



arc training
centre for
**information
resilience**

Information Resilience and the Role of Data Governance

Javad Pool

Thursday 11 July 2024





arc training centre for information resilience

- CIRES is an Australian Research Council (ARC) Industrial Transformation Training Centre
- Commenced at the end of July 2021 and will run for 5 years with a total funding of \$13.8M
- CIRES hosts over 50 staff including research leaders from computer science, mathematics, and business, research fellows, industry and international experts, PhD and Masters Students
- CIRES will train PhD students, Research Fellows, and Data Engineers, working collaboratively with our Partners to build workforce capacity in Australian organisations and help them achieve information resilience.



Source: cires.org.au/about/

Allianz  **Partners**

Mantel
group 



Queensland Government
Queensland Health

 **MaxKelsen**

astral[★]



 **health+wellbeing**
Queensland *making healthy happen*

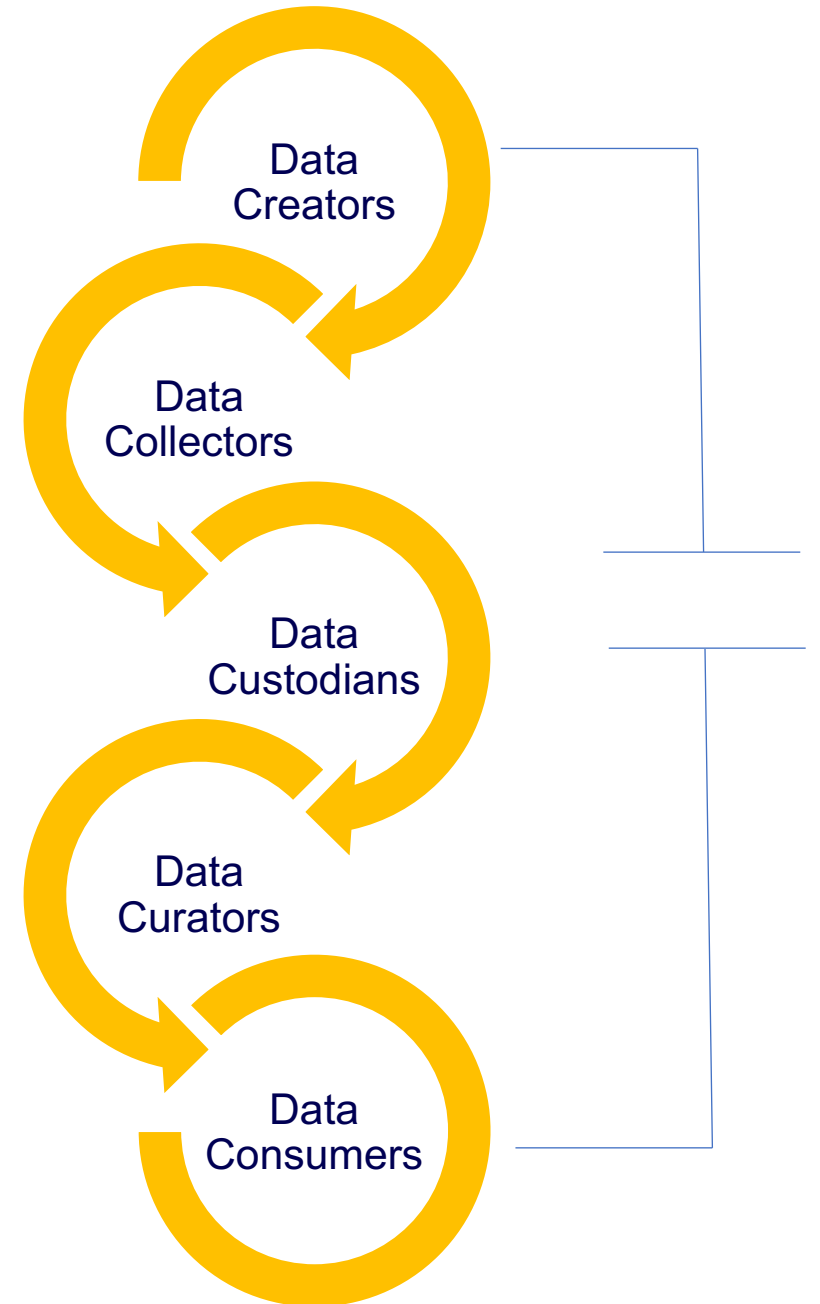


Queensland Government
Department of Education and Training

Information Resilience

... the capacity of organisations to build, protect, and sustain agile data pipelines, capable of detecting and responding to failures and risks across the value chain in which the data is sourced, shared, transformed, analysed, and consumed.

Source: Sadiq et al. (2022)



Information Resilience Themes

1. Responsible use of data
2. Data curation at scale
3. Algorithmic transparency
4. Trusted data partnerships
5. Agility in value creation from data



1. Responsible Use of Data

By '**responsible use of data**', we mean *'ensuring that data is utilised ethically and for intended purposes, with a focus on social acceptance, clear analytics agendas, and fostering data literacy within organisations'*.

➤ Example:

*A healthcare agency can conduct community meetings to discuss data use transparency, gaining stakeholder support and **social acceptance** for medical practices.*

2. Data Curation at Scale

By '**data curation at scale**', we mean *'the systematic approach to managing, preparing, and maintaining extensive data sets, ensuring their quality, usability, and relevance across various organizational functions and processes'*.

➤ Example

To ensure data quality discovery, organizations can develop methods to measure the quality of datasets according to specific needs, enhancing understanding of data fitness for various purposes.



3. Algorithmic Transparency

By '**algorithmic transparency**' we mean *'the incorporation of interpretability, unbiasedness, and transparency into learning algorithms, ensuring explainability and fairness in their predictions and outcomes'*.

➤ *Example*

Organisations can use methods to check if algorithms are fair, like testing how changing certain factors affects predictions. By ensuring that outcomes aren't unfairly influenced by specific factors, they ensure fairness and transparency.

4. Trusted Data Partnerships

By '**trusted data partnerships**' we mean *'collaborative efforts aimed at fostering data literacy, trust and competence in sharing data. This involves ensuring privacy and security while facilitating efficient data discovery, robust data linkage, data provenance, and effective data governance'*.

➤ *Example*

Organisations prioritise privacy and security when sharing data by implementing techniques like anonymisation to protect sensitive information.



5. Agility in Value Creation from Data

By '**agility in value creation from data**', we mean *'agile adaptation and utilisation of data-driven solutions within IT landscapes and business processes, focusing on organisational structures and mechanisms to maximise value creation'*.

➤ *Example*

Organisations can restructure teams to integrate data analytics capabilities into every part of the organization, fostering collaboration between analytics and business groups to maximize value creation and data-driven decision-making.





arc training
centre for
**information
resilience**

Role of Data Governance for Information Resilience



Information Resilience and Data Governance



What Is Data Governance?

Data Governance is ... “the structured integration of Data Management practices (procedures and methods) in the organizational structures and processes of a company”.

**APPLYING THE PRACTICES OF DATA MANAGEMENT
DATA QUALITY MANAGEMENT, DATA ENGINEERING, DATA ANALYTICS &
REPORTING, AND MORE**

**EMBEDDING DATA MANAGEMENT PRACTICES INTO THE
ORGANIZATIONAL AND OPERATIONAL STRUCTURE
DATA GOVERNANCE ROLES**

Source: Bollweg (2022)



Decision Domains

➤ **Data Principles**

- *Clarifying the role of data as an asset*

➤ **Data Quality**

- *Establishing the requirements of intended use of data*

➤ **Metadata**

- *Establishing the semantics or “content” of data so that it is interpretable by the users*

➤ **Data Access**

- *Specifying access requirements of data*

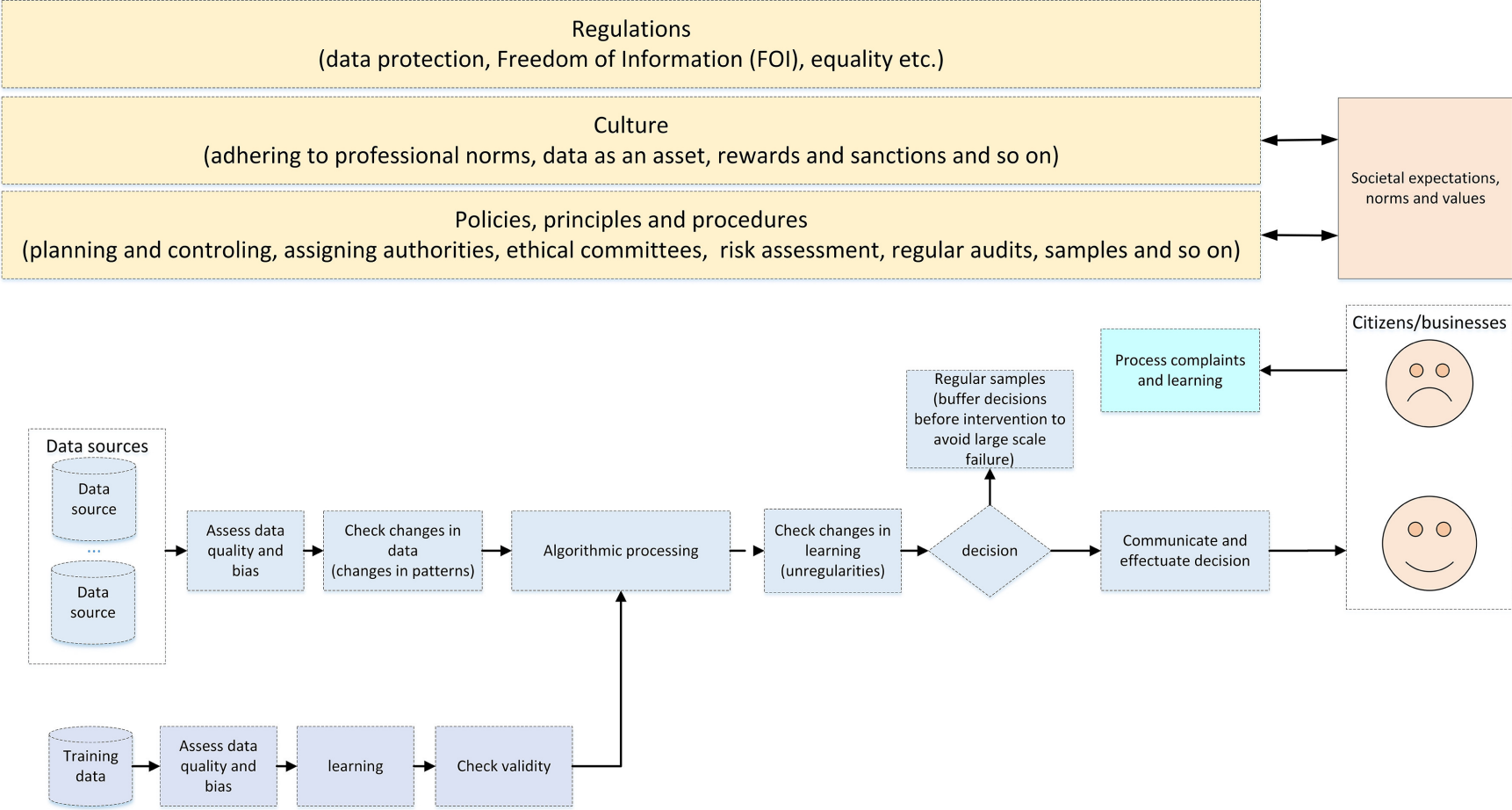
➤ **Data Lifecycle**

- *Determining the definition, production, retention and retirement of data*

Source: Khatri & Brown (2010)

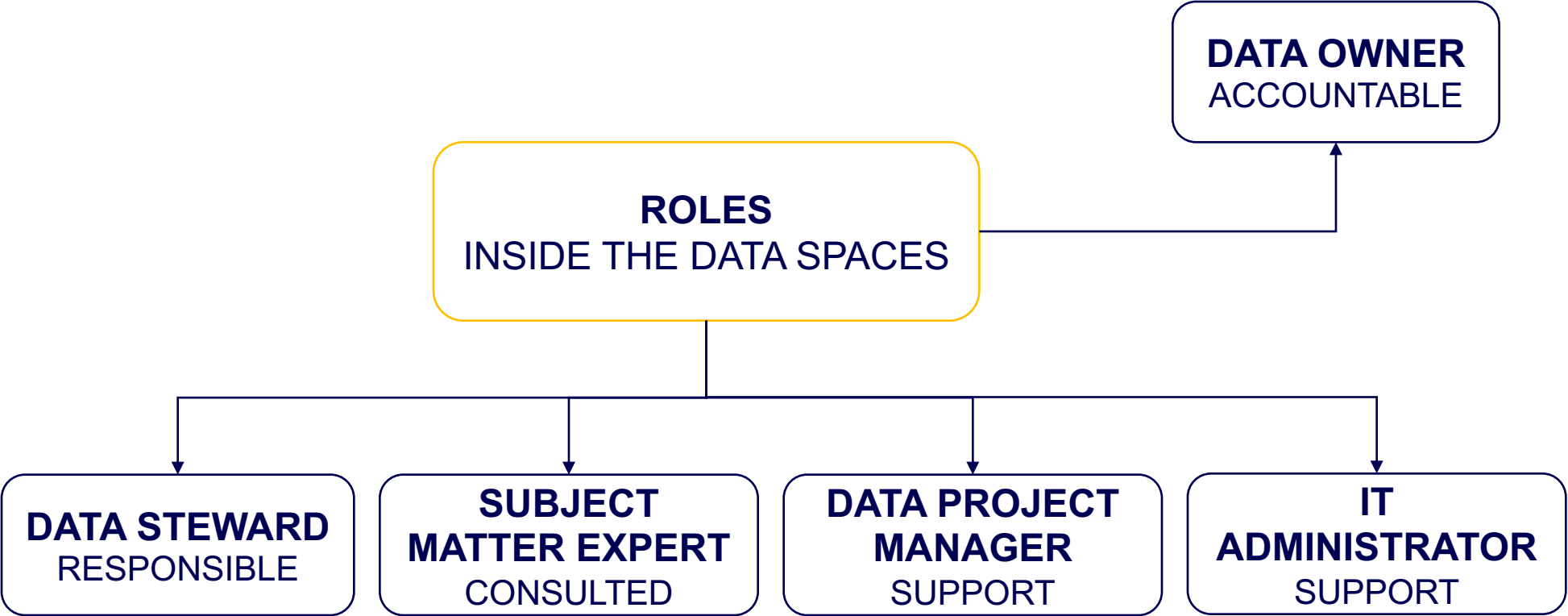


System-Level Governance Model of Big Data



Source: Janssen et al. (2020), p.4

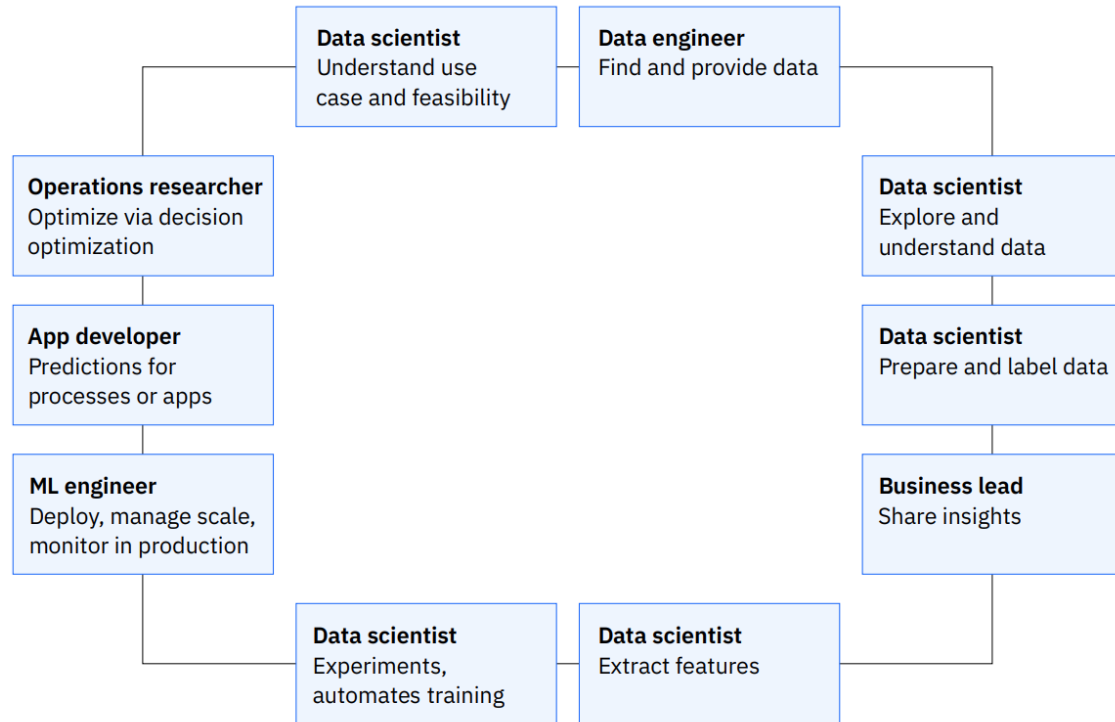
Roles of Data Governance



Source: Bollweg (2022). p.26



Roles Across the AI Lifecycle



Encourage collaboration with key stakeholders and understand their top concerns:

Chief Financial Officer (CFO), risks to profitability

Chief Marketing Officer (CMO), risks to brand

Chief Risk Officer (CRO), risks to enterprise

Chief Data Officer (CDO), efficient data operations

Chief Human Resources Officer (CHRO), potential talent impacts

Chief Executive Officer (CEO), organizational accountability

Chief Privacy Officer (CPO), regulatory accountability

‘AI models are not created equally. But all models must be governed.’

By defining clear roles and responsibilities and engaging with stakeholders from different departments, we can build and govern AI systems that are both innovative and ethical.

Source: IBM, [watsonx.governance](https://www.ibm.com/watsonx/governance)



References

- *Bollweg, L. M. (2022). What Is Data Governance?. In Data Governance for Managers: The Driver of Value Stream Optimization and a Pacemaker for Digital Transformation (pp. 9-45). Berlin, Heidelberg: Springer Berlin Heidelberg.*
- *Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. Government information quarterly, 37(3), 101493.*
- *IBM (2023), Build responsible AI workflows with AI governance, Retrieved July 2024, from <https://www.ibm.com/downloads/cas/G9KXO4WK>*
- *Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148-152.*
- *Sadiq, S., Aryani, A., Demartini, G. et al. (2022), Information Resilience: the nexus of responsible and agile approaches to information use. The VLDB Journal, 31, 1059–1084.*



Thank You!



Australian Government
Australian Research Council



THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA



cires.org.au  