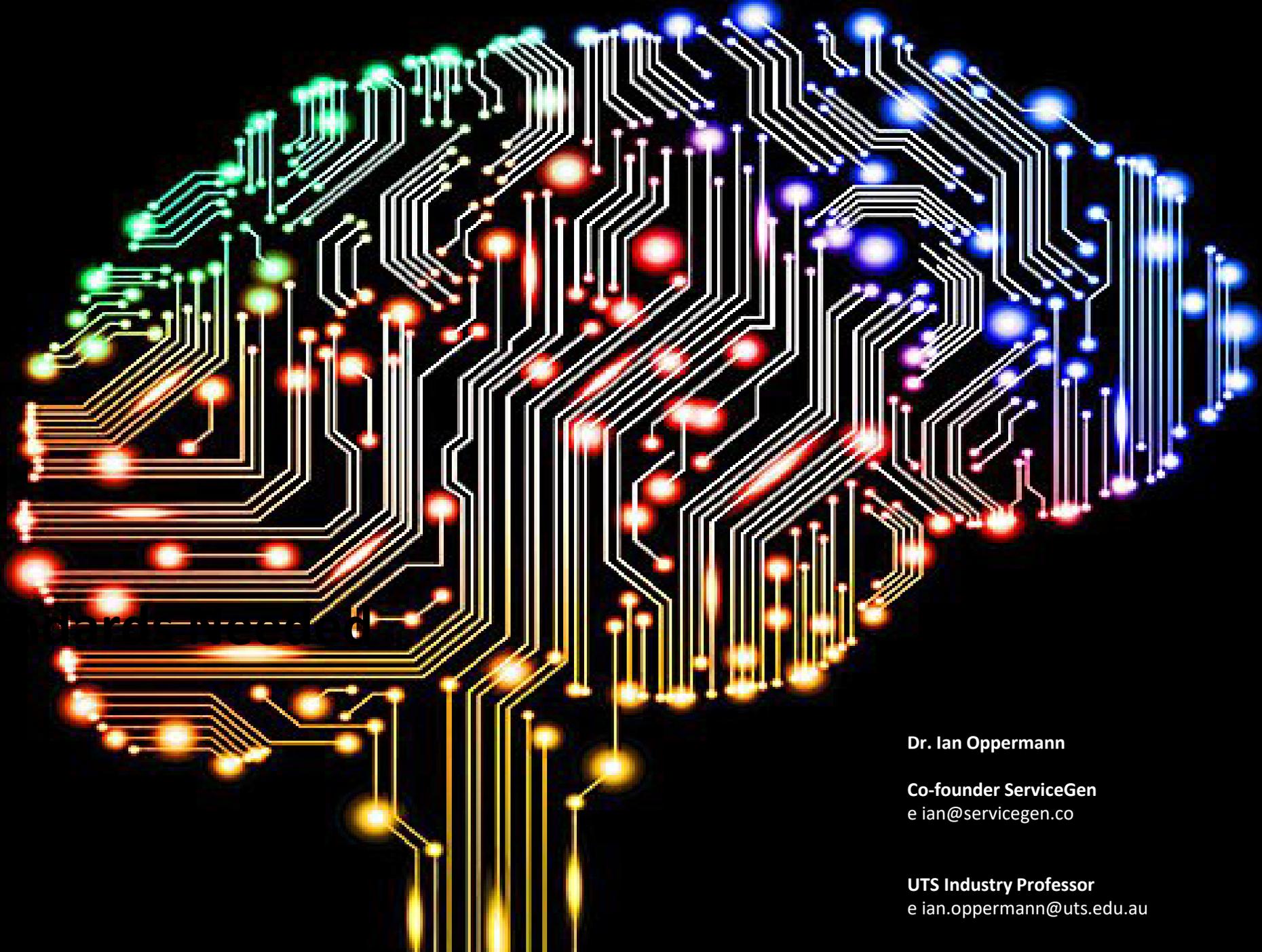
Some other useful standards – Data Quality for AI and ML

ISO/IEC AWI 5259-1, Data quality for analytics and ML — Part 1: Overview, terminology, and examples

ISO/IEC AWI 5259-2, Data quality for analytics and ML — Part 2: Data quality measures

ISO/IEC AWI 5259-3, Data quality for analytics and ML — Part 3: Data quality management requirements and guidelines

ISO/IEC AWI 5259-4, Data quality for analytics and ML — Part 4: Data quality process framework

Data and Digital
Standards Landscape

JULY 2022

STANDARDS Australia
CELEBRATING 100 YEARS 1922-2022

Welcome to the new normal

What if the fridge said: "No"?

warwick.ac.uk/definitelypossible
#definitelypossible

Dr. Ian Oppermann

Co-founder ServiceGen
e ian@servicegen.co

UTS Industry Professor
e ian.oppermann@uts.edu.au

ServiceGen