



December 2021



# Sharing data in trusted frameworks

An ACS Technical White Paper



# About the editor



## Dr Ian Oppermann FACS

ACS President

Ian currently holds the role of NSW Chief Data Scientist and is an Industry Professor at UTS. Ian has 30 years' experience in the ICT sector and has led organisations with more than 300 people, delivering products and outcomes that have impacted hundreds of millions of people globally. He has held senior management roles in Europe and Australia as Director for Radio Access Performance at Nokia, Global Head of Sales Partnering (network software) at Nokia Siemens Networks, and then Divisional Chief and Flagship Director at CSIRO.

Ian is considered a thought leader on digital economies and is a regular speaker on big data, broadband-enabled services and the impact of technology on society. He has contributed to six books and co-authored more than 130 papers that have been cited more than 4,000 times. Ian has an MBA from the University of London and a Doctor of Philosophy in Mobile Telecommunications from the University of Sydney.

Many people and organisations worked to make this series of white papers a reality. We'd like to thank these partners:



# Foreword



## **The Hon Victor Dominello MP**

### **NSW Minister for Digital and Customer Service**

For the past two years, NSW and the world have faced an unprecedented challenge. The impacts of COVID-19 have been felt in every sector, in every household, and for many of us in government it was among the greatest public policy challenges of our careers.

As COVID spread, we were confronted with a requirement to produce and share data in ways and at velocities that we never had before. We were generating complex, detailed reports every day, based on data drawn from thousands of sources. In some instances that data was of a sensitive type and somehow we had to sift through it, share it safely and use it to drive our technical solutions and decision-making processes as we worked through the crisis.

We were able to do that because of a new breed of approaches to data sharing. These formalised approaches to the safe custody and sharing of data were critical to our response to the pandemic and paved the way to increased data sharing in the future. Having been put through the toughest crucible imaginable, we can now start to work on how we can apply these methods to increase the value of our other datasets and truly start to see the promise of government data sharing fulfilled.

As we move into smart cities and IoT, digital identities, digital currency and assets attached to blockchains, into open data on weather and traffic, into AI systems and automated decision-making processes – as we enjoy the benefits of all these new systems and technologies – data is going to be the fuel that drives it, and the management of risk is going to be critical to that process.

We have to ensure we have the processes and frameworks in place that will ensure the benefits of shared data don't come with a loss of individual liberty and privacy. That's where this work and other work being done within industry and government is so crucial.

This is the fourth of these papers I've had the privilege to introduce. It's amazing to see all the work that so many people in Australia are putting into making a data-enabled future a possibility, and I'm tremendously excited about the possibilities this work can open up for our country. It's critical work, the kind that can and will ensure Australia remains both free and prosperous, and I'm looking forward to seeing where this work can be applied and built upon in the years to come.

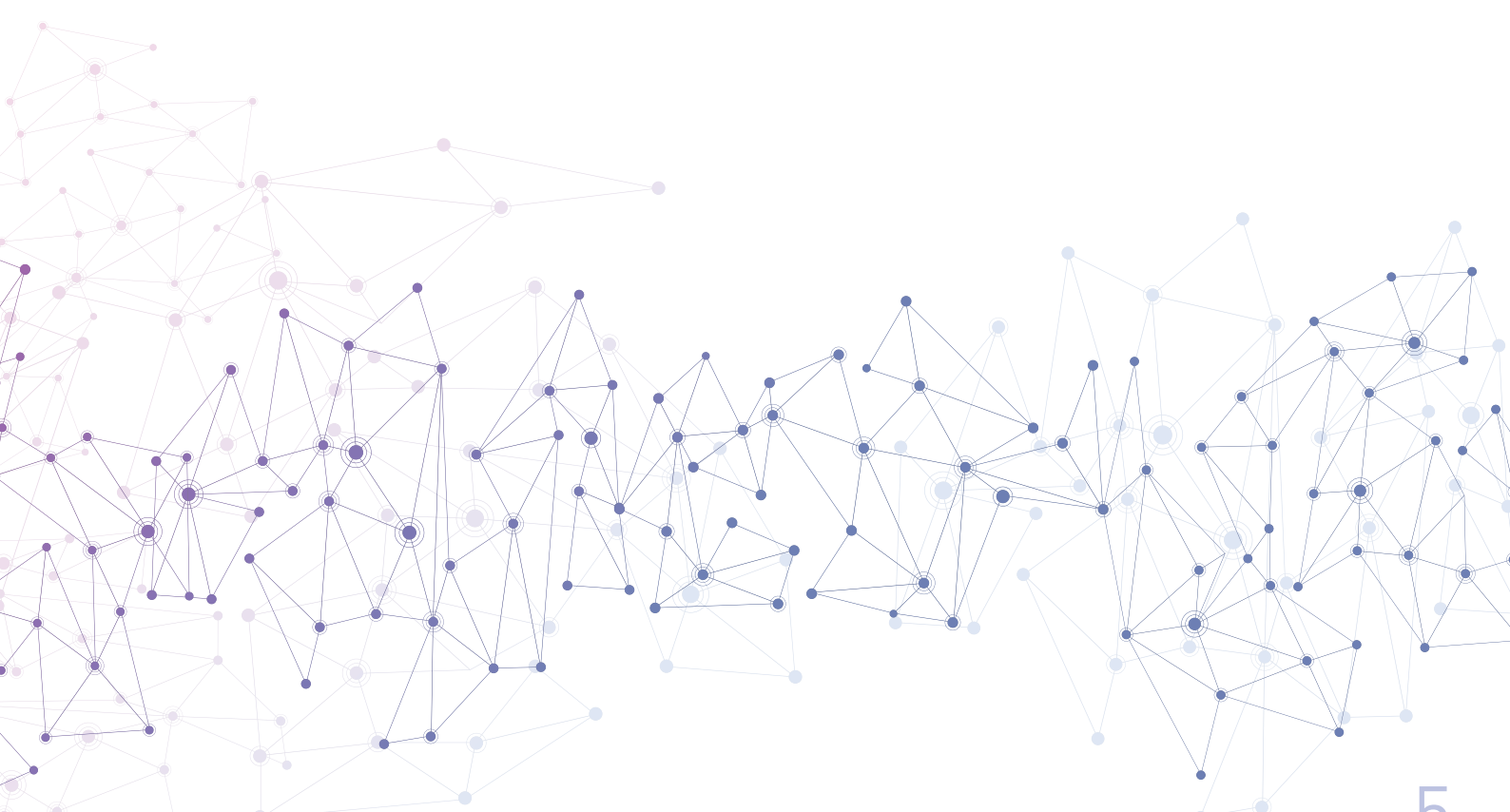
# Contents

- EXECUTIVE SUMMARY..... 6
- FRAMEWORK SUMMARY – STRUCTURE OF THIS PAPER..... 7
- 1. INTRODUCTION ..... 10
  - 1.1 The problem ..... 10
- 2. CONSIDERATIONS FOR DATA SHARING AND USE ..... 12
  - 2.1 The Five Safes model..... 13
  - 2.2 ‘Why’ matters ..... 16
  - 2.3 Data sharing is a form of data use ..... 16
  - 2.4 Modes of sharing data..... 17
- 3. SEPARATING SENSITIVITY FROM PERSONAL INFORMATION..... 20
  - 3.1 Personal information (PI) and personally identifiable information (PII) ..... 23
  - 3.2 A comment on the 2019 definition of Personal Information Factor (PIF) .... 24
  - 3.3 How many bits do you need to uniquely identify an individual in a population?..... 26
  - 3.4 Time, space, personal features and relationship features ..... 27
- 4. CONSIDERING THE WHOLE DATA LIFE CYCLE – QUALITY, METADATA AND HUMANS ..... 30





4.1	Data quality requirements are dependent on use .....	34
4.2	Humans and machines at each stage of the data life cycle .....	35
5.	GOVERNANCE ACROSS THE DATA LIFE CYCLE .....	36
6.	BRINGING IT ALL TOGETHER.....	38
6.1	Application of controls based on risk – considerations and controls.....	39
6.2	Characterising levels of control .....	40
6.3	Determining the level of control required.....	43
6.4	What is a Safe Person? .....	45
6.5	Determining the state of control at each stage of the data life cycle .....	46
7.	DISCUSSION .....	48
7.1	The work on PIF is continuing .....	49
7.2	The need for standards .....	50
8.	CONCLUSIONS.....	52
9.	THANKS.....	54
10.	APPENDIX – RESOURCES.....	56



# EXECUTIVE SUMMARY

This paper is the culmination of an effort to identify frameworks that can be used to safely share and use data. It sets out frameworks for data sharing that consider the level of personal information in data, sensitivities associated with the use of the data itself, and sensitivities in use of outputs of analysis of data. These sensitivities are addressed by variable controls at appropriate points in the data life cycle.

The work identifies controls to ensure that data is treated appropriately along its entire life cycle. It is this, often unknown, life cycle that creates so much concern for data custodians and others involved in the data ecosystem, including data subjects themselves.

The controls identified in this paper are linked to demonstrated capability, assessable governance, and clear lines of authority at each phase of the data life cycle. These link the purpose of data sharing (the 'why') with the mode of data sharing (the 'how') and provide a method to ensure sufficient governance in the circumstances.

## Key messages

---

1

Data sharing remains a challenge in many organisations. Very often the concerns raised are around privacy, but in practice, the real concerns relate to sensitivities of the data, data quality and the impact of decisions made from insights generated from the data. This paper attempts to articulate those concerns and identify mitigations for them.

---

2

The complexity of data life cycles is also identified as a limiting factor for systematic data sharing. This paper attempts to identify conditions required to be in place before data is used, and before data and data products are on-shared.

---

3

This paper presents simplified frameworks of controls for data sharing along portions of the data life cycle. These frameworks identify points of control to address sensitivities and inherent risks of data sharing and use for different types of data.

# FRAMEWORK SUMMARY – STRUCTURE OF THIS PAPER

This paper is designed to walk through the various elements that must be factored in when planning out risk management over the life cycle of a shared data set. It is organised as a loosely structured framework that covers the various considerations of a safe data set and how to manage those risks. It builds upon previous ACS white papers that cover the individual issues in more depth:

*Data Sharing Frameworks (2017)*<sup>1</sup>

*Privacy in Data Sharing: A Guide for Business and Government (2018)*<sup>2</sup>

*Privacy-Preserving Data Sharing Frameworks (2019)*<sup>3</sup>

**Chapter 1** introduces the core problem that needs to be solved.

**Chapter 2** provides an overview of core considerations of shared data. Building on a modified Five Safes framework, it looks at how to factor in modes of sharing data and the key elements of 'safe' data sharing.

**Chapter 3** expands on that, covering measures of sensitivity that include but extend upon personally identifying information. It looks at standardised models of determining sensitivity, which will inform the controls applied at each stage of the data life cycle.

**Chapter 4** examines that data life cycle. As the data moves through its life cycle, the sensitivities and therefore required controls must change. This chapter walks through the process of mapping out that data life cycle and applying appropriate controls and metadata at each stage of the life cycle.

**Chapter 5** briefly touches on governance and the management of the data through its life cycle.

**Chapter 6** brings it all together and looks at how you can take these elements and develop a unified plan for data-sharing controls over the entire life cycle of a data set. At the end, you should have a usable framework for the application of controls on the data, which will guide decision-making on the safety and usability of the data. An example of a control track can be seen in Figure 1 on page 8.

1 Available at <https://www.acs.org.au/insightsandpublications/reports-publications/data-sharing-frameworks.html>.

2 Available at <https://www.acs.org.au/insightsandpublications/reports-publications/privacy-in-data-sharing.html>.

3 Available at <https://www.acs.org.au/insightsandpublications/reports-publications/privacy-preserving-data-sharing-frameworks.html>.

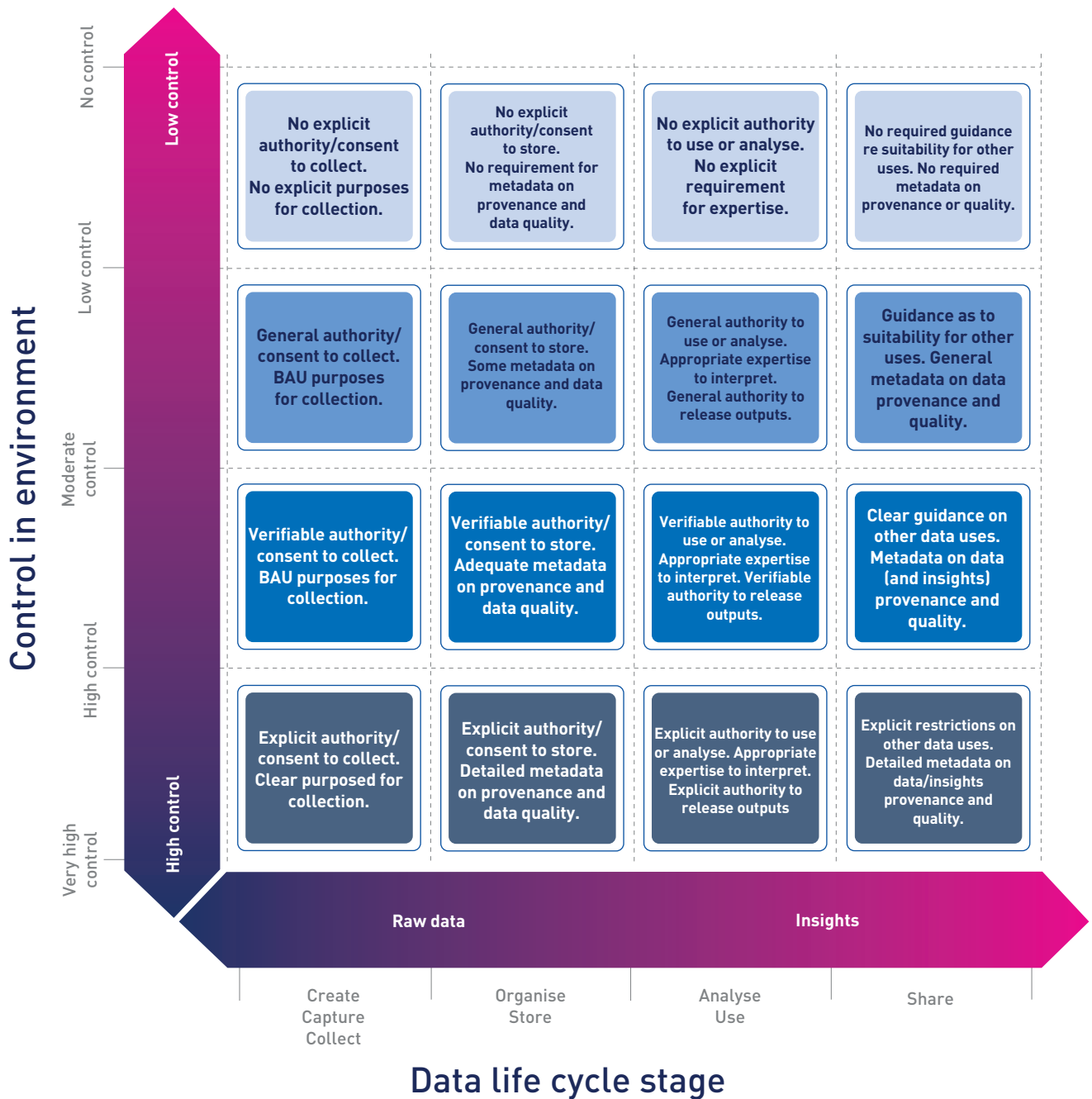


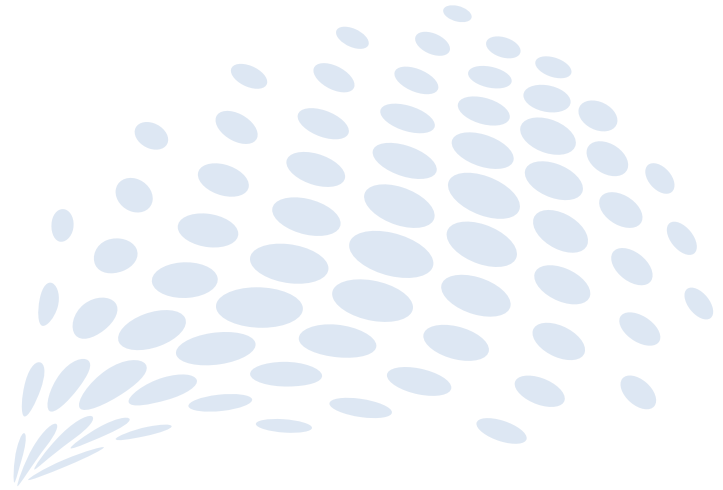
Figure 1. Characterising control layers through the data life cycle







# 01



## INTRODUCTION

### 1.1 THE PROBLEM

Data is the lifeblood of the modern economy. It impacts, enables and personalises how we work, play and engage socially and is also crucial for the operation of government and the economy. Banks and financial services companies can be described as data and digital services organisations with some bricks and mortar operations. Value comes from creating, using, protecting and sharing data. Use of data is a very wide and vague topic, incorporating analysis, storage, aggregation, dissemination and deletion.

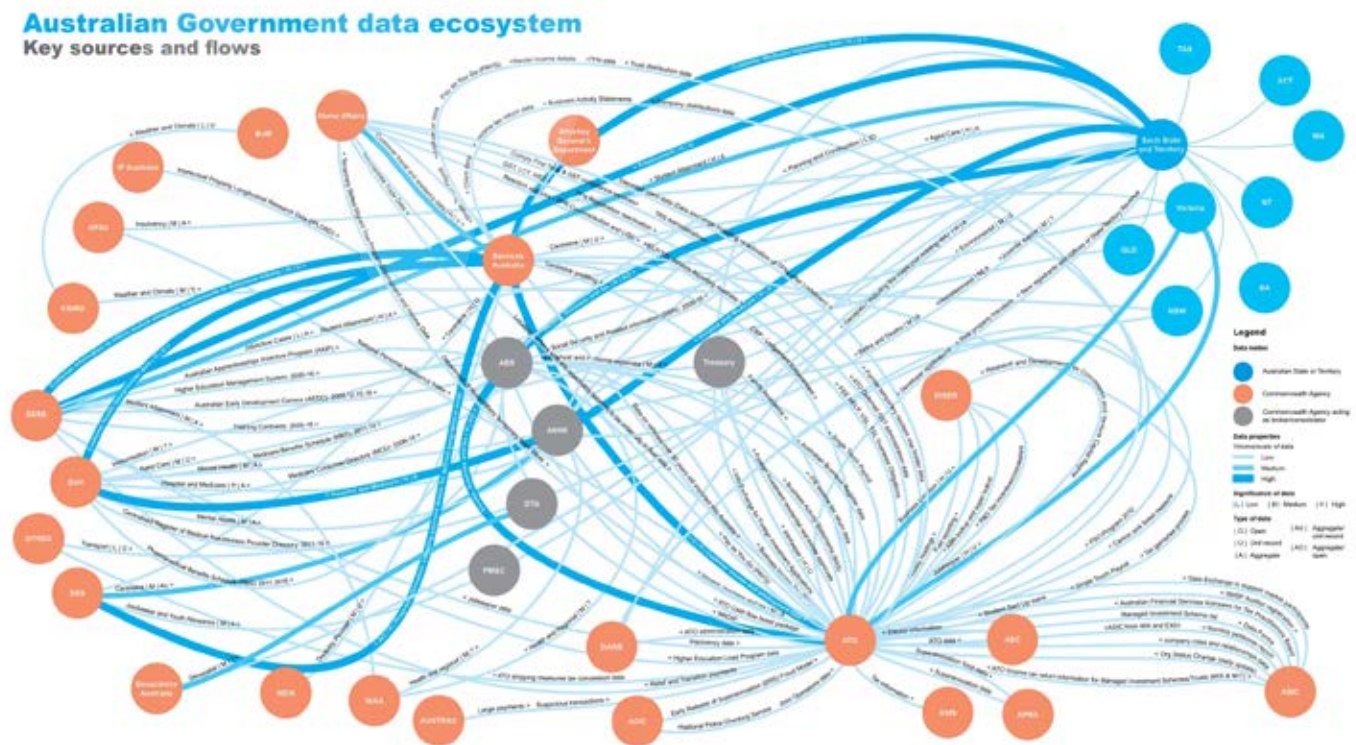


Figure 2. Example real world data sharing network (source: Australian Tax Office)

The life cycle of data can have many twists and turns, and it can involve diverse actors. Figure 2 gives an example of a real world data life cycle, with a view of the Australian Government's data ecosystem. It features many entities, many connections between these entities, multiple jurisdictional regulatory environments, data used and shared in many forms, and many different uses for the data once received. The complexity, unknown overall pathway and unknown implications of data sharing in a real world environment makes many data custodians hesitant to share data. Not sharing is their one guaranteed point of control.

The data flows themselves are of different formats and different levels of sensitivity, contain different levels of personal information and are shared in different volumes. In this example, data is likely being shared in episodic transactions rather than continuous streams, it also likely to be historical curated data rather than real time data, and it will be shared within formalised, bespoke data sharing agreements along with basic metadata. In modern systems, it is also likely to be data in digital format, rather than data shared as paper documents. The combination of these factors and dimensions can be used to describe data sharing methods for data of different inherent sensitivities.

Every time data is transmitted, used or analysed, it changes. That change may be a change in context, a change in the history of the data, a change in who knows the content of the data, or possibly a change in the data itself (for example, from compression or error-prone transmission). At every step and with every action, these changes create metadata that describes the journey so far. If this metadata can be captured, we can begin to address some of the major concerns that 'upstream' data custodians have about 'downstream' use for data shared through different methods.

This critical question is whether all the possible ways of accessing and using data, including sharing and analysis, can be mapped to a finite number of repeatable frameworks that consider:

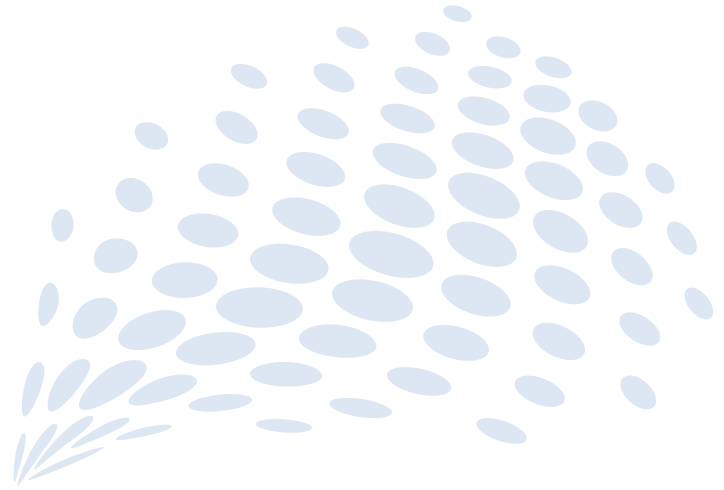
- tracing and assessing the chain of authority to receive and use data
- following the flow and use of data in digital or non-digital formats
- capturing and enhancing the metadata on provenance and consent (or permission) to process and on-share
- capturing and enhancing the metadata on data quality
- following the impact on the data itself as it moves between entities.

Data is often described as either being 'open', meaning it can be accessed by anyone with few (or no) restrictions, or 'closed', meaning that specific restrictions must be placed on the access to the data and the use of the data, including use of insights generated from the data. Very often, entities develop the view that there are few ways of using data between these open or closed frameworks. This white paper will also introduce ways of describing degrees of 'trust' or control that reflect the sensitivities associated with the data itself and the level of technical and domain competence of the intended users of the data. These degrees of trust interact with the governance capabilities the authorising framework required for each different level.

We will start with simple frameworks and slowly work to expand and integrate the key elements into an overarching ecosystem. The goal is to develop practical data sharing frameworks, with identifiable controls, which operate in practical environments.

This paper assumes all analysis is performed using data that has been deidentified, meaning the data has no unique identifiers. It is also assumed that the deidentified data is not subject to any national security classification.

# 02



ULTIMATELY DATA SHARING IS AN ACT OF TRUST, AND TRUST IS EITHER DEVELOPED WITHIN A TRUSTED RELATIONSHIP OR THROUGH DEMONSTRATION OF TRUSTWORTHY CAPABILITY THAT ENCOMPASSES TECHNICAL AND GOVERNANCE CAPABILITY, AS WELL AS AUTHORISATION FRAMEWORKS AND CLARITY OF PURPOSE.



# CONSIDERATIONS FOR DATA SHARING AND USE

It is sometimes conceptually convenient to think of data as having a simple, linear life cycle with a data analysis, or other single use at the centre of that life cycle. As Figure 2 shows, in practice data can be used and reused many times. It can pass through many hands, or algorithms; be used to generate insights; or be combined with other data and insights. Copies of the data and associated metadata and insights can be recombined or archived. The unknown nature of the total data life cycle and the lack of controls that can be activated or scrutinised by data custodians can lead to a culture of hesitancy to share data.

The dilemma often faced by people who want access to data is how to build a trusted data sharing framework in the absence of one. The question of 'Can I have access to your data?' will very often be met with a firm, polite but negative response of 'No', often backed by the statement 'because of the Privacy Act' – the BOTPA reason. This is particularly true if the data is about people.

Ultimately data sharing is an act of trust, and trust is either developed within a trusted relationship or through demonstration of trustworthy capability that encompasses technical and governance capability, as well as authorisation frameworks and clarity of purpose. Recalling Figure 2, data sharing and use is not a single transaction, but parties who share data are a step in what may be a very complex data life cycle. As the number of stages of the life cycle increase, trust between parties becomes increasingly difficult to maintain. Trust between parties can be replaced with controls and scrutiny to ensure appropriate use of data across the stages of the data life cycle.

## 2.1 THE FIVE SAFES MODEL

The Fives Safes model was introduced in the early 2000s to try to address concerns around data sharing and use. It identifies five core risk areas. Several organisations around the world, including the Australian Bureau of Statistics, use the Five Safes framework to help make decisions about effective use of data that is confidential or sensitive. The dimensions of the framework are:

**Safe People** – refers to the knowledge, skills and incentives of the users to store and use the data appropriately. In practice, a basic technical ability is often necessary to understand training or restrictions and avoid inadvertent breaches of confidentiality; an inability to analyse data may lead to frustration and increases incentives to 'share' access with unauthorised people.

**Safe Projects** – refers to the legal, moral and ethical considerations surrounding use of the data. This is often specified in regulations or legislation, typically allowing but limiting data use to some form of 'valid statistical purpose', and with appropriate 'public benefit'. Grey areas might exist when 'exploitation of data' may be acceptable if an overall 'public good' is realised.

**Safe Setting** – refers to the practical controls on the way the data is accessed. At one extreme, researchers may be restricted to using the data in a supervised physical location. At the other extreme, there are no restrictions on data downloaded from the internet. Safe Setting encompasses both the physical environment (such as network access) and procedural arrangements such as the supervision and auditing regimes.

**Safe Data** – refers primarily to the potential for identification in the data. It may also refer to the quality of the data and the conditions under which it was collected (accuracy), the percentage of a population covered (completeness), the number of features included in the data (richness), or the sensitivity of the data.

**Safe Outputs** – refers to the residual risk in publications built from sensitive data.



The Five Safes framework is relatively easy to conceptualise when considering cases of 'extremely' safe, although it does not unambiguously define what this is. An extremely safe environment may involve researchers who have had background checks, projects that have ethics approval and rigorous vetting of outputs from that data environment. Best practice may be established for such frameworks, but none of these measures is possible to describe in unambiguous terms as they all involve judgement.

Figure 3 shows the dimensions of an adapted Five Safes framework taken from the 2018 ACS Technical White Paper *Privacy in Data Sharing: A Guide for Business and Government*.<sup>4</sup> The adapted model explores different, quantifiable levels of 'safe' for each of People, Projects, Setting, Data and Outputs, as well as how these different safe levels could interact in different situations. It also tries to place these five risk dimensions into a larger context that considers more of the data life cycle and the consequences of use of insights generated from data analysis.

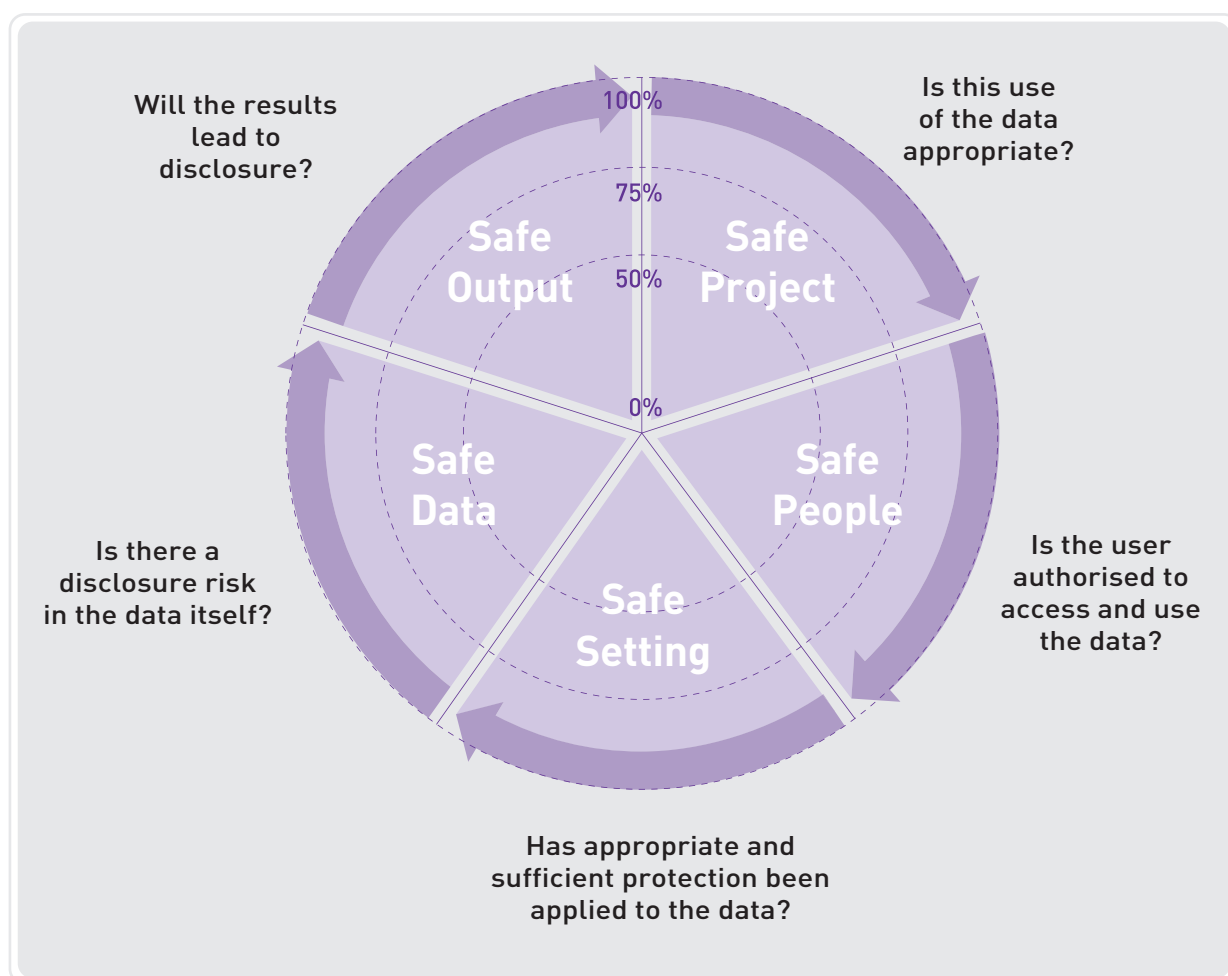



Figure 3. Modified Five Safes framework

One of the great challenges of this model is the interaction between the risk dimensions. The Project or purpose can impact People, Data, Setting and Output; and Data can impact People, Setting and Output. The ability to work out which risk dimensions are fixed and which need to be adapted in response to these risk frameworks makes the approach an iterative process at best.

## A REAL WORLD EXAMPLE: THE FIVE SAFES

---

 **A local council** wants to develop a water temperature heatmap for an environmentally sensitive lake. Data measurement will be performed through a network of water-based sensors that are sparsely spread, many of them located near to isolated lakeshore homes. The data therefore has the potential to reveal information about occupancy of the homes or activities taking place within the homes. Some of the basic aspects to consider are:

- **Project (fixed):** the merits of the project may well provide a strong motivation to proceed
- **Data (fixed):** the location of sensors near isolated homes means that the data is highly likely to contain personal information
- **People (variable):** a high likelihood of personal information in data means protections must be put in place to limit the people who access the data or carry out the project
- **Setting (variable):** a high likelihood of personal information in data means protections need to be put in place to limit access to data and outputs of analysis
- **Outputs (variable):** the project requires only aggregated output so the results of analysis can be treated to reduce the level of personal information before release.

In the lake temperature example (see 'A real world example: the Five Safes'), the high-level output may be aggregated in a temporal or spatial sense to reduce the reidentification risk and reduce the amount of information released. The questions the Five Safes models leave unanswered include:

- What happens to the Output?
- Are there unintended consequences associated with the use of that Output?
- Who is responsible for any harms that arise from use of that Output?
- Could someone still be reidentified from the Output after the analysis?
- Do people who use the Output have sufficient context or expert knowledge to correctly interpret the Output?
- What biases in the data or the analysis prevent the Output from being generalised beyond the scope of the Project?
- Who is responsible when harms arise from the use of the Output or release of the data?

While it is a reasonable set of considerations for an individual project in isolation, the model fails to address much larger concern of 'safe for whom?' It also focuses on just one stage in the data life cycle, which is when data is to be analysed, without real consideration for the journey of the data to that point, or the stages after analysis.

## 2.2 'WHY' MATTERS

In surveys of data custodians and the general public, the intended use of the data was frequently identified as a very significant factor when determining the risk framework for data sharing and use.

The summary from the *Australian Community Attitudes to Privacy Survey 2020* from the Office of the Australian Information Commissioner states:<sup>5</sup>

*Our comfort with certain data practices depends on the type of information collected, the purpose behind it, and the level of trust in the organisation involved. Australians appear more comfortable with data practices where the purpose is clearly understood – for example, law enforcement using facial recognition and video surveillance to identify suspects.*

Ethics committees will often ask the 'why' question related to human research projects, but ethics committees are not used in all people-centred data projects.

A formal definition of 'data use' and 'use case' would bring clarity about what is intended for the data and what can be done with the results. Work is underway within standards bodies to try to formalise use cases for data (ISO/IEC/JTC 1 SC 32/WG 6).<sup>6</sup> Very often, however, a use case is described in terms of:

- who wants access to the data
- why they want to access data
- consideration of the level of personal information in data
- consideration of aspects of sensitivity of the data and the results of analysis
- concerns about the level of granularity of access of data
- concerns related to the use of insights and decisions generated from analysing data.

The sensitivity of any dataset relates to the level of personal information, the possible harms arising from the use of the data, and the concerns around unintended consequences of data availability. Depending on the sensitivity of the data and how likely an individual is to being identified in the data, being able to explain 'who' and 'why' is becoming increasingly important. The safeguards required to be put in place also increase with sensitivity, levels of personal information being used and the risk of reidentification of individuals.

## 2.3 DATA SHARING IS A FORM OF DATA USE

The term 'data sharing' is often used together with 'data use'; however, sharing is actually a subset of use. Data may be used in many ways, as seen below:

- for analysis: tallying, visualising, describing, diagnosing, showing relationships, predicting or modelling
- for event detection: monitoring or alerting
- to trigger actions: based on thresholds or as a consequence of event detection
- for historical record: observing, recording and archiving
- for storage: files for photos, videos, programs presentations or spreadsheets

<sup>5</sup> Available at <https://www.oaic.gov.au/engage-with-us/research/australian-community-attitudes-to-privacy-survey-2020-landing-page/2020-australian-community-attitudes-to-privacy-survey/>.

<sup>6</sup> See [https://www.iec.ch/ords/f?p=103:7:512258326175321:::FSP\\_ORG\\_ID,FSP\\_LANG\\_ID:3406,25](https://www.iec.ch/ords/f?p=103:7:512258326175321:::FSP_ORG_ID,FSP_LANG_ID:3406,25).

- to create data products: copies, aggregates, subsets, perturbed versions, insights or outputs in other formats (for example, printing of digital documents)
- in transmission: broadcasting, transferring or connecting
- for deletion: destruction or rendering data inaccessible to the current entity.

These different uses imply different operations on data including static analysis, real time use, observing or alerting, passive storage, or interactive two-way use. These myriad possibilities means that many different frameworks may be relevant. The roles of entities (people, devices, systems) change depending on where they sit in the frameworks and which phase of the data life cycle they are operating on.

When data is shared, a current data holder transmits data or data products to the next entity in some form. That next entity then uses the data for their purpose within their authorising framework and may then on-share data and data products including insights. Given all the all data can be used and shared, issues of trust stretched across long chains of entities become increasingly difficult to manage. We instead turn to controls that may be put in place along sections of the life cycle.

## 2.4 MODES OF SHARING DATA

Data sharing and use can involve more than taking a copy of data and using or analysing without oversight. Different degrees of access can be provided, from none (most extreme), allowing access to prepared data products (including insights or aggregations), limited analysis access, to providing a copy of the data without restriction (see Figure 4). These various modes of sharing allow increasing (or decreasing) levels of control depending on the sensitivities or risks associated with the data.

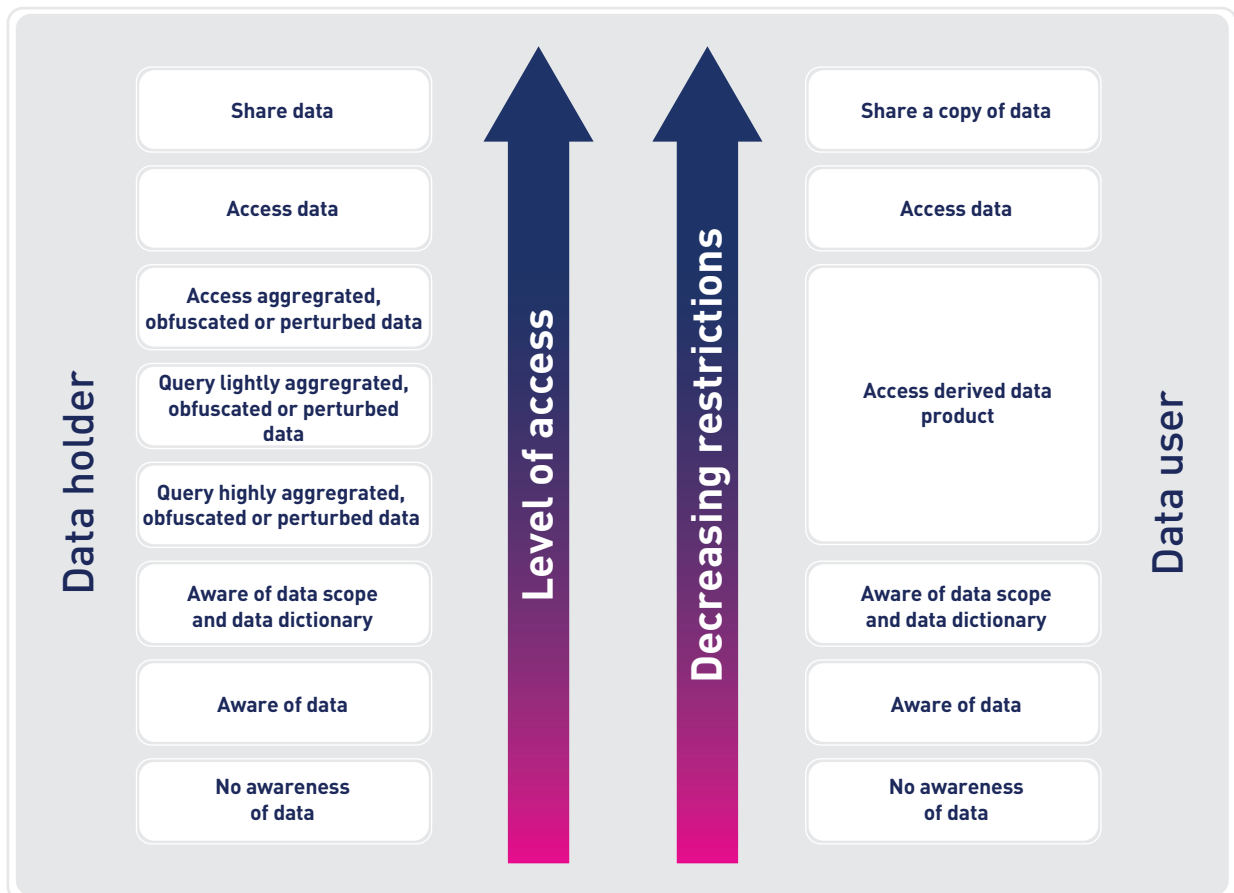


Figure 4. Framework for data sharing and use

Further, a number of general scenarios for data sharing and use can be used as an interaction between an entity that holds data (holder) and an entity wishing to use data (user), as shown in Figure 5:

- **Scenario 1:** holder shares actual data or data products with user, but user cannot modify data
- **Scenario 2:** user can query data and gain insights but not directly access data (vault model)
- **Scenario 3:** holder shares actual data or data products with user, and user can modify data
- **Scenario 4:** holder and user employ the services of a trusted third party to process and analyse the data.

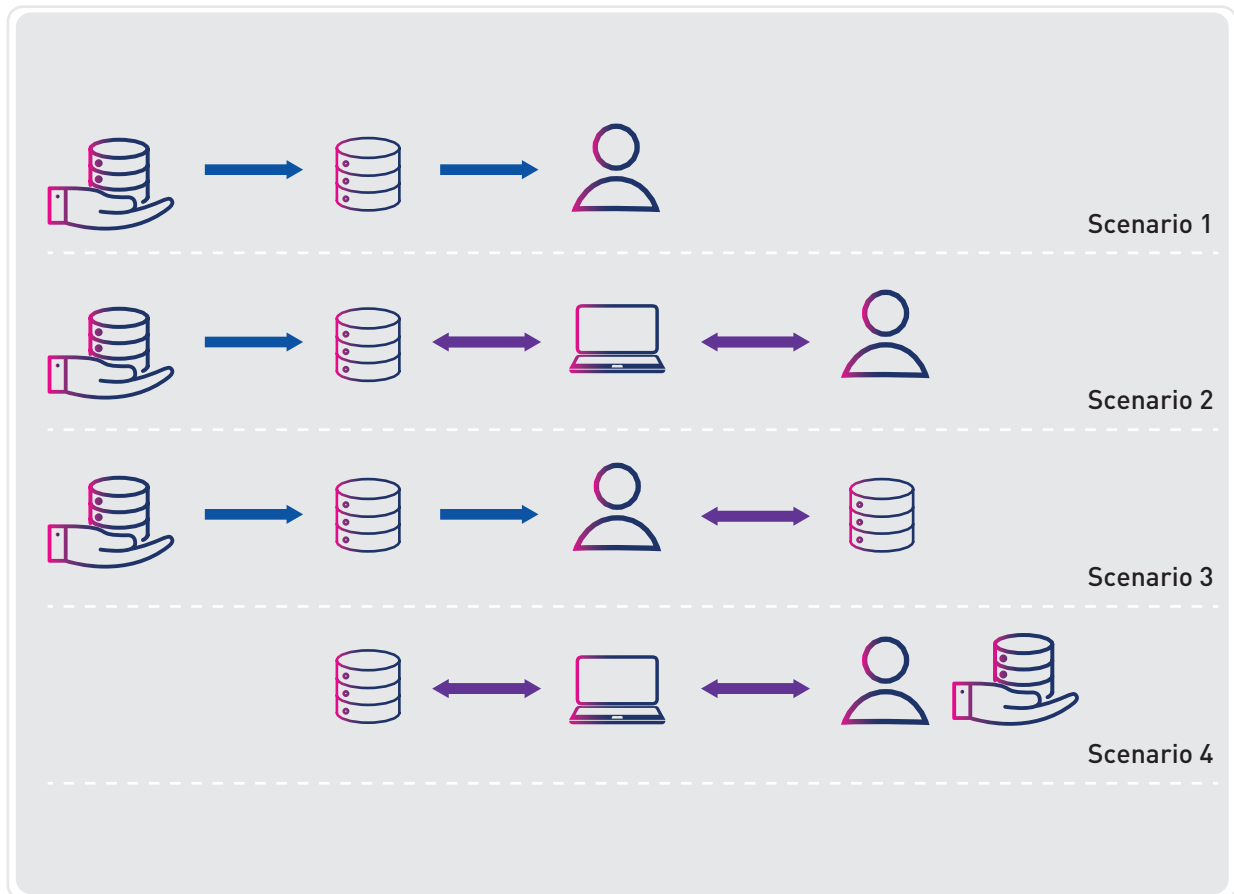
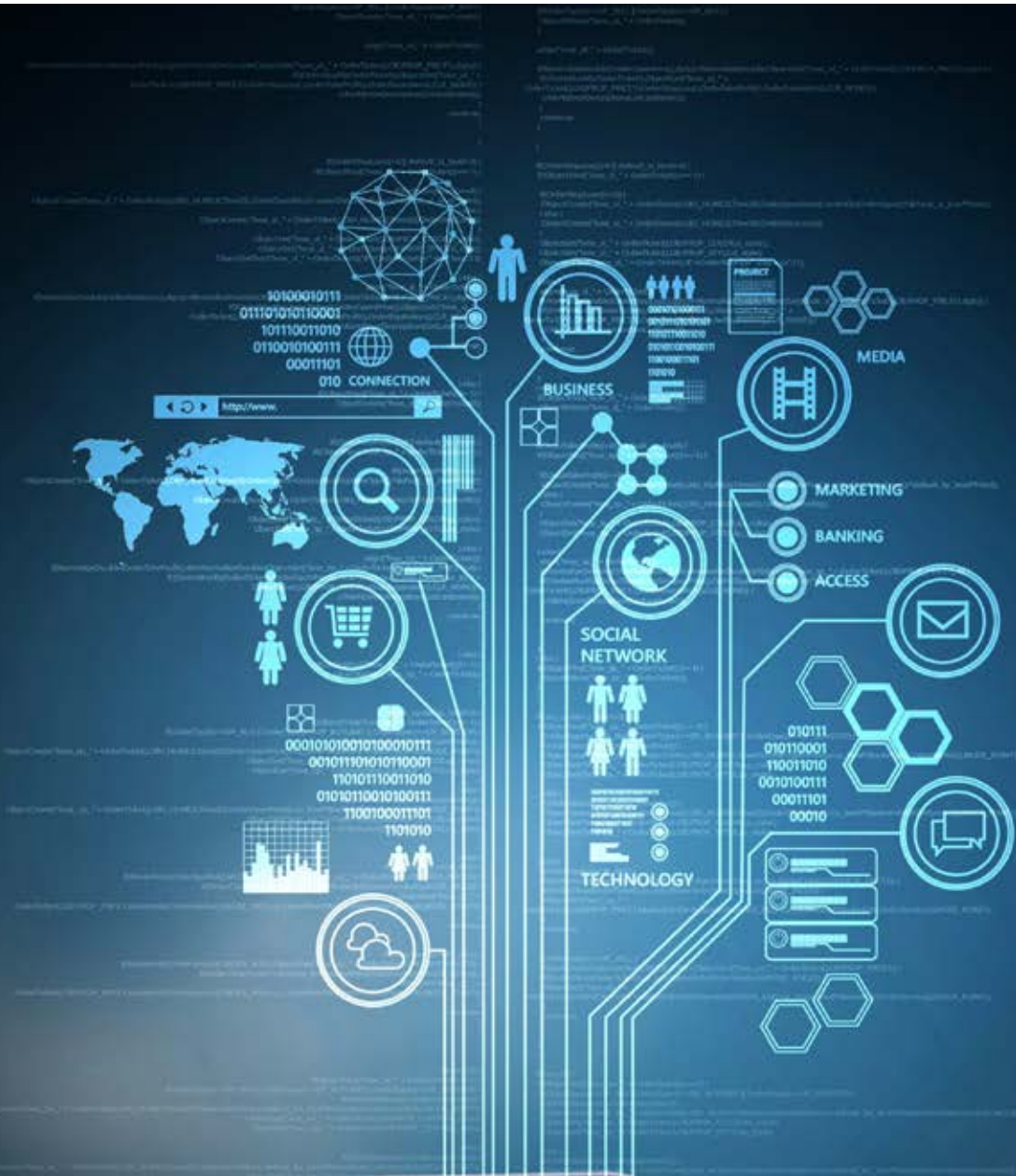


Figure 5. Scenarios for sharing or accessing data

These broad scenarios can be further broken down, but for the sake of simplicity, we will use these to describe modes of sharing.

The challenge for a data sharing use case is to determine which dimensions are set by the nature of the problem and which need to be adjusted in response to the nature of the problem.

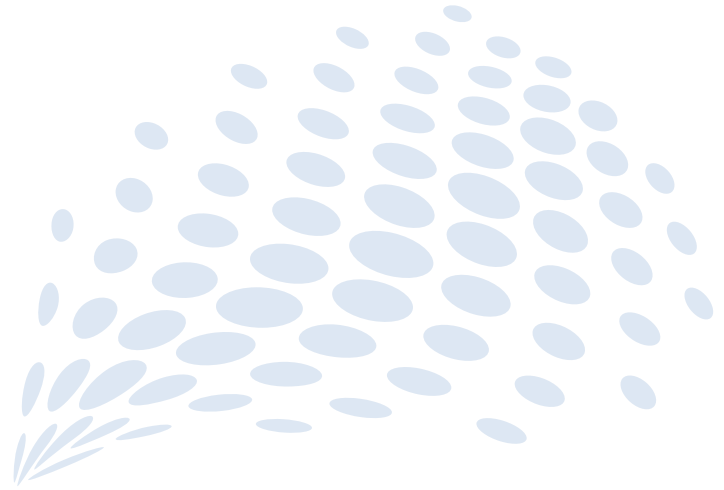




THE SENSITIVITY OF ANY DATASET RELATES TO THE LEVEL OF PERSONAL INFORMATION, THE POSSIBLE HARMS ARISING FROM THE USE OF THE DATA, AND THE CONCERNS AROUND UNINTENDED CONSEQUENCES OF DATA AVAILABILITY. DEPENDING ON THE SENSITIVITY OF THE DATA AND HOW LIKELY AN INDIVIDUAL IS TO BEING IDENTIFIED IN THE DATA, BEING ABLE TO EXPLAIN 'WHO' AND 'WHY' IS BECOMING INCREASINGLY IMPORTANT.



# 03



# SEPARATING SENSITIVITY FROM PERSONAL INFORMATION

While concerns related to privacy are often the reason for restrictions on data sharing, there are many other concerns related to unintended consequences of the use of data. These include concerns about:

- release of data about vulnerable individuals
- reidentification of individuals
- loss of exclusive access to insights from data
- appropriate use of the data
- appropriate use of insights gained from data
- unexpected or embarrassing results found from analysis of data
- recipients' lack of expertise to analyse or interpret the data
- data age and quality
- use of data without the contextual knowledge of its collection or data quality.

These concerns, along with privacy-related issues, are shown in Figure 6. Conceptually, it is possible to think of 'high', 'medium' or 'low' levels of concern around each of these sensitivities. In each case, mitigations of different strength may be applied to data itself, to the governance framework in place, to the requirement for technical and domain expertise when data is used (including for analysis) or to prohibitions for use of data products, including secondary use.

In many cases, more than one sensitivity will exist, in which case multiple mitigations will need to be employed. It is important to note that these mitigations must apply across the entire data life cycle to be effective. This may have the effect of limiting the extent of that life cycle.

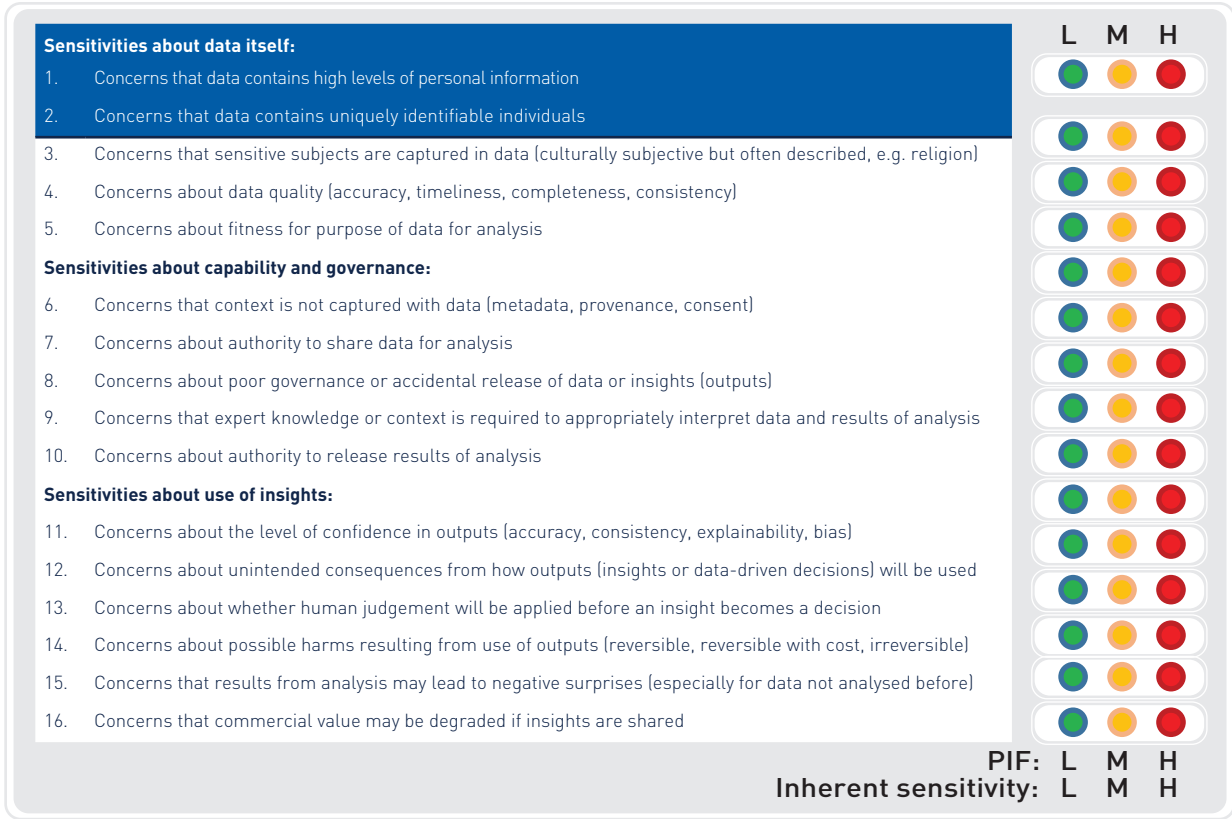


Figure 6. Examples of sensitivities relevant to data sharing and use



## A REAL WORLD EXAMPLE: OUT-OF-HOME CARE

As an example, consider the reform of out-of-home care (OOHC) in New South Wales, Australia (see Figure 7).<sup>7</sup> The OOHC scheme works with children who have been identified as being at risk of significant harm, placing such children into protective environments. The scheme reform is underpinned by the creation of longitudinal datasets linking data from many government agencies on an individual (child centric) basis. All data is deidentified before linkage. Nonetheless, concerns persist about privacy and sensitivity about the use of data, the nature of the project and use of outputs.

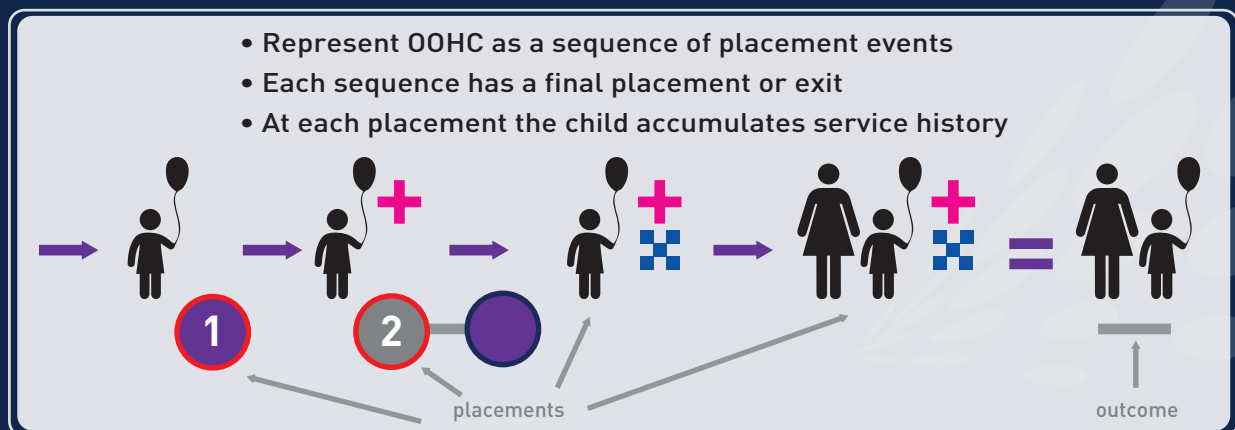


Figure 7. Motivating example, OOHC reform

In the OOHC example (see 'A real world example: out-of-home care'), concerns identified included:

- What if a machine/algorithm generates insights (outputs)? Can the results be trusted?
- Who can access this data?
- Who are outputs shared with?
- What are the consequences of sharing or using these insights (outputs)? Can this make things worse (outcomes)?
- Does linked deidentified data actually contain sufficient personal information to reasonably identify individuals?
- Could poor data quality lead to inaccurate insights?
- Is there a human-in-the-loop so that a machine or an algorithm is not empowered to automatically act on the insights generated?
- Is there appropriate access and authorisation to data and analytical insights?

The sensitive nature of the reform program, the potential impact on children and families, and the sensitive subject of the very rich datasets trigger every almost concern identified in Figure 6. Strong governance is required along with a particular authorising framework to access data and insights. The people who access the data must have significant technical capability, domain expertise and an understanding of appropriate use within the authorising framework and governance environment.

<sup>7</sup> For more details of data assets used for reform, see <https://www.theirfuturesmatter.nsw.gov.au/investment-approach/tfm-human-services-data-set>.

### 3.1 PERSONAL INFORMATION (PI) AND PERSONALLY IDENTIFIABLE INFORMATION (PII)

The concepts of personal information versus personally identifiable information are not clearly differentiated in regulatory frameworks. The term 'personal information' is typically used very broadly and is described differently in different parts of the world. The website of the Office of the Australian Information Commissioner states:<sup>8</sup>

*Personal information is information that identifies or could reasonably identify an individual. The Privacy Act 1988 and the FOI Act define 'personal information' in the same way:*

*Personal information means information or an opinion about an identifiable individual, or an individual who is reasonably identifiable:*

- a. whether the information or opinion is true or not and*
- b. whether the information or opinion is recorded in material form or not*

While not uniquely identifiable, eye colour, hair colour and shoe size are all PI (information about an identifiable person). The threshold question is then: When is the person identifiable?

In general, it is expected that the level of PI in a linked, deidentified dataset will increase as more people-centred datasets are linked. Conceptually shown in Figure 8, as more datasets containing PI are linked, a point may be reached where an individual is personally identifiable, or 'reasonably' identifiable. The dataset is then considered to have PII. The epsilon in this figure is an indication of the difference represented by the gap before the 'reasonable' threshold is met.

This raises the question: Can this threshold of PII and the definition of 'reasonable' be quantified? The answer depends on context.

Some of the dimensions of this context that matter are:

1. Can an individual in a dataset (rows of people and columns of features) be identified as unique, based on a single feature or combinations of features?
2. Can the unique row be identified in other datasets and so link information between datasets (for example, unidentified online browsing records)?
3. Can the unique row of features be mapped to an actual person or small group of people, based on access to other data?
4. Could someone observing the unique row spontaneously identify the actual person from the unique feature or feature combination, based on their own knowledge?
5. Is an individual known to be in a dataset, and could their row be identified based on a subset of features?
6. Is an individual known to be in a dataset, and could knowledge of the nature of the dataset (for example, patients with cancer) lead to inferred information about an individual?

A similar logic can apply to a small number of rows with the same feature values. Being able to narrow down to a small number of identical rows may introduce some uncertainty, but many of the contextual considerations above remain relevant.

These contextual considerations require different controls for different environments to preserve privacy and avoid PI becoming PII. This includes screening who has access to data, controlling access to linkable datasets and providing prohibitions on use (and secondary use) of data and data products.

<sup>8</sup> Available at <https://www.oaic.gov.au/freedom-of-information/frequently-asked-questions/what-is-personal-information-and-how-does-it-interact-with-the-freedom-of-information-act-1982/>.



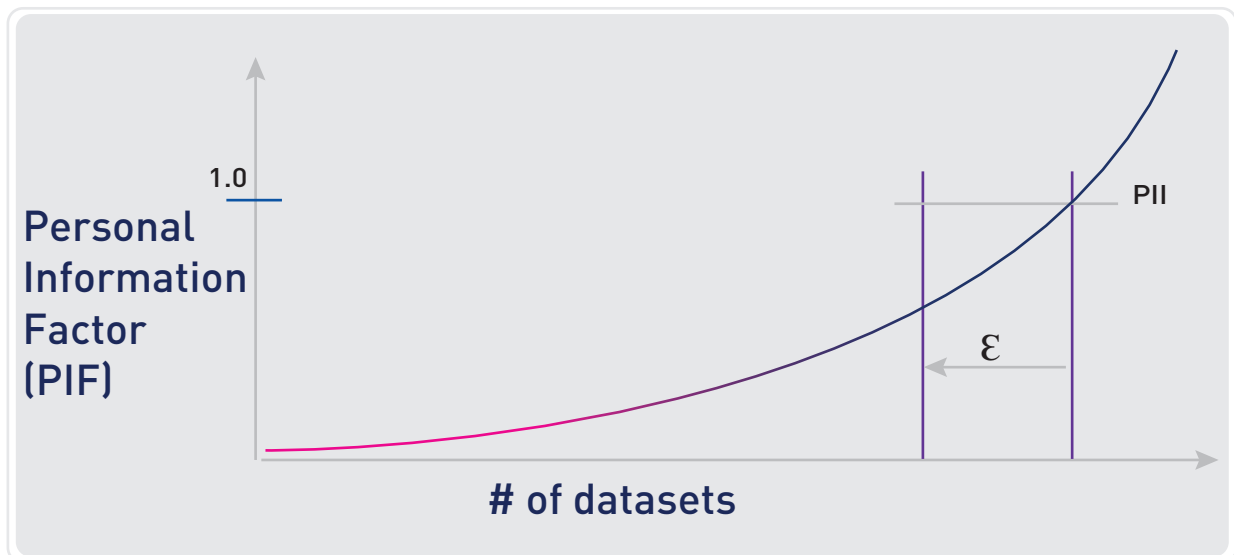


Figure 8. Conceptualisation of a normalised Personal Information Factor (PIF) and the threshold point of reaching personally identifiable information (PII)

### 3.2 A COMMENT ON THE 2019 DEFINITION OF PERSONAL INFORMATION FACTOR (PIF)

The 2019 ACS Technical White Paper *Privacy-Preserving Data Sharing Frameworks* proposed a way to calculate an important parameter, a 'Personal Information Factor' (PIF), which was the measure of information gain an 'attacker' would gain for an individual known to be in a dataset (rows of individuals and columns of features). The information gained for any given feature for the known individual was referred to as the 'cell information gain' (CIG). The sum of all of the CIGs for a row became the 'row information gain' (RIG). The PIF for the dataset was defined to be the highest RIG within the dataset when normalised by the number of rows that were identical with that RIG. This meant that, if one row was unique and had the highest RIG, it determined the PIF for the dataset.

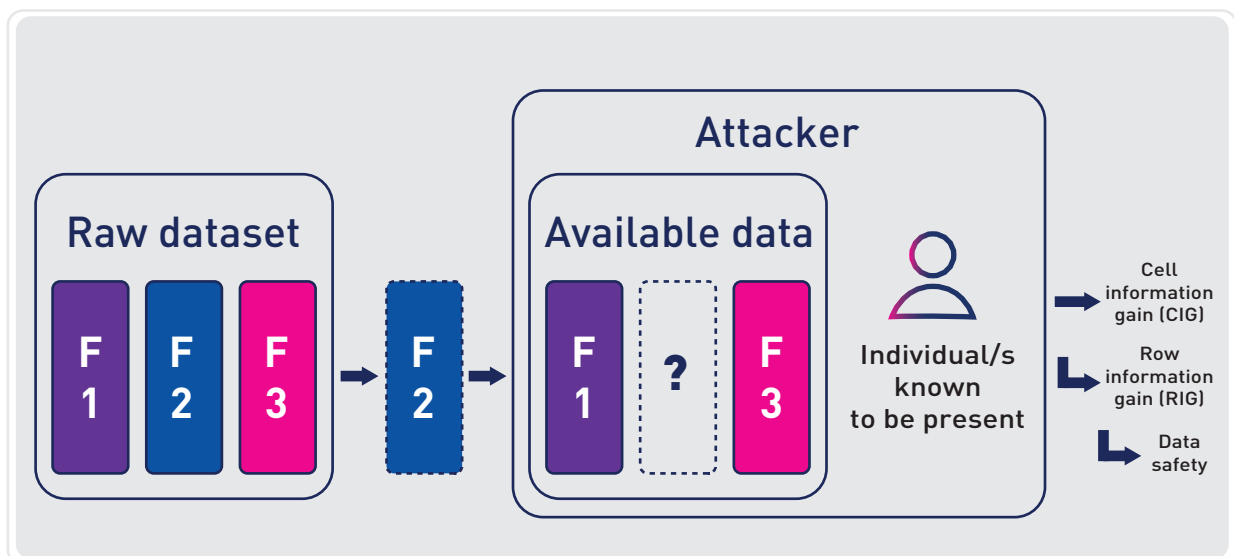


Figure 9. Conceptual model for information gain by an attacker

## EXCERPT FROM THE 2019 WHITE PAPER PRIVACY-PRESERVING DATA SHARING FRAMEWORKS

The PIF for the dataset is driven by both the minimum identifiable cohort size (MICS) and the amount of information that would be revealed if individuals in this cohort were reidentified. The definition of PIF is still a work in progress, but the current working definition is given as:

$$PIF = \text{maximum of } (RIG_{(x)} / (MICS \text{ at } RIG_{(x)}))$$

At any given RIG threshold, the MICS at that value is the smallest number of rows with the same column values. For example, if the number of rows with a RIG at  $RIG_{max}$  is 1, then the PIF is equal to  $RIG_{max}$ . If the number of rows with a RIG of  $RIG_{max}$  is 2, and there are no other unique rows in the dataset, then the PIF is  $RIG_{max} / 2$ . If there is a unique row at a threshold RIG less than  $RIG_{max}$  (for example,  $RIG_{(x)}$ ) and the number of rows at is  $RIG_{max}$  is 2, then the PIF is  $RIG_{(x)}$  provided  $RIG_{(x)}$  is greater than  $RIG_{max} / 2$ .

This 2019 definition of the PIF for a dataset has a number of limitations in that:

1. The information gain is measured against a priori knowledge of the feature. For example, if it was assumed based on real world knowledge that there are four equally likely values for a particular feature, the a priori knowledge is that any feature value has 25% chance of being the actual value. Once the actual value is known, there is 100% certainty, representing 2 bits of information gain for that feature. The question remains: How many bits are required to uniquely identify an individual as being the person in the dataset? We will return to this below.
2. It does not allow for mutual information between features, meaning each feature is considered independent of all others and no feature values can be derived or even narrowed down from others. So again, the sum of the CIGs for a row is an upper bound of total information gain.
3. It assumes the attacker knows that a person of interest is in the dataset, and none of the features of the dataset are already known and could be used to identify which row the person is in. Any known features used to identify the exact row would not lead to information gain by the attacker, again leading to this PIF being an upper bound.
4. It is not normalised or bounded, with the consequence that the PIF of 1.0 imagined in Figure 8 has not yet been achieved. The PIF of a dataset can be many bits of information gain for an attacker who knows an individual is in a dataset. It does not give an absolute measure of identifiability.

Nonetheless, the 2019 definition of PIF for a dataset has been useful when considering relative values of PIF. If a dataset to be shared is historically considered to be safe to share within certain trusted (or controlled) environments, then future datasets can be measured against the historically satisfactory level and either maintained or reduced by data protection means such as breaking tables apart or suppression or aggregation of data into lower specificity ranges.

### 3.3 HOW MANY BITS DO YOU NEED TO UNIQUELY IDENTIFY AN INDIVIDUAL IN A POPULATION?

One of the fundamental questions posed by the original conception of the PIF was how many data deidentified sets could be linked before an individual was uniquely (or reasonably) identifiable.

Various attempts have been made to quantify the answer to this question, ranging from the famous '33 bits' reported in 2010 as the number of bits required to hold a unique register for every person on the planet at the time,<sup>9</sup> to more sophisticated recent attempts.<sup>10</sup> The answer to this question comes from an understanding of the population within which this individual sits and what else is known about the individual.

As an example, if the state of New South Wales has approximately 8 million inhabitants, then 23 bits would be sufficient to create a record system that could uniquely index every person. However, that is different from the ability to identify 'John Smith' from a set of data.

If John Smith had a unique identifier (such as a driver's licence or a unique height measured in centimetres) and this was recorded in a dataset, and it was known John lived in NSW and there was a way to connect this unique identifier to John, then this dataset undoubtedly contains personally identifiable information. If the distribution of unique identifiers was uniform, then John's unique identifier would be calculated to have:

$$\text{Log}_2(8,000,000/1) = 22.9 \text{ bits of information using the 2019 PIF model}$$

The difference between uniquely identifying John as a deidentified row in a dataset and connecting to the actual John Smith remains the mapping between identifiers (or rows) and people. That connection also needs to include the information that John resides in NSW (such as recorded on his NSW driver's licence). These unique identifiers to people mappings must be separately controlled to avoid the rows in the data from being personally identifiable. The possibility of connecting such mappings to datasets (even if not actively done) has been considered to be a breach of privacy legislation in NSW with Opal travel cards.<sup>11</sup>

If, however, there were 256 possible height measures for the dataset, then the number of bits gained by knowing John's height in centimetres depends on the distribution of height within the population within these 256 values. If John is of very unusual height, then the number of bits of information gained by learning John's height is approximately:

$$\text{Log}_2(8,000,000 / \text{number of occurrences of John's height bracket}) < 22.9 \text{ bits (unless John is unique)}$$

A mapping is still needed to connect a rare (or unique) height to John Smith. That mapping may be a formal population height mapping, or someone who sees a dataset and knows John has a unique height. This leads to the complicating factor of what individuals know about members of a population, how familiar they are within individuals in that dataset, and how their access may reasonably be controlled. In this example, however, it is possible to see that John could be identified by an individual with knowledge of John with (far) less than 23 bits of information.

The example can be extended as more fields are added. In each case, if John's feature values are increasingly less unique, there is a more ambiguity as to whether one particular row is John. Nonetheless, contextual information held by a person who knows John and the combination of not-unique feature values can quickly lead to a reasonable assumption that a particular row refers to John Smith.

9 See *The Wall Street Journal* article at [https://www.wsj.com/articles/BL-DGB-16975?reflink=desktopwebshare\\_permalink](https://www.wsj.com/articles/BL-DGB-16975?reflink=desktopwebshare_permalink).

10 Sweeney L (2000) 'Simple Demographics Often Identify People Uniquely', *Data Privacy Working Paper 3*, Carnegie Mellon University, Pittsburgh.

11 See *Waters v Transport for NSW* [2018] NSWCATAD 40. Available at <https://www.caselaw.nsw.gov.au/decision/5a8351f1e4b074a7c6e1c492#amendments>.

Throughout this series of white papers, we have assumed that if a row is unique (in its combination of feature values), then it can be mapped to an individual with sufficient effort or contextual knowledge. If a row is unique, the PIF approach describes the worst-case information release for an individual. In this sense, it remains a useful measure.

### 3.4 TIME, SPACE, PERSONAL FEATURES AND RELATIONSHIP FEATURES

When thinking about how a person could be identified, it is not just personal features that could be used. Information about where a person was, when they were there and who (or what) they interacted with could all be used in combination to create a unique record in a dataset (and so, with effort, be used to identify an individual).

These dimensions of time, space, personal features, and relationships are not completely independent (one feature value may be inferable from others); however, they do represent quite different dimensions when considering how to protect data. Aggregation, suppression or perturbation can be applied equally to the entire dataset, or applied with different levels to temporal, spatial, personal or relationship features. The intention is to maintain utility of the dataset in one or more of these feature domains while preferentially protecting features in the other domain (and so reducing utility of the data in these domains). Figure 10 attempts to show how cohorts can be defined based on these different dimensions.

The PIF for this dataset is again determined by the size of the minimum identifiable cohort size (MICS) and how much information is gained if an attacker knew an individual was in this cohort. Mutual information between temporal, spatial, personal features and relationships is still not considered, so the PIF remains an upper bound.

Developing standard aggregation, suppression or perturbation approaches in each of these domains would assist when analysing data from different sources. It is certainly possible to imagine standard protection approaches for numerical features (such as latitude and longitude or age in days), but more challenging for categorical features (eye colour or hair colour). An example of how this may be done is given below in 'Example of PIF in action: COVID-19 data release considerations'.

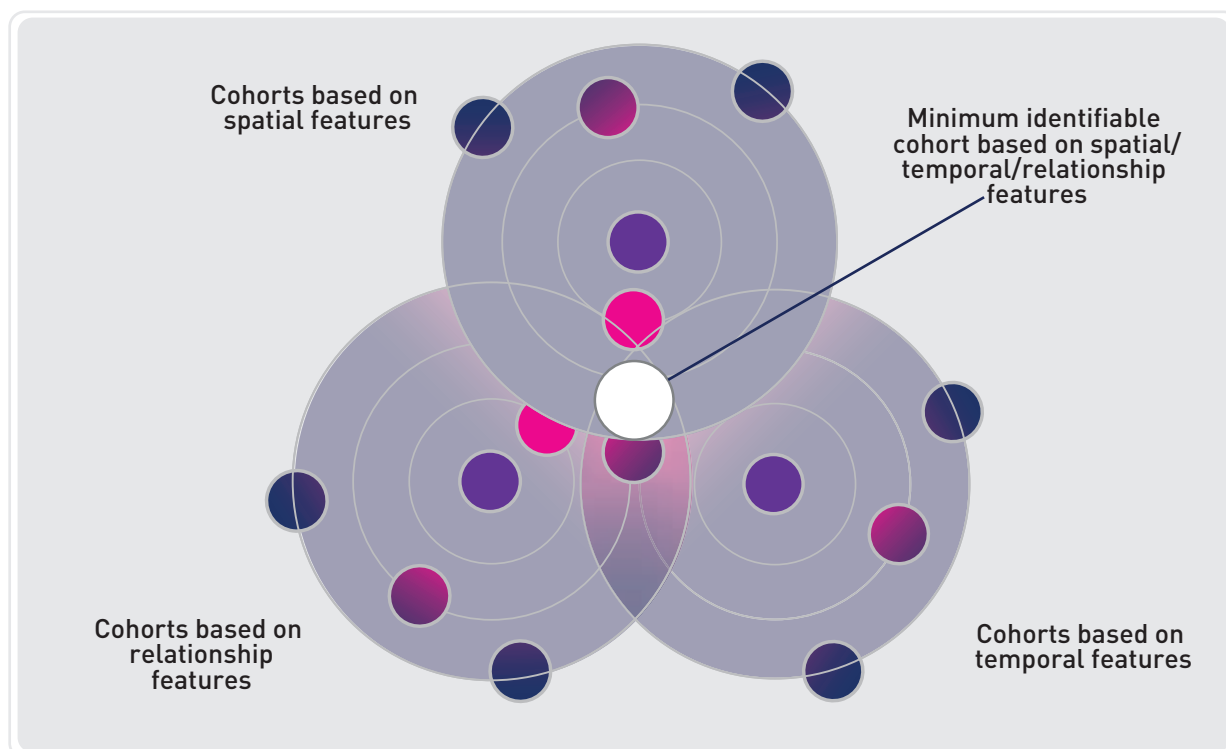


Figure 10. Cohort identified by temporal, spatial and relationship features

## EXAMPLE OF PIF IN ACTION: COVID-19 DATA RELEASE CONSIDERATIONS

**In 2020**, the NSW Government committed to release information about the developing number of confirmed COVID cases on a daily basis at postcode level. Issues of the level of personal information and the sensitivity of the data were of foremost concern. This was balanced with the strong desire for the public to be informed about the developing COVID situation. A complete set of possible fields for release was collated from NSW Health (raw data) and then tested for the total amount of information that would be revealed if released.

A series of conversations was undertaken regarding the balance of data being released 'in the public interest' versus data that was merely 'of interest to the public'. The government also considered the risks associated with reidentification of individuals and how much information could be associated with an individual who was identified. A Personal Information Factor (PIF) was assessed to determine an upper limit measure of the worst-case information that would be released if an individual were identified.

This tool and measurement process was used to design additional protections (principally disconnecting temporal and spatial features, as well as aggregation) for the data before releasing it as open data. The data in the reduced feature tables was analysed each day to ensure the PIF is reduced to an agreed level before release. The dataset was in the form of rows (unique individual) and columns (features related to that individual). The data released was also used to create spatial maps for those who do not want to access the data directly.

Table 1 has records back to first recorded COVID cases in NSW. Tables 2, 3 and 4 commence from a later date to prevent relinkage of these tables.

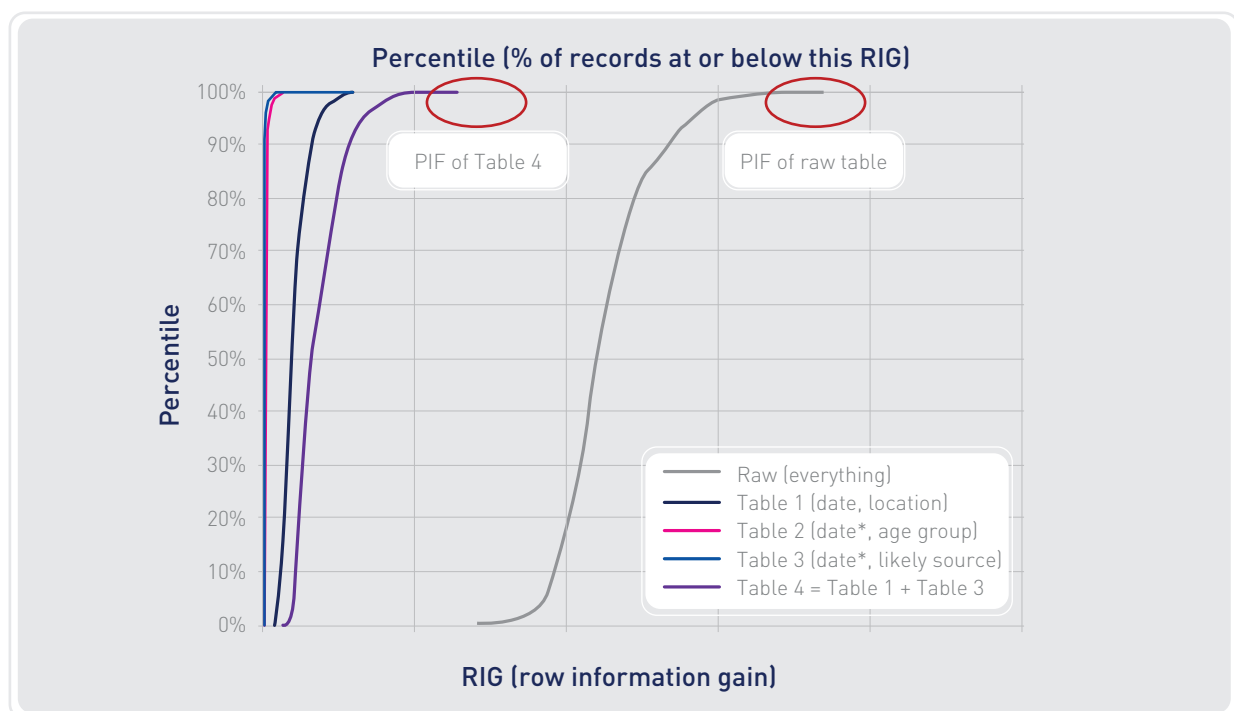
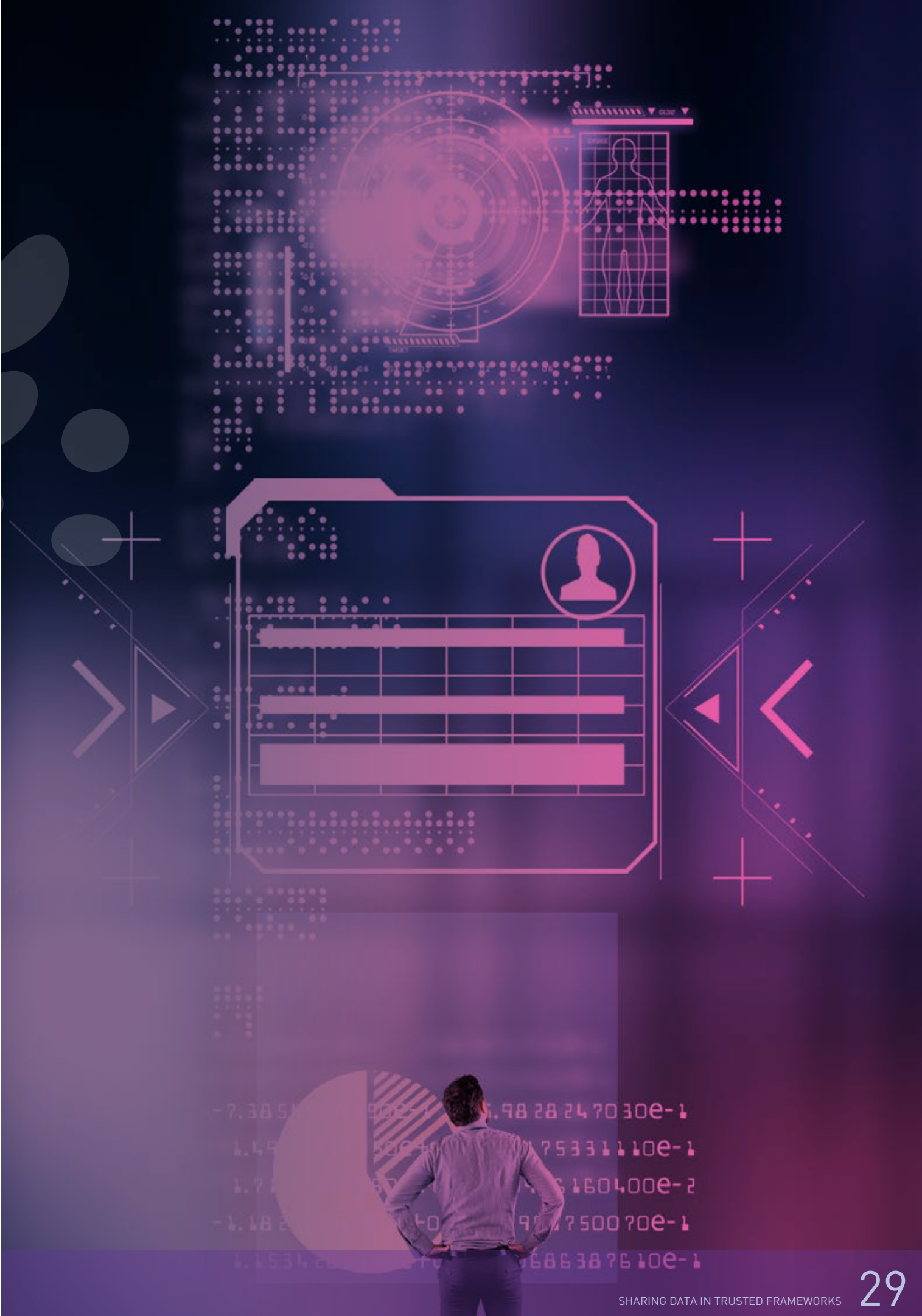
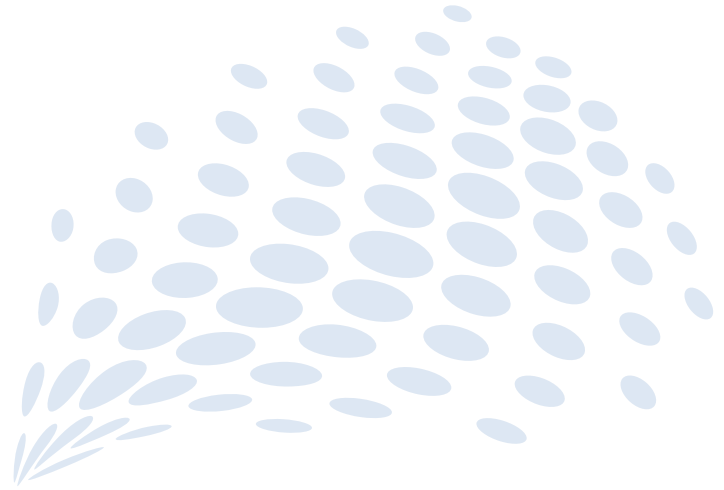


Figure 11. Example of reducing the PIF of a raw table by creating subtables





# 04



DATA THAT IS FINALLY 'USED' MAY WELL BE DIFFERENT FROM THE DATA THAT WAS ORIGINALLY CREATED/COLLECTED.



# CONSIDERING THE WHOLE DATA LIFE CYCLE – QUALITY, METADATA AND HUMANS

Figure 2 shows that, once created, data and data products may be used (and reused) many times in many forms. This makes identifying a simple series of controls that are effective over such an elongated life cycle a significant challenge.

Figure 12 shows a simplified data life cycle that will allow us to explore controls that may be considered from the point of data creation to collection, storage and then use by the receiving entity. This 'use' may be analysis of the data. The data or data products are then shared and finally archived. The simple life cycle can be expanded at any phase to more explicitly show the range of activities that take place during that phase.

Along the way, the original data is assumed to be modified from its original form – from when it was captured ( $D_1$ ), transmitted ( $D_2$ ), stored ( $D_3$ ), used (as  $D_4$ ) and then stored (as  $D_5$ ). The types of factors that can impact data during these stages include:

- subsampling of raw data or reduction in data precision before transmission
- loss of data, lossy data compression<sup>12</sup> or data corruption during transmission
- loss of data, lossy data compression or data corruption during storage
- lossy data decompression or data corruption when importing data, removal of low-quality data before use
- loss of data, imperfect data compression or data corruption during archiving.

As a consequence, the data that is finally 'used' may well be different from the data that was originally created/collected. In modern digital information management systems, data loss and corruption are rare. However, if data is captured from a camera on a drone, transmitted wirelessly and then compressed on storage before analysis, many more data loss or data quality events may occur. Once data is used, an incomplete dataset may then ultimately be archived.

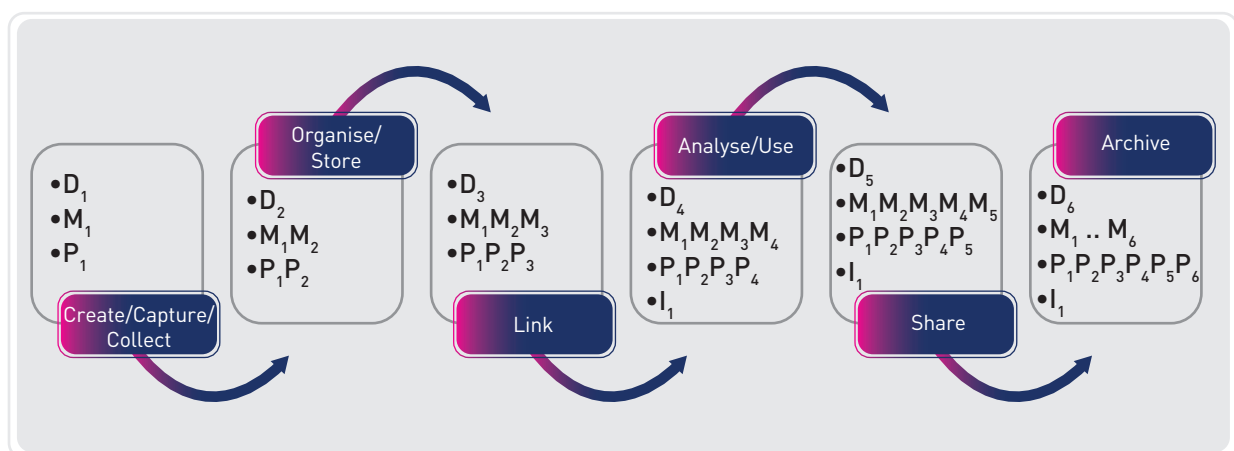


Figure 12. A simplified data life cycle

12 In information technology, lossy compression or irreversible compression is the class of data encoding methods that uses inexact approximations and partial data discarding to represent the content.

As data moves along the different stages of the life cycle, metadata can also be collected.

Metadata ( $M_1 \dots M_N$ ) can be collected that describes:

- data quality including accuracy, timeliness, completeness and consistency
- conditions under which data is collected/created: context and environmental conditions
- data format: electronic data, paper-based data, data captured in other formats, and data encoding.

Special metadata ( $P_1 \dots P_N$ ) on data provenance can be also collected that describes the journey of the data to the point of use:

- authorising environment: regulations or policies under which data is captured, transmitted, stored, used and shared
- which entities have held the data
- which entities have accessed the data and for what purpose
- what transformations have been performed on the data.

Finally, as data is used for analysis, insights are generated ( $I_1$ ), which can accompany the data for subsequent uses. Insights are a form of data product derived from data and may have an independent life cycle from the data itself. Insights can be used, reused or combined with other data or insights, as demonstrated in Figure 2.

The complexity of potential data life cycles makes the simple model of Figure 12 more likely to be an exception rather than the general model. Figure 13 focuses on access, use of and sharing of data (or data products) with the implications for repeated access to data (and metadata and insights), use of data (and metadata and insights), and sharing of data (and metadata and insights). Taking Figure 13 as the more general model, the considerations for data use become:

- an evaluation of the authority to access data/insights/metadata based on an understanding of provenance data
- an evaluation of the appropriateness of the quality of data for the intended use
- an understanding of the format in which data will be accessed and used
- an evaluation of the authority to use data/insights/metadata, based on an understanding of provenance data
- an evaluation of the authority to share data/insights/metadata, based on an understanding of provenance data
- providing guidance on use of insights and data products created through updated metadata.



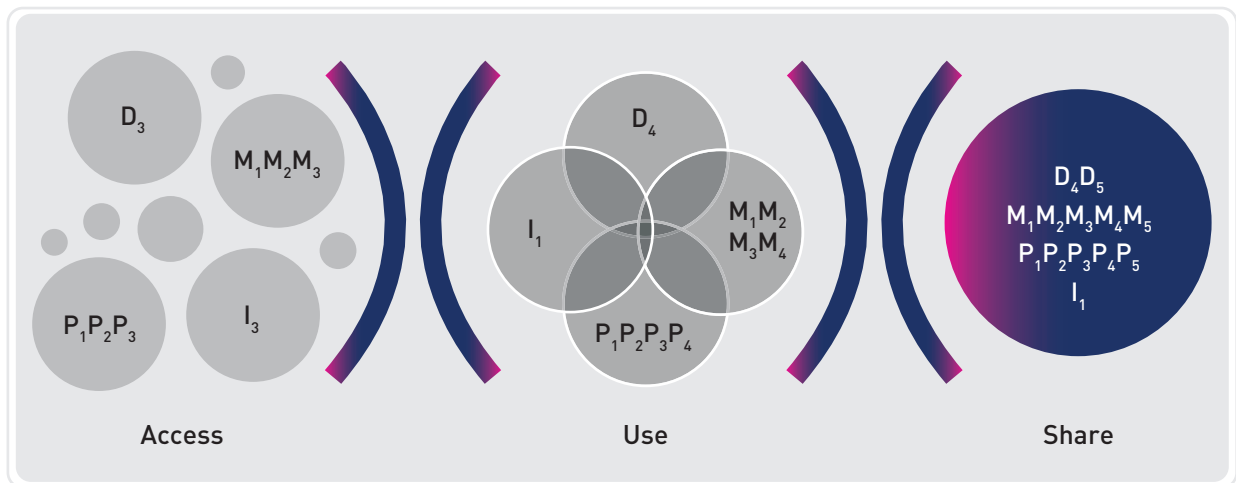


Figure 13. Data life cycle focused on access, use and share

	L	M	H
<b>Sensitivities about data itself:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
1. Concerns that data contains high levels of personal information	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Concerns that data contains uniquely identifiable individuals	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Concerns that sensitive subjects are captured in data (culturally subjective but often described, e.g. religion)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Concerns about data quality (accuracy, timeliness, completeness, consistency)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Concerns about fitness for purpose of data for analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Sensitivities about capability and governance:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Concerns that context is not captured with data (metadata, provenance, consent)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Concerns about authority to share data for analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Concerns about poor governance or accidental release of data or insights (outputs)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Concerns that expert knowledge or context is required to appropriately interpret data and results of analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Concerns about authority to release results of analysis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<b>Sensitivities about use of insights:</b>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Concerns about the level of confidence in outputs (accuracy, consistency, explainability, bias)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. Concerns about unintended consequences from how outputs (insights or data-driven decisions) will be used	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. Concerns about whether human judgement will be applied before an insight becomes a decision	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. Concerns about possible harms resulting from use of outputs (reversible, reversible with cost, irreversible)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. Concerns that results from analysis may lead to negative surprises (especially for data not analysed before)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. Concerns that commercial value may be degraded if insights are shared	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	PIF: L	M	H
	Inherent sensitivity: L	M	H

Figure 14. Sensitivities (most) relevant to different stages of the data life cycle

Returning to the issue of sensitivities around data sharing, Figure 14 attempts to align the major sensitivities. It is now possible to associate sensitivities with the stages of data access, use and sharing.



## 4.1 DATA QUALITY REQUIREMENTS ARE DEPENDENT ON USE

Data quality underpins many concerns about data being released for use, ranging from concerns about the data reflecting poorly on the data custodian to concerns about poor-quality insights or data products being generated from poor-quality input data. If the data quality is not known, then appropriate care may not be taken with data products or insights generated, and how they are used.

Data quality was described earlier as including four dimensions: accuracy, timeliness, completeness and consistency. Figure 15 details a more general two-layer data quality standard with detailed data quality indicators.<sup>13</sup>

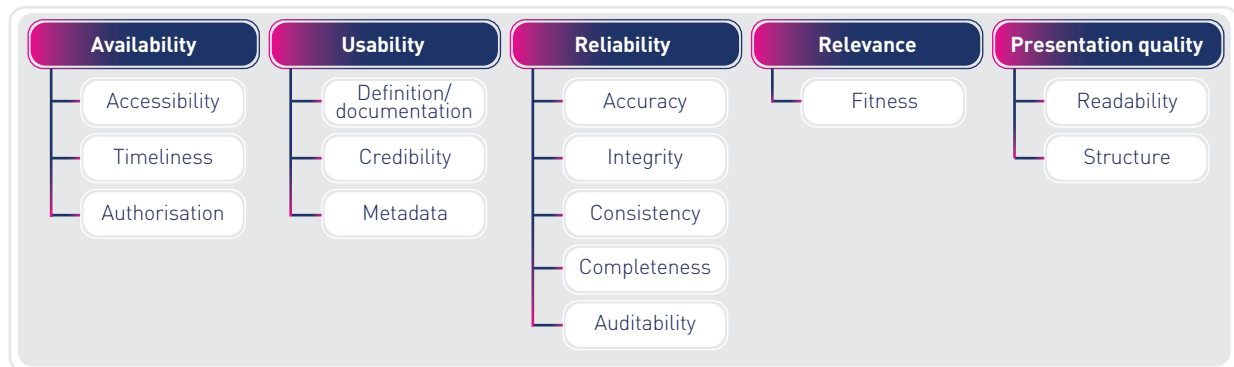


Figure 15. Data quality framework (source: *Data Science Journal*<sup>13</sup>)

This data quality framework is composed of five dimensions of data quality – availability, usability, reliability, relevance, and presentation quality. For each dimension, the authors identified one to five elements to quantify data quality. The first four quality dimensions are regarded as indispensable, inherent features of data quality and the final dimension is additional properties that improve ease of use. The characteristics of these five dimensions can be seen below:

- Availability is defined as the degree of convenience for users to obtain data and related information, which is divided into the three elements of accessibility, authorisation, and timeliness.
- Usability refers to whether the data is useful and meets users' needs, including data definition/documentation, data reliability and metadata.
- Reliability refers to the level of trust in the data; this consists of accuracy, consistency, completeness, adequacy and auditability elements.
- Relevance is used to describe the degree of correlation between data content and users' expectations or demands; adaptability is its quality element.
- Presentation quality refers to a valid description method for the data, which allows users to fully understand the data. Its dimensions are readability and structure.

Data quality has been described in terms of user needs or use case, which makes generic data quality standards difficult to achieve. The international standards body JTC 1's subcommittee 42 on artificial intelligence (AI) is working on draft standards in the form of 'ISO/IEC AWI 5259-1 Artificial intelligence – Data quality for analytics and machine learning'.

<sup>13</sup> Cai L and Zhu Y (2015) 'The Challenges of Data Quality and Data Quality Assessment in the Big Data Era', *Data Science Journal*, 14:2, doi: <http://doi.org/10.5334/dsj-2015-002>.

In the meantime, data quality can be assessed against the intended use. For some data uses, the data quality can be described in relative terms. For example:

**Analysis type:** count (histogram, PDF, CDF, benchmark)

- Data quality requirements: the data field to be counted must be accurate to within counting limit of resolution. Other quality parameters limit use of analysis.

**Analysis type:** thresholding, discriminator, classifier

- Data quality requirements: the value of the data field to be classified must be closest to the correct class value within the classification limit of resolution. Other quality parameters limit use of analysis.

**Analysis type:** prediction

- Data quality requirements: data quality limitations incrementally impact the principal components of the prediction. The data quality of the principal components must be improved to improve algorithm accuracy. Other quality parameters limit use of analysis.

## 4.2 HUMANS AND MACHINES AT EACH STAGE OF THE DATA LIFE CYCLE

The metadata that can be captured will differ depending on whether a person or device undertakes the tasks associated with each stage of the data life cycle (see Figure 16) and how prepared the person or device is to capture this metadata.

Many historical data capture systems do not capture rich metadata, which makes data quality determinations more challenging, the context of data capture difficult to understand, and may cause challenges when determining authority to use or release data and data products. Overall, these challenges combine to hinder data sharing and use.

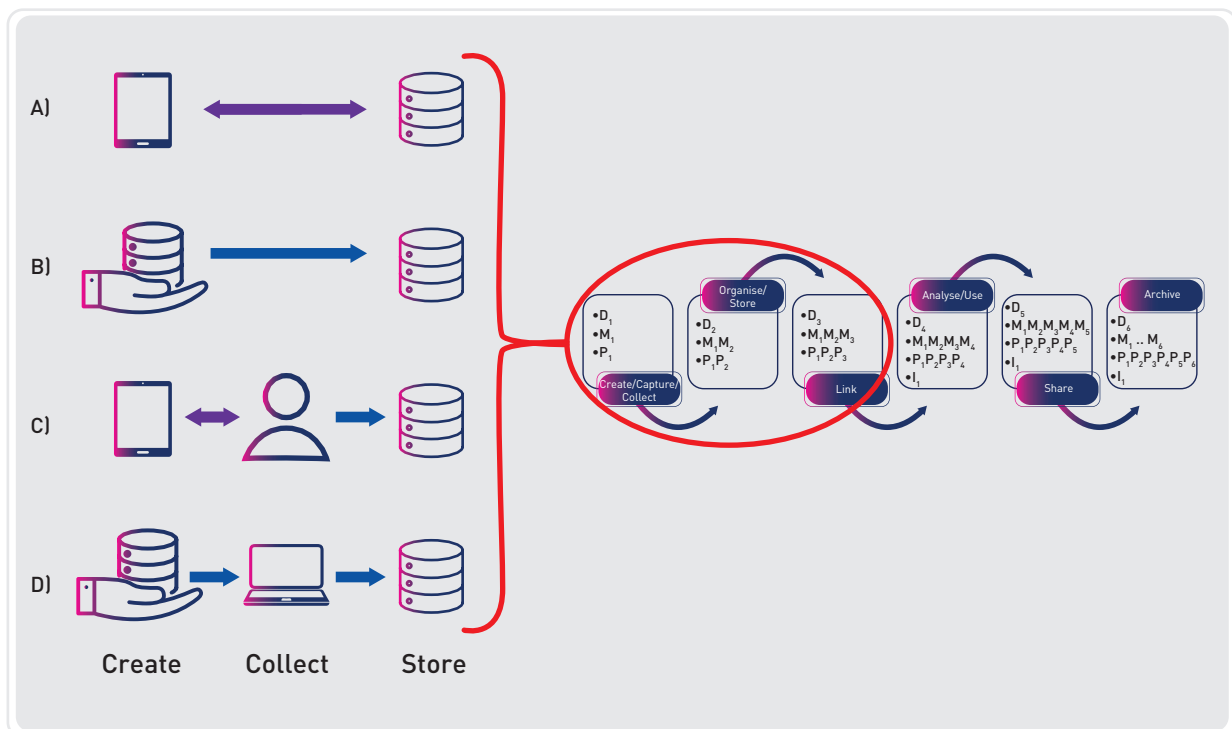
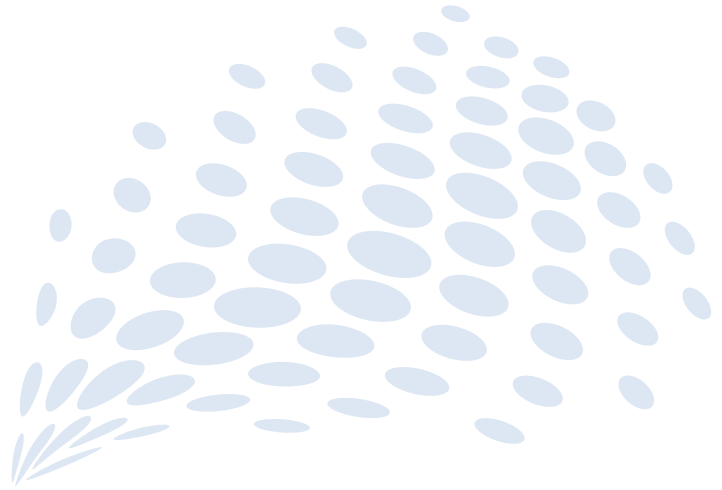


Figure 16. The metadata captured will differ depending on whether a person or device undertakes the tasks associated with each stage of the data life cycle

# 05

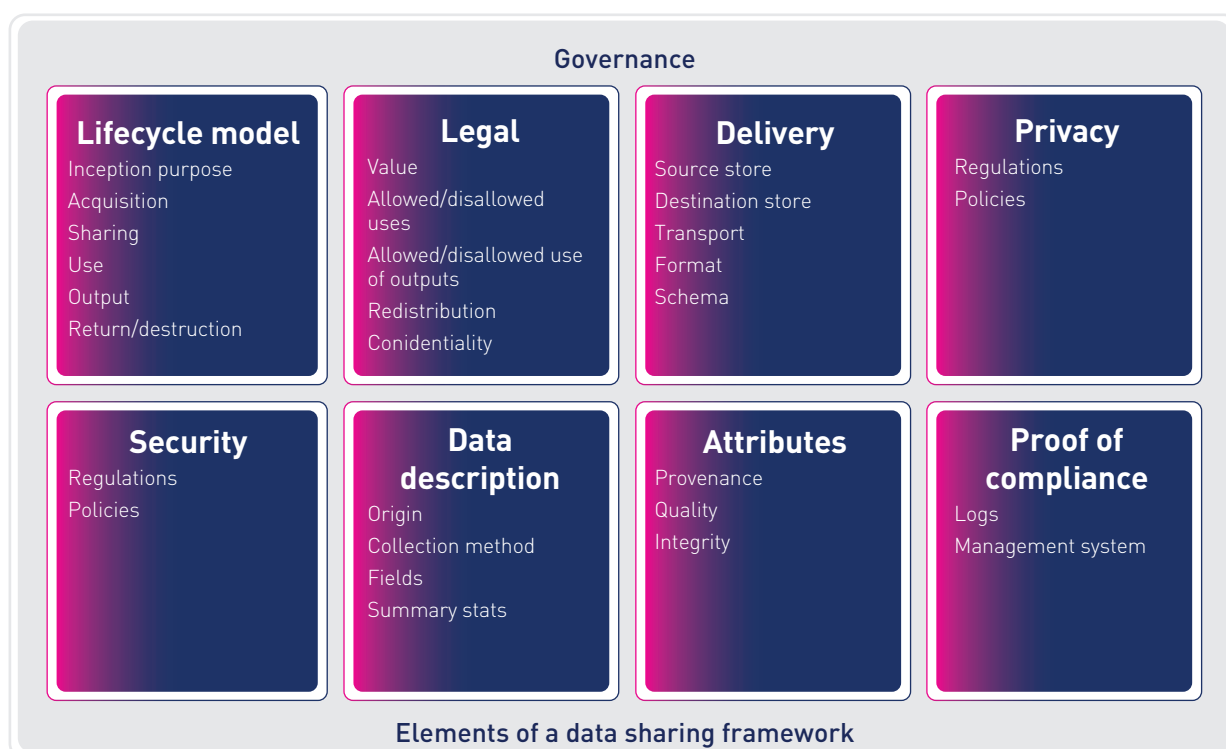


IF THE INHERENT 'RISK' OF DATA USE INCREASES DURING THE LIFE CYCLE OF A PROJECT, THEN THE PROTECTIONS NEEDED MUST ALSO INCREASE OVER TIME TO ENSURE THE PROJECT REMAINS 'SAFE'.



# GOVERNANCE ACROSS THE DATA LIFE CYCLE

Issues of data sharing and use are acknowledged to exist throughout the life cycle of data creation, collection, storage, use, analysis, archival and deletion. If the inherent 'risk' of data use increases during the life cycle of a project, then the protections needed must also increase over time to ensure the project remains 'safe'. The aspects of governance which need to be considered at different stages of the life cycle are shown in Figure 17. In this figure, PIA refers to a privacy impact assessment, a point-in-time evaluation of the potential impact on privacy.<sup>14</sup>



**Figure 17. Governance aspects to be considered for data use at different phases of the data life cycle (source: JTC 1 Advisory Group 9 report)**

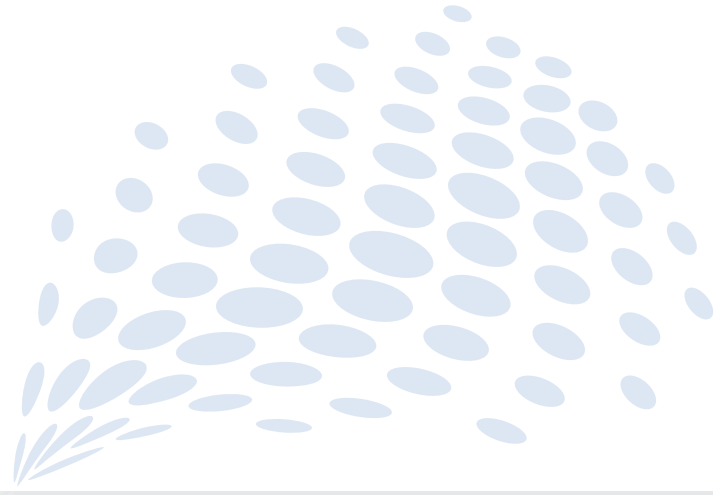
A general framework for data sharing and use is shown in Figure 4. This figure highlights sharing from the most restrictive to least restrictive. The most restrictive is not sharing knowledge that the dataset exists. The least restrictive is allowing access to data where the data user can take a copy of the data. The 'data products' described are intended to include metadata, provenance data (a specific form of metadata), aggregated data and modified versions of the underlying data. The different levels of access provide opportunities for different controls for risks identified during different phases of a project.

This paper assumes all analysis is performed using data that has been deidentified (had unique identifiers removed). It is also assumed that the deidentified data is not subject to any national security classification.

<sup>14</sup> See the Office of the Australian Information Commissioner, at <https://www.oaic.gov.au/privacy/guidance-and-advice/guide-to-undertaking-privacy-impact-assessments/#introduction-to-privacy-impact-assessments>.



# 06



IN EACH PHASE OF A DATA LIFE CYCLE, MULTIPLE ACTIONS TAKE PLACE THAT CHARACTERISE THE STATE OF THE DATA AT THE END OF THE PHASE. EXPLICIT EFFORTS TO CAPTURE THIS STATE, AND COMMUNICATE IT ALONG WITH THE DATA/DATA PRODUCTS TO THE NEXT PHASE OF THE DATA LIFE CYCLE WILL HELP DETERMINE WHETHER THE APPROPRIATE CONTROLS HAVE REMAINED IN PLACE DURING THE PHASES OF THE LIFE CYCLE UP UNTIL THE POINT OF NEXT 'USE'.





# BRINGING IT ALL TOGETHER

## 6.1 APPLICATION OF CONTROLS BASED ON RISK – CONSIDERATIONS AND CONTROLS

In this section, we will bring the respective pieces together and describe the ways to address the sensitivity versus privacy matrix through controls based on identified risk. After assessing a project for sensitivities, 'considerations' help to address these sensitivities and identify appropriate use of controls, based on the dimensions of the 'Five Safes' and the larger risk framework related to use of outputs. Figure 18 shows broad focus areas based on sensitivity of the data (or data use) based on Figure 14. The calculation of a PIF positions the project on the privacy axis.

The framework of controls being examined relies on the ability to determine the level of sensitivity of the information captured in the data, the level of PI in the data (PIF). Different controls can be applied at different phases of the project using data (such as collection, analysis, outputs and use of outputs).

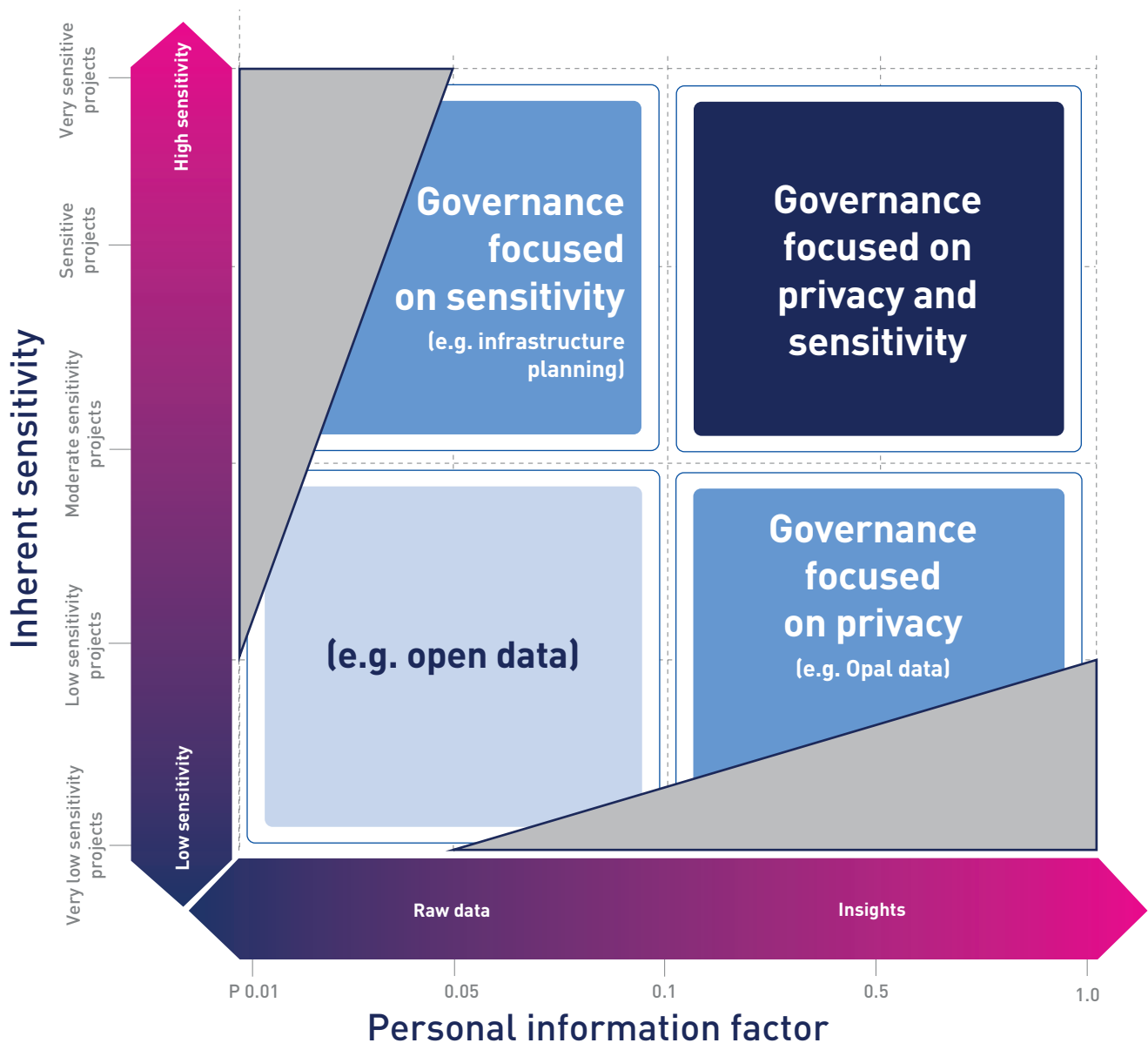


Figure 18. Governance with an emphasis on levels of personal information in data, or inherent sensitivity

## 6.2 CHARACTERISING LEVELS OF CONTROL

In the simplest of data life cycles, two entities may trust each other and establish protocols for data sharing and use with the characteristics discussed earlier in this paper. Once multiple stages of life cycle exist with data or data products on-sharing, more formal structures are required that allow confirmation of:

- authority to receive and use data
- authority to share data or data products
- confirmation of governance capability, systems and processes
- confirmation of technical capability
- confirmation of appropriate domain experience.

Figure 19 shows that some of these aspects interact to create 'very high control' environments, to 'no control' (or open) environments with no limitations on data sharing and use.

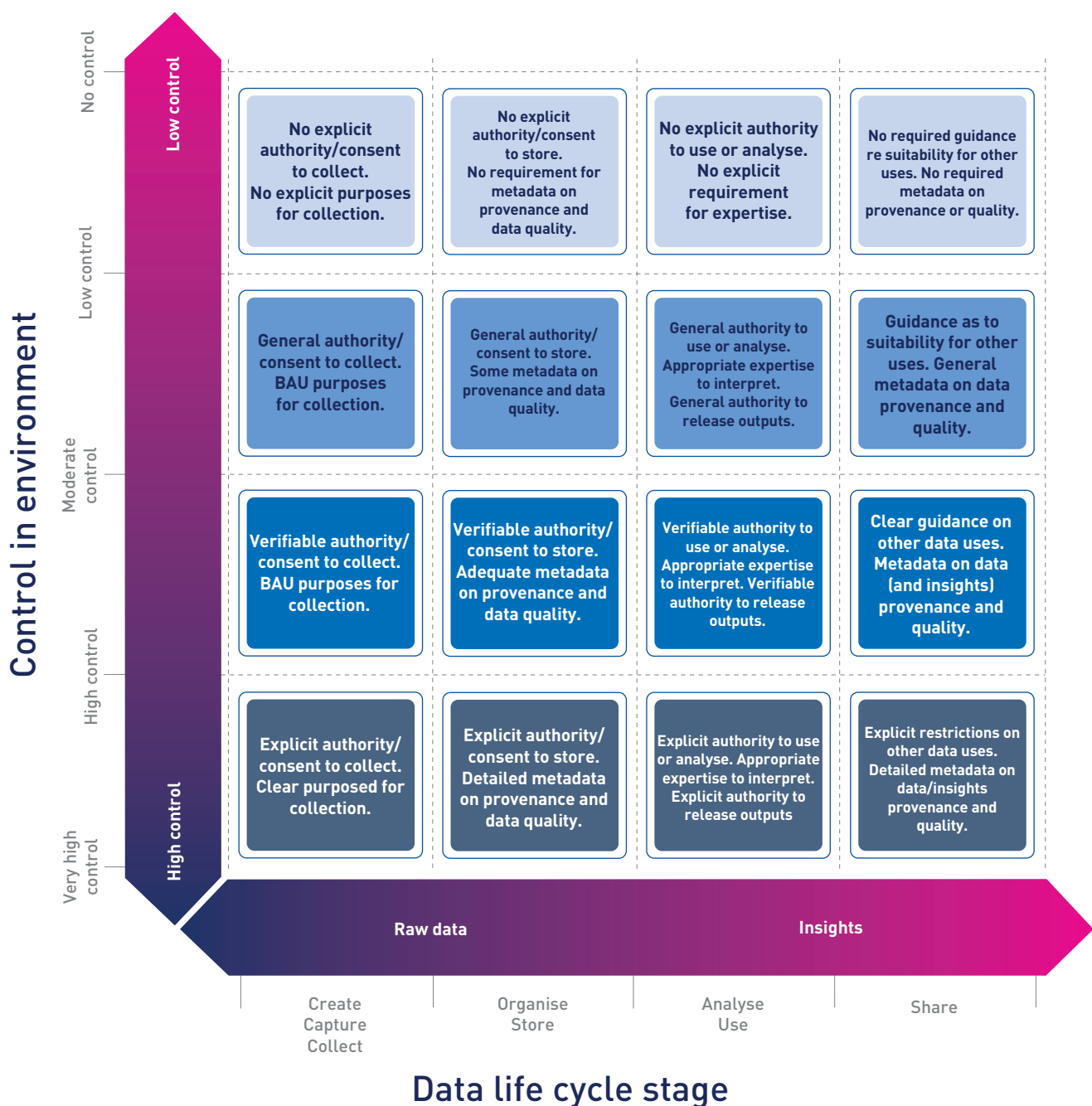


Figure 19. Characterising control layers for first stages in a simplified data life cycle



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

Capability includes skill in all stages of the data life cycle – data analysis, data provenance, governance and security.

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

No control = no assessments or no restriction on people accessing or utilising data

Each of these controls requires an objective, repeatable, standardised assessment of:

- capability
- governance
- purpose
- data quality and provenance
- sensitivity of data
- degree of personal information contained in datasets.

These different control environments can be characterised as follows.

A **very high control** environment

**must have:**

- explicit purpose and authority to access and use data
- expert users experienced with the data of the quality provided and with associated metadata
- expert analytical capability and domain expertise
- strong governance and security at each stage of the life cycle
- explicit restrictions on release of data and insights, or secondary use of data and insights
- people who have met general expertise requirements as well as project-specific requirements for a 'Safe Person' and agree to be bound by limitations on data access and use.

**is suitable for:**

- data that can only be accessed under an external instrument such as a Public Interest Direction (PID)
- data that is reasonably personally identifiable
- data that contains sensitive subject matter
- data that has a well-quantified quality.

## A **high control** environment

### **must have:**

- explicit purpose and authority to access and use data (although it may not have project-specific requirements)
- expert users experienced with the data of the quality provided and with associated metadata
- very skilled analytical capability and domain expertise
- strong governance and security at each stage of the life cycle
- explicit restrictions on release of data and insights, or secondary use of data and insights
- people with access who have met general expertise requirements for a 'Safe Person' and agree to be bound by limitations on data access and use.

### **is suitable for:**

- data that is not reasonably personally identifiable
- data that contains sensitive subject matter
- data that has a well-quantified quality.

## A **moderate control** environment

### **must have:**

- general purpose and authority to access and use data (such as an authorising regulatory framework)
- experienced users dealing with the data of quality provided and with associated metadata
- skilled analytical capability and domain expertise
- strong governance and security at each stage of the life cycle
- general restrictions on release of data and insights, or secondary use of data and insights
- people with access who have met general requirements for a 'Safe Person' and agree to general conditions on data access and use.

### **is suitable for:**

- data that is not reasonably personally identifiable
- data that contains some sensitive subject matter
- data that is of sufficiently high quality for the intended use.

## A **low control** environment

### **may have:**

- no explicit authority to collect and use data, but no known restrictions to use data
- users with some experience dealing with data of the quality provided
- users with some analytical capability and domain expertise
- appropriate governance and security at each stage of the life cycle
- may not have restrictions on release of data and insights, or secondary use of data and insights.

is suitable for:

- data that is not reasonably personally identifiable
- data does not contain sensitive subject matter
- data that is of sufficiently high quality for general use.

A **no control** environment

may have:

- no controls in place.

is suitable for:

- data that has been approved for release as open data
- data that is of sufficiently high quality for general use.

### 6.3 DETERMINING THE LEVEL OF CONTROL REQUIRED

The question to ask now is: What level of control do I require for data sharing and use? Taking the characteristics of the different control environments in reverse, a series of questions can be asked to help identify the level of control required.

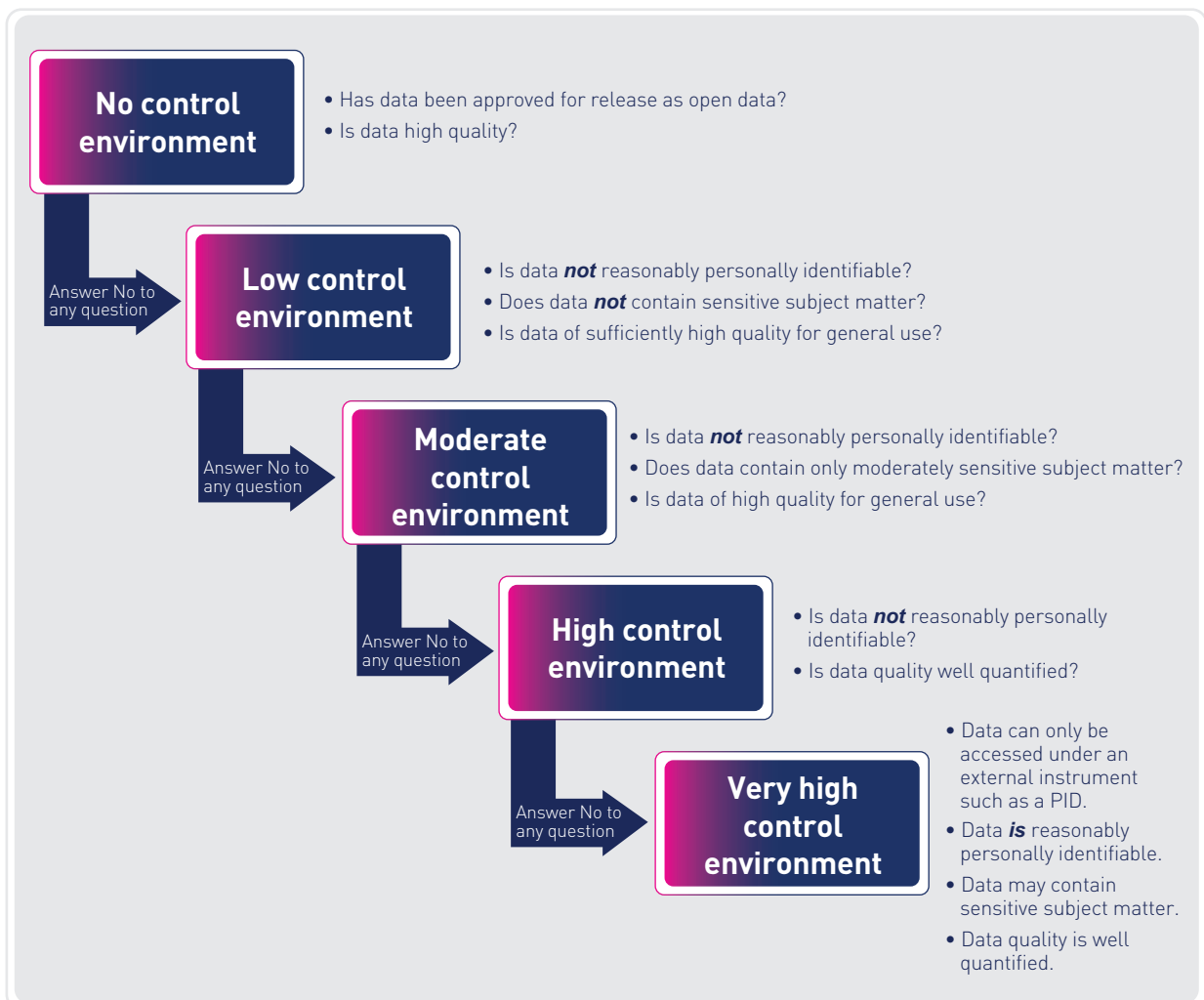


Figure 20. Level of control required for data sharing



# A REAL WORLD EXAMPLE: RELEASE OF COVID-19 DATA

Each day the NSW Ministry of Health analysed its data on confirmed COVID cases to determine if it could be released. From the raw data (referred to as Table 0), data products were created with lower PIF (Table 1 through Table 4), and it was these data products that were assessed for release each day.

The data in the reduced feature tables was analysed each day to ensure the PIF was reduced to an agreed level before release. The dataset was in the form of rows representing unique individuals and columns (with features related to that individual). The data released was also used to create spatial maps for those who did not want to access the data directly.

These were the main characteristics of the activity in terms of the controls for data sharing and use environments:

- data and metadata were collected in digital form by a high-capability team, working with explicit authority
- data was stored with explicit authority from the NSW Ministry of Health
- data was analysed by the NSW Data Analytics Centre (DAC) team under authority from the NSW Ministry of Health
- data analysis calculated the PIF of Table 0, and the PIF of the data products (Table 1 through Table 4) of daily confirmed COVID cases
- if the targeted reduction in PIF values was met, data products (Table 1 through Table 4) were released with some metadata into a 'no control' environment – specifically as open data – with no ability to restrict access, qualify users or limit use of the data.

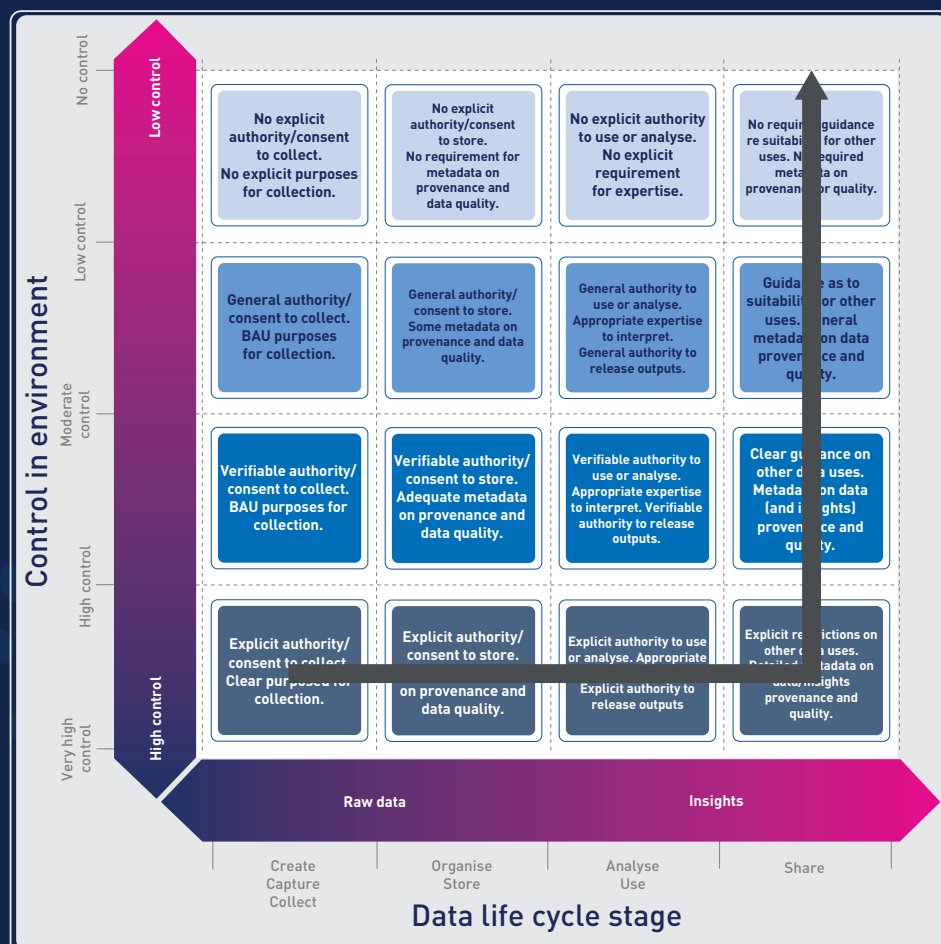


Figure 21. Example of how COVID data products were released daily

## 6.4 WHAT IS A SAFE PERSON?

The Five Safes model has an element referred to as a 'Safe Person'. The general requirements for a Safe Person to work in a project is someone who:

- is verifiably skilled and experienced in their domain's techniques – for example, an analytical expert, governance expert or cyber expert
- has been screened or endorsed by independent authorities – for example, someone who has been endorsed by an executive manager or has completed a police check or working with children check
- understands and agrees to be bound by legal frameworks such as privacy protection legislation and health record protection legislation
- understands and agrees to follow formal governance processes used in the analytical environment
- understands the roles of others in the analytical chain and governance process, and agrees to respect and work with these roles
- understands and is able to use the specific tools and processes in the analytical environment.

From a project-specific standpoint, they:

- are expressly authorised to work with the subject data for an authorised project
- understand and agree to be bound by project legal agreements or restrictions such as a Public Interest Direction (PID) or other project-specific restrictions.

### Individual privacy considerations

As discussed earlier, the knowledge held by a person viewing a dataset or insight may lead to reidentification. Understanding the connection an individual has to a dataset could be an additional consideration.

Additional measures for a Safe Person may include the following elements:

- personal connection to the dataset – understanding the degree of separation between the people represented in the dataset, or the region represented, and the analyst
- accountability – the personal consequences for the analyst in the event that reidentification does occur (PII is attained), PII is released or that PII is used inappropriately by the analyst.

Figure 22 shows the different roles of Safe Persons in the analysis/use phase of the data life cycle. The different roles in the analysis phase include identifying:

- someone with the authority to receive the data and bring it into the analysis phase
- someone with security/governance responsibility
- someone with the required analytical skill at the level of expertise required by the level of the control environment
- someone with domain expertise
- someone with delegated authority to release data and insights, along with restrictions on use and secondary use.

In practice, several of these roles may be held by one person. The roles highlighted in orange are those that may have project-specific requirements, depending on the level of control of the environment. The other roles are generic for any project involving people centric data.

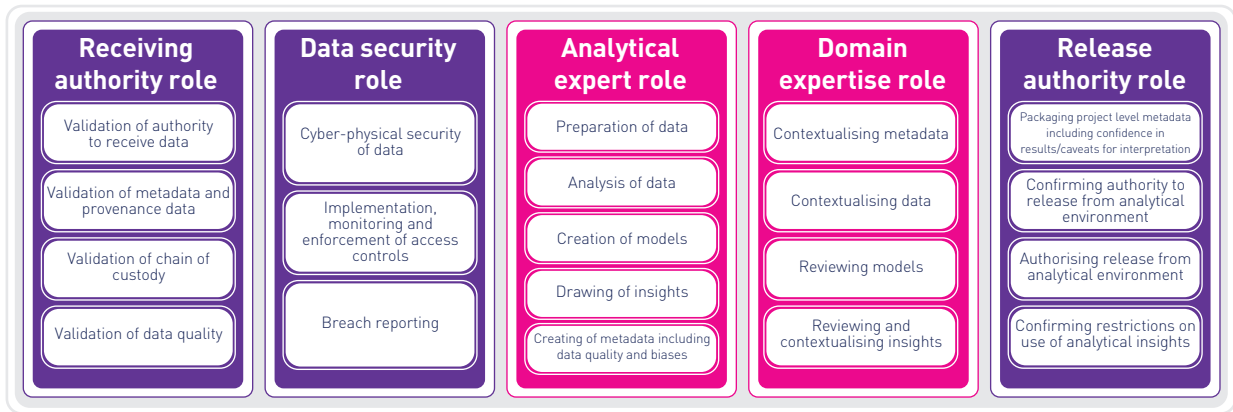


Figure 22. Safe People roles in the analyse/use stage

## 6.5 DETERMINING THE STATE OF CONTROL AT EACH STAGE OF THE DATA LIFE CYCLE

This section explores the stages (or phases) of the data life cycle, the interaction of people and devices to arrive at different possible states from create to use. In each phase, a number of possibilities exist depending on whether:

- authority to bring data into that phase (or create for the first phase) exists
- data is created in digital or non-digital (for example, paper-based) form
- metadata is captured and extended during the phase
- data, data products and insights are released from the phase in the life cycle.

Figure 23 to Figure 25 shows a series of discrete pathways through the create/capture/collect phase and analyse/use phase. In each phase of a data life cycle, multiple actions take place that characterise the state of the data at the end of the phase. Explicit efforts to capture this state and communicate it along with the data/data products to the next phase of the data life cycle will help determine whether the appropriate controls have remained in place during the phases of the life cycle up until the point of next 'use'.

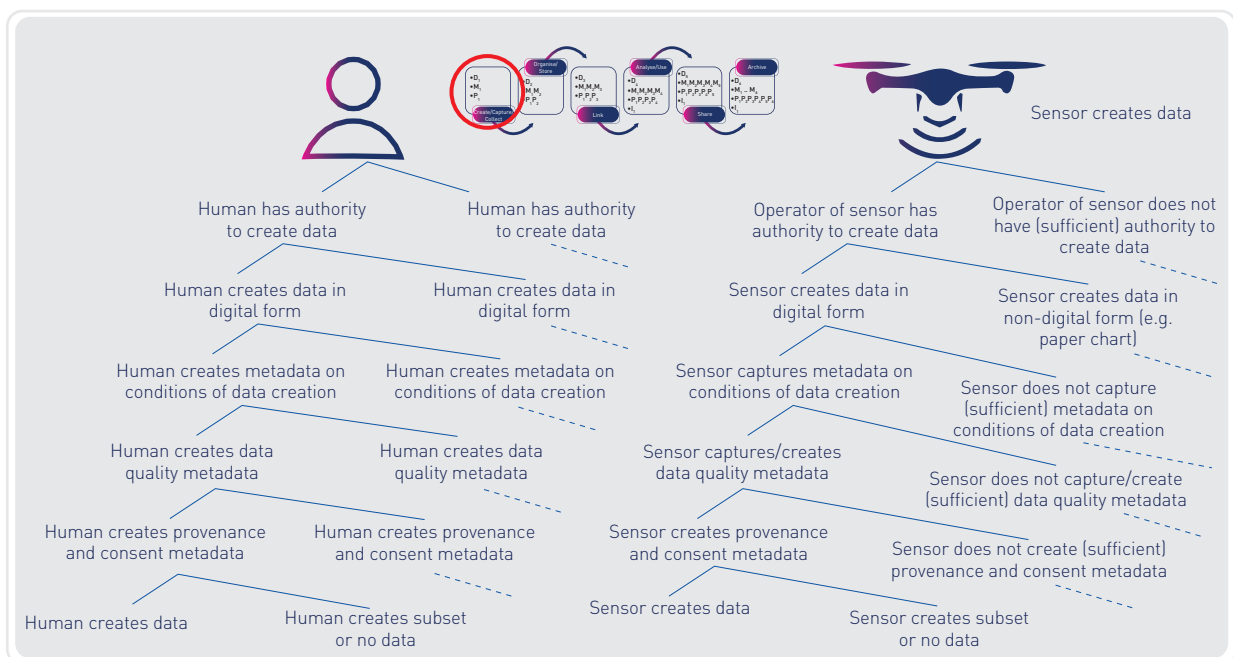


Figure 23. Possible pathways through the create stage

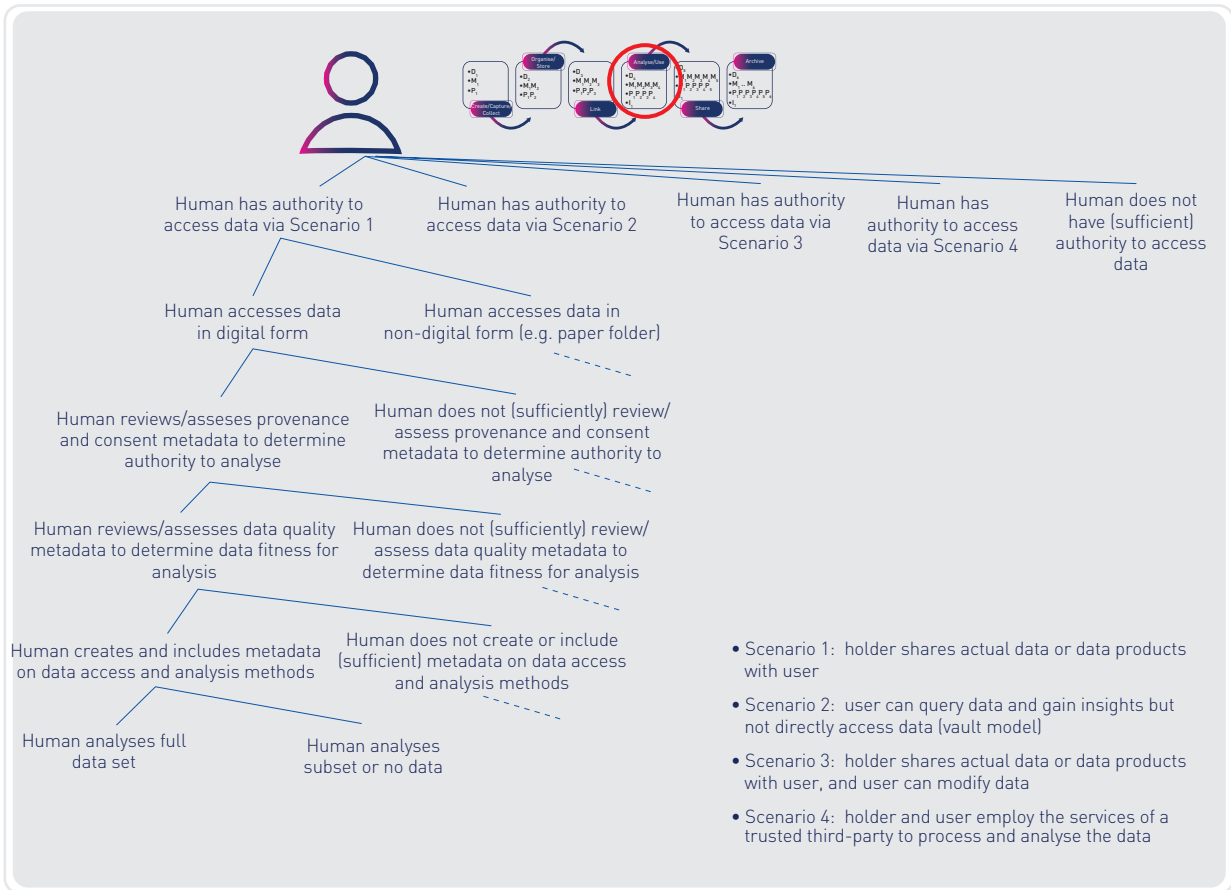


Figure 24. Possible pathways through the analyse stage when analysis is performed by a person

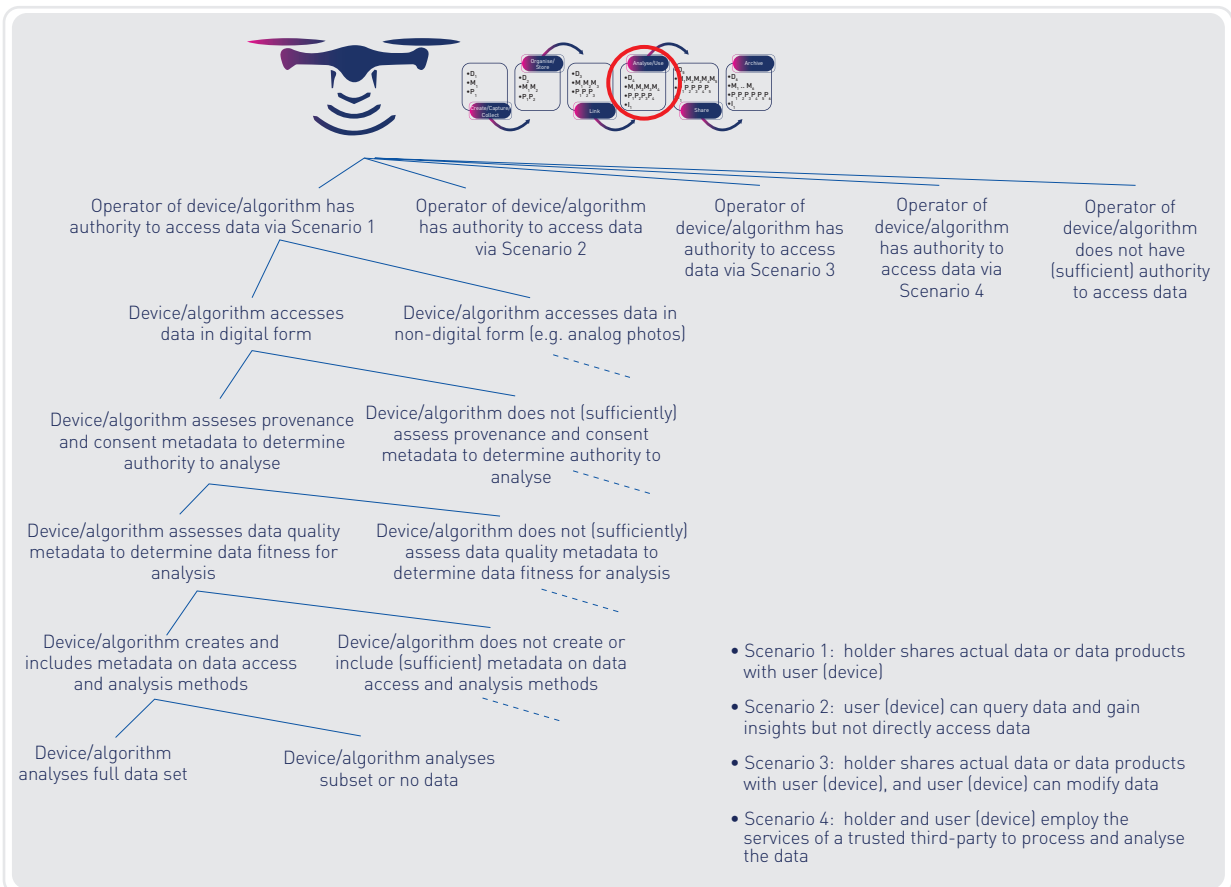
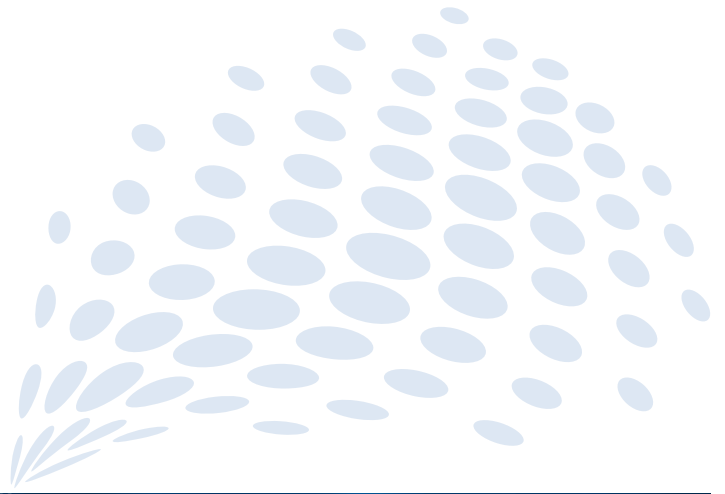


Figure 25. Possible pathways through the analyse stage when analysis is performed by an algorithm

07





# DISCUSSION

The work presented in this paper is an ongoing effort to identify frameworks to safely share and use data. The work identifies controls required to ensure that data is treated appropriately along the entire data life cycle. It is this, often unknown, life cycle that creates so much concern for data custodians and others involved in the data ecosystem, including data subjects themselves.

## 7.1 THE WORK ON PIF IS CONTINUING

The PIF as described in the 2019 ACS Technical White Paper *Privacy-Preserving Data Sharing Frameworks* was a first attempt at defining this parameter and creating a practical tool. The PIF uses information theory to compute privacy risk in a dataset. The tool suggests the associated risks and proposes recommendations for sharing data; for example, suppression of certain attributes. The analysis results are also displayed as visuals, which makes interpretation easier. Based on the associated risks, the tool uses a provable privacy technique (for example, differential privacy) to perturb data.

The Cybersecurity CRC, led by CSIRO's Data61, has continued to develop the original PIF tool and build it into data sharing frameworks. Unlike traditional tools that choose design parameters in an ad hoc fashion, the new AI-based framework considers various attack vectors, user risk appetite and the required level of accuracy to select the design parameters (Figure 26).

The evolved PIF Tool assesses privacy risk in a dataset and provides recommendations while publishing or sharing data. The proposed AI-enabled framework automatically transforms the data to mitigate the identified risks (where possible) using provable privacy techniques like differential privacy.

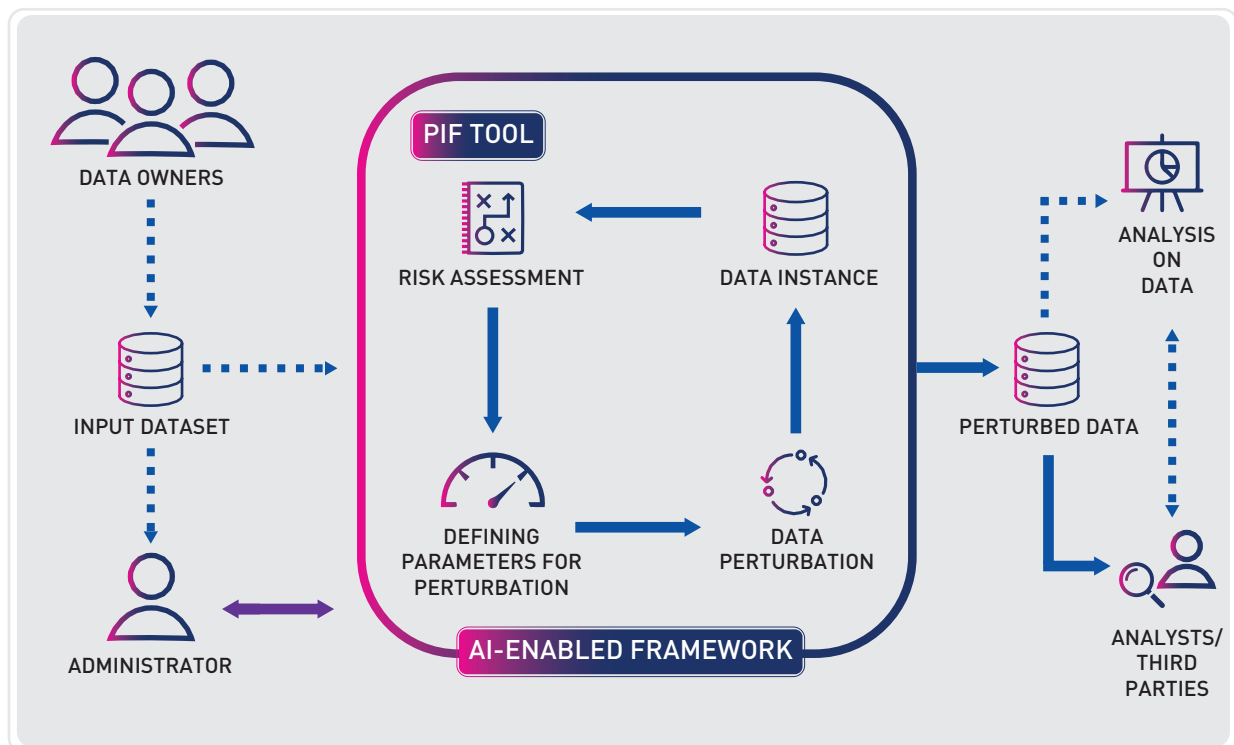


Figure 26. Overview of ongoing work to evolve the PIF (source: Data61)

## 7.2 THE NEED FOR STANDARDS

Standards are fundamental to systematic data sharing. Standards on terminology, use cases, ways of sharing, roles, and responsibilities as well as governance and security are all important elements to ensure safe data sharing and use. There is a great deal of work taking place in the world of standards, which provides useful resources for data sharing frameworks. Standards Australia is the national member body at ISO<sup>15</sup> and the IEC,<sup>16</sup> and JTC 1, their joint technical committee focused on intersectional information technology.

The most relevant groups within the IEC/ISO/JTC 1 family include subcommittees (SC) for data sharing and use include:

- SC 27 – Information security, cybersecurity and privacy protection
- SC 32 – Data management and interchange
  - within SC 32, Working Group 6 (WG 6) on data usage
- SC 38 – Cloud computing and distributed platforms
- SC 40 – IT service management and IT governance
- SC 41 – Internet of things and digital twin
- SC 42 – Artificial intelligence.

The subcommittee on AI (SC 42) and on data usage (WG 6) are two important groups to watch as they develop their work programs. SC 42 explores generic data quality standards and issues of bias in data and algorithms. WG 6 explores terminology, use cases and ways of mitigating sensitivities of data sharing. A more comprehensive list of relevant standards is provided in the Appendix – Resources.

---

15 International Organization for Standardization <https://www.iso.org/home.html>.

16 International Electrotechnical Commission <https://www.iec.ch/homepage>.

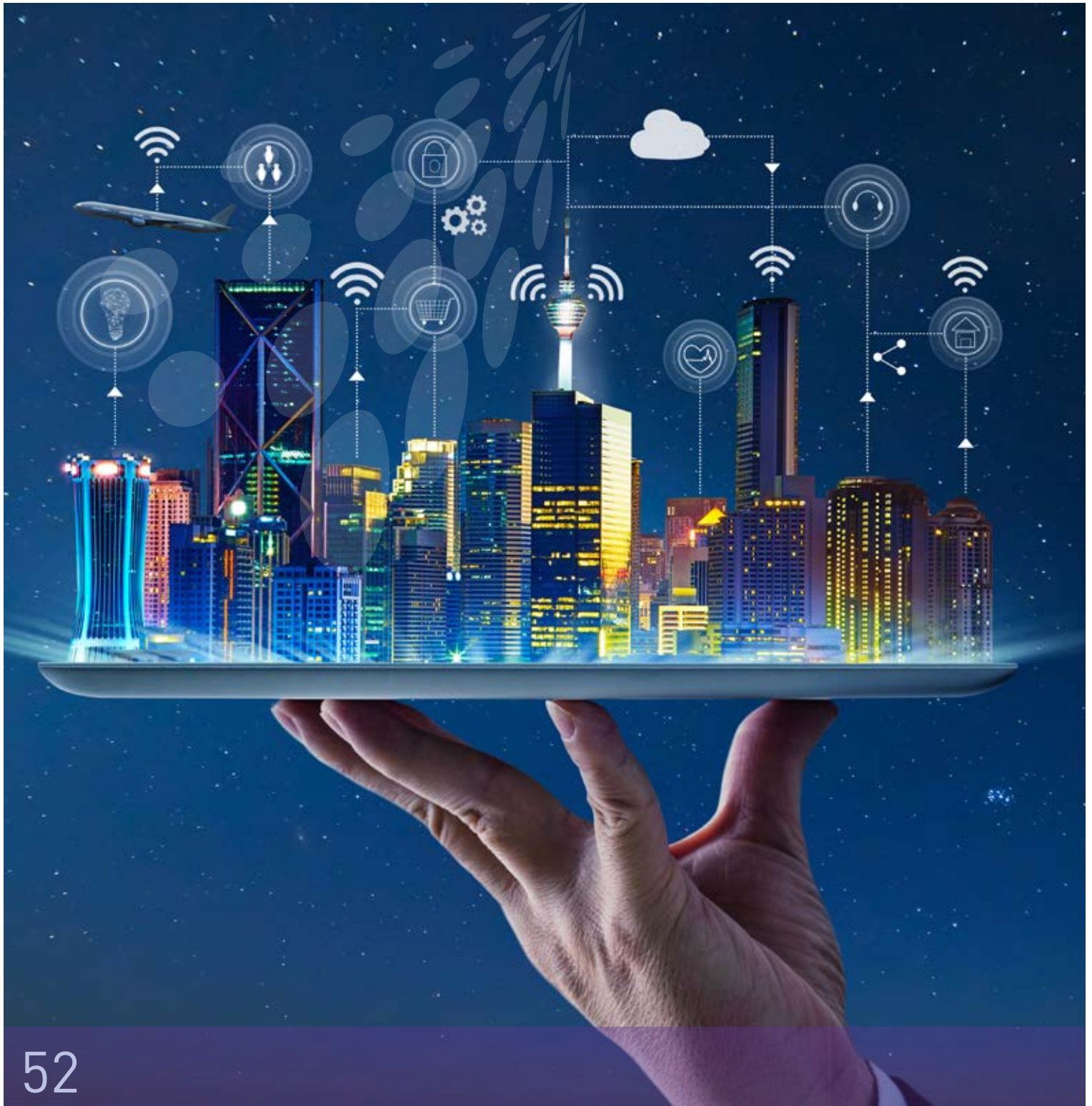
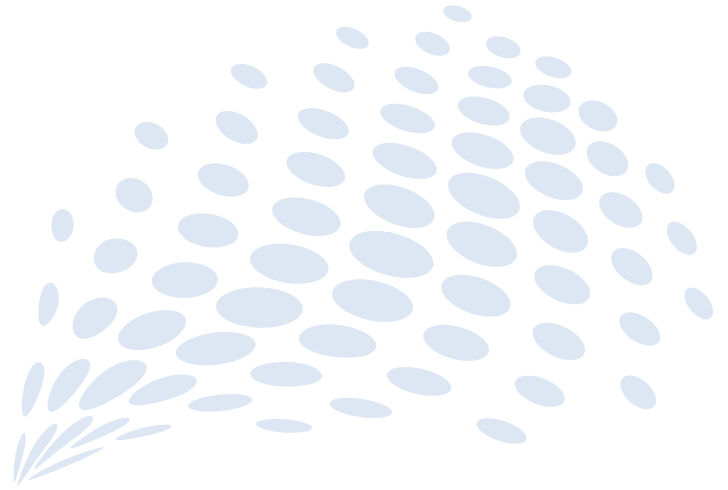


STANDARDS ARE  
FUNDAMENTAL TO  
SYSTEMATIC DATA  
SHARING.





08



# CONCLUSIONS

For decades, most economies have become more services-dominated, and these services economies are increasingly digital, online and driven by data. In all sectors, services are increasingly created, delivered and consumed via digital means, driven by the increasing adoption of online, mobile, digital technologies. Recent years have witnessed dramatic changes in the way we consume music, watch movies or arrange dinner. While less obvious, there are dramatic changes underway in all industry sectors, and in government, driven by changing consumer expectation, reduced barriers to new entrants and an increasingly borderless world of information flow.

Major drivers for this change are coming from the intersection of a growing, ageing and urbanising population and a globally changing climate and political landscape, overlaid with our expectation of an ever-improving quality of life. In most economies these drivers of change are also giving rise to sectoral challenges such as rising healthcare costs, increasing household energy prices and pressures on government to do more with less.

In a world with finite resources, these challenges must be met with ever-greater productivity increases and technological enhancements. Continued technological advancement and rapid adoption are central to our ongoing response to these major drivers.

As technology and digital solutions increasingly play a role in driving the economy and society forward, they become pervasively embedded into business operations, across key service offerings and into our personal lives. New developments spur more innovative business models, products and services, which are crucial in responding to our current challenge but also lead to accelerating use of and generation of data and digital services.

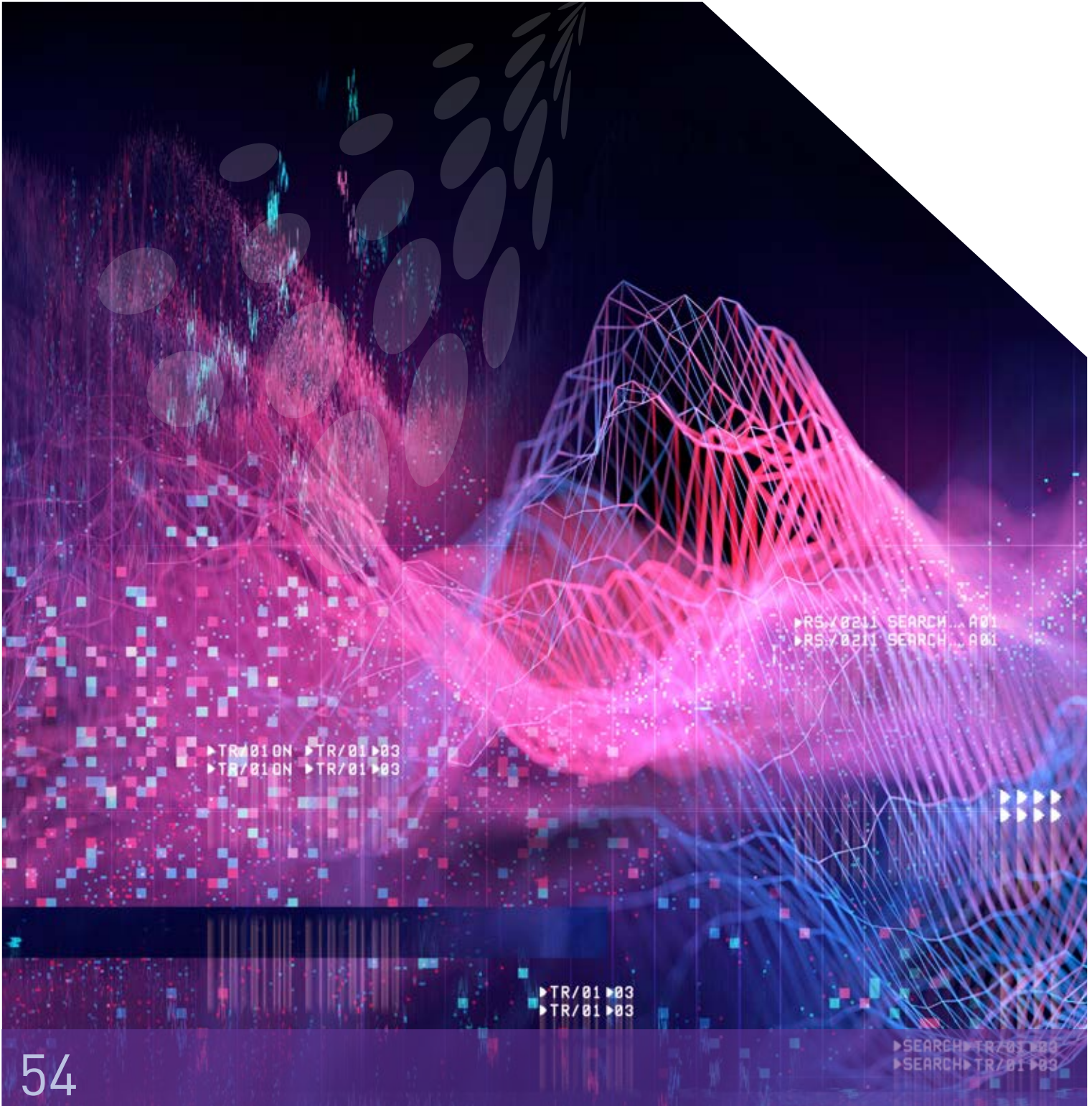
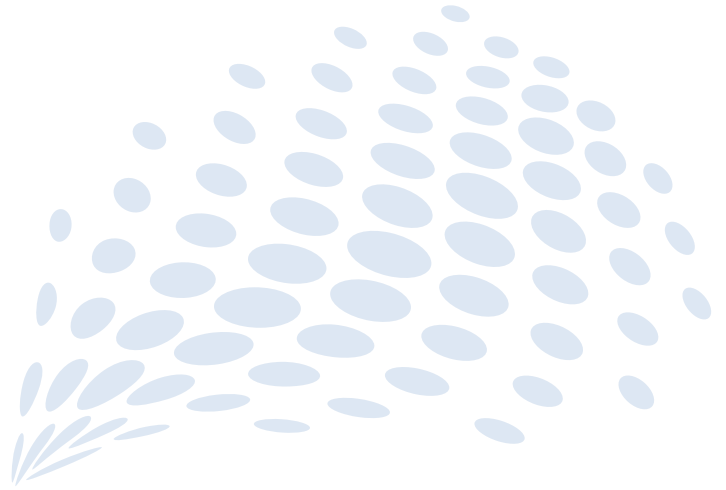
In the near future, it will become a self-reinforcing process accelerated by increased use of AI to make sense of the rising tide of data in continuing to locally optimise services delivery and to increasingly personalise.

As systems develop, the appropriate handling of deidentified personal data, operating within appropriate authorising frameworks, will need to be considered alongside technological capability. New methods for managing, enhancing and evaluating metadata will assist, and new frameworks for sharing and using data will need to be considered.

The frameworks presented here are a part of that work. They provide a working, if not fully complete, model for how to reduce the risks associated with the sharing of data while still enabling the benefits. Combined with the three previous ACS white papers, we hope they can provide a workable foundation for business and government to enable the sharing of data with confidence, and thereby reap the benefits that shared data can deliver.



# 09



▶TR/010N ▶TR/01▶03  
▶TR/010N ▶TR/01▶03

▶RS/0201 SEARCH...A01  
▶RS/0201 SEARCH...A01



▶TR/01▶03  
▶TR/01▶03

▶SEARCH▶TR/01▶03  
▶SEARCH▶TR/01▶03



# THANKS

The 2021 ACS paper was the culmination of more than five years work by a Taskforce that included ACS, the NSW Data Analytics Centre (DAC), Standards Australia, the office of the NSW Privacy Commissioner, the NSW Information Commissioner, the Australian Government's Digital Transformation Agency (DTA), CSIRO, Data61, the Department of Prime Minister and Cabinet, the Australian Institute of Health and Welfare (AIHW), SA-NT DataLink, the South Australian Government, the Victorian Government, the Western Australian Government, the Queensland Government, the Communications Alliance, the Internet of Things Alliance Australia, Ambiata, Data Synergies, Creator Tech, Objective, EY, Microsoft, Clayton Utz and several other companies.

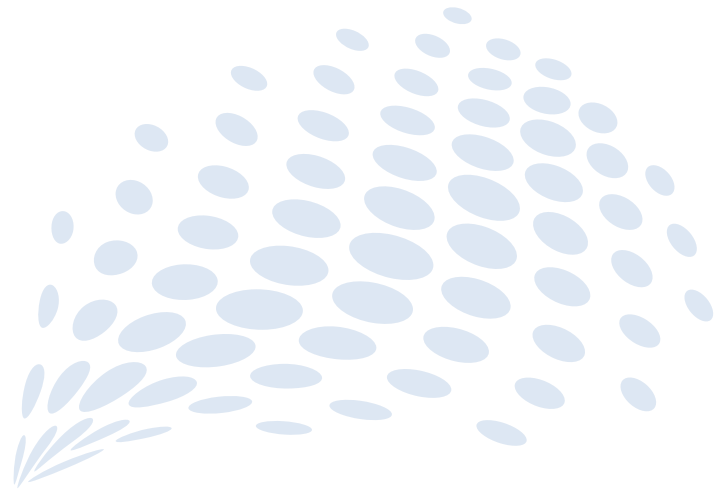
Thanks go to the contributors to many, many 'Safe' workshops over five years including:

Lyria Bennett Moses, Kimberlee Weatherall, Stephen Hardy, Peter Leonard, Chris Radbone, Geof Heydon, Sonya Sherman, Mathew Baldwin, Geoff Neideck, Frank Zeichner, Malcolm Crompton, Geoff Clarke, Kate Cummings, Ghislaine Entwisle, Ghazi Ahamat, Ben Hogan, Scott Nelson, Adrian Watson, Rachael Fraher, Alex Harrington, Andy West, Angelica Paul, Ashton Mills, Brian Thorne, Bridget Browne, Cassandra Gligora, Chris Mendes, Daniel Marlay, Dominic Guinane, Kelda McBain, Liz Bolzan, Luke Giles, Marilyn Chilvers, Matthew Roberts, Matthew McLean, Michael Wright, Mike Willett, Peter Hatzidimitriou, Rick Macourt, Robin van den Honert, Roulla Yiacoumi, Shona Watson, Suyash Dwivedi and Tiffany Roos.

And finally, thanks to all others who have provided, and continue to provide, contributions and feedback.



# 10





# APPENDIX – RESOURCES

## FURTHER INFORMATION ON THE PIF TOOL

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>.

## DATA PRIVACY AND GOVERNANCE – TRAINING, TEMPLATES AND TOOLS

### Office of the Australian Information Commissioner (OAIC)

These training resources are designed to help organisations and agencies develop or improve their privacy training programs.

<https://www.oaic.gov.au/privacy/training-resources/>

## DDMM DIGITAL AND DATA TRUST PRINCIPLES

The Australian Data and Digital Council (Digital and Data Ministers' Meeting) has committed to using data and digital technologies to improve the lives of Australians, now and into the future. This includes projects to drive smarter service delivery and improved outcomes for you. Through bringing together ministers from the Australian Government and all state and territory governments, the Council seeks to improve the way jurisdictions work together to deliver data and digital projects.

<https://pmc.gov.au/sites/default/files/publications/trust-principles.pdf>

## NSW DATA GOVERNANCE TOOLKIT

The Toolkit contains 12 modules that are designed to help agencies improve their ability to govern their data.

- Module 1: Introduction to Data Governance
- Module 2: Legal and Policy Context
- Module 3: Data Governance Model
- Module 4: Strategy and Planning
- Module 5: Organisational Structures
- Module 6: Assigning roles and responsibilities
- Module 7: Leadership
- Module 8: Data-driven Culture
- Module 9: Workforce Skills and Capability
- Module 10: Technology
- Module 11: Data Management
- Module 12: Data Governance Checklist

<https://data.nsw.gov.au/data-governance-toolkit-0>

## SMART CITIES – TRAINING, TEMPLATES AND TOOLS

### **Smart Places Masterclass (recorded video)**

A 10-part Smart Places Masterclass Series developed in the lead up to the NSW Smart Places Summit in August 2021 focused on the core drivers of human trust, privacy preserving data management and cyber security strategies, and how government might incorporate these principles in driving better engagement with communities:

- Session 01 – Building Trust when creating Smart Places
- Session 02 – Smart Place Standards
- Session 03 – Digital Twins
- Session 04 – Data Sharing and Use, AI and Data Governance
- Session 05 – Smart Places Maturity Models
- Session 06 – 5G and Smart Places
- Session 07 – Precincts and Public Spaces
- Session 08 – Cyber Security in Smart Places
- Session 09 – An overview of IoT in Smart Places
- Session 10 – Investing in Smart Places

<https://www.dpie.nsw.gov.au/our-work/strategy-and-reform/smart-places/smart-places-masterclass>





## RELEVANT STANDARDS FOR DATA USAGE – IEC, ISO, JTC 1 AND BSI

Publisher	Designation	Title
ISO/IEC	19944-2:2020	Cloud and distributed platforms – Cloud services and devices: data flow, data categories and data use – Part 2: Guidance on application and extensibility
ISO/IEC	19944-1:2020	Cloud and distributed platforms – Data flow, data categories and data use – Part 1: Fundamentals
ISO/IEC	15489-1:2016	Information and Documentation – Records management – Part 1: Concepts and principles
BSI	BS 30301:2019	Information and Documentation. Management Systems for records. Requirements (British Standard)
ISO/IEC	BS 30301:2019	Information technology – Governance of IT for the organization
ISO/IEC	24368	Information technology – Artificial intelligence – Overview of ethical and societal concerns
ISO/IEC	24668	Information technology – Artificial intelligence – Process management framework for big data analytics
ISO/IEC	20546:2019	Information technology – Big data – Overview and vocabulary
ISO/IEC	20547-3:2020	Information technology – Big data reference architecture – Part 3: Reference architecture
ISO/IEC	20547-1:2020	Information technology – Big data reference architecture Part 1: Framework and application process
ISO/IEC	20547-4:2020	Information technology – Big data reference architecture Part 4: Security and privacy
ISO/IEC	15944-8:2012	Information technology – Business Operational view – Part 8 Identification of privacy protection requirements as external constraints on business transactions
ISO/IEC	15944-1:2011	Information technology – Business Operational View – Part 1: Operational aspects of open-edi for implementation
ISO/IEC	15944-12:2020	Information technology – Business operational View – Part 12: Privacy protection requirements (PPR) on information life cycle management (ILCM) and EDI of personal information (PI)
ISO/IEC	22624:2020	Information technology – Cloud computing – Taxonomy based data handling for cloud services

<b>Publisher</b>	<b>Designation</b>	<b>Title</b>
ISO/IEC	19583-23:2020	Information technology – Concepts and usage of metadata – Data element exchange (DEX) for a subset of ISO/IEC 11179-3
ISO/IEC	19583-1:2019	Information technology – Concepts and usage of metadata – Part 1: Metadata concepts
ISO/IEC	38508	Information technology – Governance of IT – Governance of data – Guidelines for data classification
ISO/IEC	38505-1:2017	Information technology – Governance of IT – Governance of data – Part 1: Application of ISO/IEC 38500 to the governance of data
ISO/IEC	38505-2:2018	Information technology – Governance of IT – Governance of data – Part 2: Implications of ISO/IEC 38505-1 for data management
ISO/IEC	11179-3:2013	Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes
ISO/IEC	14662:2010	Information technology – Open-edi reference model
ISO/IEC	27001:2013	Information technology – Security techniques – Information security management systems – Requirements
ISO/IEC	27550:2019	Information technology – Security techniques – Privacy engineering for system life cycle processes
ISO/IEC	27701:2019	Security techniques – Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management – Requirements and guidelines
BSI	PAS 183:2017	Smart Cities. Guide to establishing a decision-making framework for sharing data and information services
ISO/IEC	10032:2003	Information technology – Reference Model of Data Management
ISO/IEC	11179-1:2015	Information technology – Metadata registries (MDR) – Part 1: Framework
ISO/IEC	11179-2:2019	Information technology – Metadata registries (MDR) – Part 2: Classification
ISO/IEC	11179-3:2013	Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes

# ACS – HELPING MOVE AUSTRALIA FORWARD

**ACS has represented Australia's technology professionals for more than 50 years as the industry's peak professional body, and currently boasts 43,000 members across every state and in every technology profession. Every year we run hundreds of professional development events and support our members as they progress in their careers.**

ACS also has online learning and collaboration tools, skills assessment tools and services, professional certification, events and conferences, roundtables and summits. Technology professionals of all types are welcome and we look forward to supporting and working with you.

**Contact us at [member.services@acs.org.au](mailto:member.services@acs.org.au) or visit our website at [acs.org.au](http://acs.org.au) to learn more on what we do for our members.**

In addition to the many services we offer to our members, ACS has a remit to advance the cause of technology and the technology industry in Australia. We believe in working for the good of our nation, and we work with our members and our partners in government and business to help Australia advance through the 21st century.

We produce thought leadership papers like the one you're reading now, run summits and events, and create opportunities for experts in different technology fields to get together to solve some of the nation's biggest problems.

If you're interested in joining us in this effort or would simply like to find out more, we recommend visiting our website at [acs.org.au](http://acs.org.au) to take a look at the many projects we're undertaking.

You can also email [governance@acs.org.au](mailto:governance@acs.org.au) if you're interested in engaging with our technical and professional advisory boards that are pushing forward this work and work like it. You don't have to be a member of ACS – we're happy to talk to anyone with an interest in making Australia better through technology.

## ACKNOWLEDGEMENT OF COUNTRY

**ACS acknowledges the Aboriginal and Torres Strait Islander peoples of this nation. We acknowledge the traditional custodians of the lands in which our Society is located and where we conduct our business. We pay our respects to ancestors and Elders, past and present.**

**ACS is committed to honouring Aboriginal and Torres Strait Islander peoples' unique cultural and spiritual relationships to the land, waters and seas and their rich contribution to society.**



# CONTACT US

## MEMBER SERVICES GENERAL ENQUIRIES

**E:** [member.services@acs.org.au](mailto:member.services@acs.org.au)

**T:** +61 (0)2 9299 3666

### Canberra

T: +61 (0)2 6143 5503

E: [acs.canb@acs.org.au](mailto:acs.canb@acs.org.au)

### New South Wales

T: +61 (0)2 9299 3666

E: [acs.nsw@acs.org.au](mailto:acs.nsw@acs.org.au)

### Northern Territory

T: +61 429 460 140

E: [acs.nt@acs.org.au](mailto:acs.nt@acs.org.au)

### Queensland

T: +61 (0)7 3316 5700

E: [acs.qld@acs.org.au](mailto:acs.qld@acs.org.au)

### South Australia

T: +61 (0)8 8363 6660

E: [acs.sa@acs.org.au](mailto:acs.sa@acs.org.au)

### Tasmania

T: +61 (0)3 6212 0225

E: [acs.tas@acs.org.au](mailto:acs.tas@acs.org.au)

### Victoria

T: +61 (0)3 9249 6700

E: [acs.vic@acs.org.au](mailto:acs.vic@acs.org.au)

### Western Australia

T: +61 (0)8 9470 4878

E: [acs.wa@acs.org.au](mailto:acs.wa@acs.org.au)





**ACS**

Level 27, Tower One  
100 Barangaroo Ave  
Sydney NSW 2000

T: 02 9299 3666

F: 02 9299 3997

E: [info@acs.org.au](mailto:info@acs.org.au)

W: [www.acs.org.au](http://www.acs.org.au)