



November 2018



Privacy in Data Sharing: A Guide for Business and Government

An ACS White Paper



About the editor



Dr Ian Oppermann, FACS CP

Ian is the NSW Chief Data Scientist, CEO of NSW Data Analytics Centre and ACS Vice President of Academic Boards.

He has over 25 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 in the area of the Digital Economy and is a regular speaker on big data, broadband-enabled services and the impact of technology on society. Ian has an MBA from the University of London and a Doctor of Philosophy in Mobile Telecommunications from Sydney University.

Many people came together to make this report a reality. We'd like to give thanks to the following organisations for their assistance and input.



CLAYTON UTZ



Objective

protiviti®
Face the Future with Confidence





Foreword

One immutable truth of the digital age is that the data of our citizens is being gathered, shared and analysed at a scale we've never seen before.

The ubiquity of social media, mobile applications, on-demand streaming services, online shopping and payment systems means that all businesses, old and new, have become both conduits and custodians of consumer data.

The value of shared and open data is immense – up to \$25 billion per year in Australia according to the 2016 Commonwealth Government report *Open government data and why it matters*. But unlocking this economic and social potential, while also maintaining robust privacy safeguards, is a challenge faced by businesses, NGOs and governments around the world.

In most cases, we are relying on privacy laws which were written in an era long before the internet, e-commerce and social media were even contemplated. This means that as government leaders and policy makers, we are playing catch up, trying to keep pace with the changing paradigm.

Increasing the speed and visibility of data sharing across our economy enables companies to better target their products and services towards the needs of customers. Equally this transfer of data enables governments to make faster and smarter decisions about the allocation of finite resources.

Over the coming decades, there are many big policy questions we need to consider when it comes to the rules governing the collection, sharing and re-use of citizens' data. As government leaders in the digital age, we must strive to enhance, wherever possible, the privacy safeguards afforded to citizens.

I commend the ACS for this thought-provoking and timely paper. It's the culmination of a significant collaborative effort involving government agencies, both state and federal, NGOs as well as ICT industry and business representatives.

This is a major milestone in our quest to unlock the incredible potential of data sharing, of open government and open business. I thank everyone who made a contribution to this ground breaking paper.

The Hon Victor Dominello MP
NSW Minister for Finance, Services and Property

Foreword by ACS

In the 2017 edition of **ACS Australia's Digital Pulse**, we highlighted 13 policy priorities that would fuel Australia's digital workforce boom and ignite the next phase of economic growth for our nation.



Yohan Ramasundara
President, ACS

Two of these recommendations included accelerating efforts to open data and building Australia's cyber capabilities to provide confidence and trust.

Across all levels, government collects and holds a significant amount of data and there is substantial economic value in making government data publicly available through the creation of new data-driven products and services.

While we regularly read that data is the new fuel for the digital economy, you will notice ACS refers to data as crude oil. It needs to be refined.

Our September 2018 report **Australia's IoT Opportunity: Driving Future Growth** reveals there is a \$308b upside opportunity if we recognise that IoT is no longer sensors and actuators; rather it's rapid real-time insight and predictive capability based on artificial intelligence. Artificial Intelligence via software applications is the data refinement process and has the potential to develop scalable Australian solutions that can be taken to the rest of the world.

We need to consider all forms of data as strategic assets capable of delivering added value for our nation. We are grateful that the ACS Data Sharing Committee has continued to commit its energies towards determining frameworks that will optimise privacy and enhance public and consumer trust in order to unleash the untapped value of data for the Australian economy.

We would like to thank our ecosystem partners for their support and assistance in developing this white paper, in particular our Vice President Dr Ian Oppermann and our Data Sharing Committee of Geof Heydon, Ghislaine Entwistle, Ben Hogan, Chris Mendes, Ghazi Ahamat, Chris Radbone, Dr Stephen Hardy and Dr Wenjie Zhang.



Andrew Johnson
Chief Executive
Officer, ACS

Contents

EXECUTIVE SUMMARY AND RECOMMENDATIONS	6
INTRODUCTION	7
1. DATA SHARING FRAMEWORKS	14
2. IDENTIFYING PERSONAL INFORMATION IN DATASETS	28
3. SAFE DATA AND PERSONAL INFORMATION FACTORS.....	32
4. ADDRESSING THE ‘REASONABLE’ CHALLENGE	46
5. QUANTIFYING THE FIVE SAFES FRAMEWORK.....	54
6. DEALING WITH AI – WHAT HAPPENS WHEN THE “PEOPLE” ARE “ALGORITHMS”?.....	66
7. MAKING IT PRACTICAL.....	74
8. LIMITATIONS OF THE APPROACH.....	80
CONCLUSIONS.....	85
APPENDIX A INTERNATIONAL EXAMPLES.....	87
APPENDIX B INCREASING THE SIZE OF THE MINIMUM IDENTIFIABLE COHORT	94
THANKS	98

Executive summary

This paper describes a framework for privacy-preserving data sharing, addressing technical challenges as well as some data sharing issues more broadly.

The paper builds on the 2017 ACS paper, *Data Sharing Frameworks*¹, expanding the concept of a Personal Information Factor and introducing a Data Safety Factor with recommendations for threshold settings.

The paper speaks to some of the challenges of trusted data sharing. These include concerns with the implications of data quality, use of outputs, the changing risk inherent in the release of results over time, and the need to develop a 'social licence to operate'.

This paper further develops the concept of a quantified 'Five Safes' data analytics framework and briefly examines the implications of such frameworks when artificially intelligent algorithms are used to analyse data. The paper provides a set of recommendations to trial the data sharing framework within the context of developing a national information governance framework.

Recommendations

-
-  1 That the Modified Five Safes Framework described in this paper be piloted for data sharing.
 -  2 That Safe Data sharing be piloted based on the Personal Information Factor, the Data Safety Factor and the data safety thresholds as described in this paper.
 -  3 That Safe Data Level 5 be described as the standard for Open Data.
 -  4 That Safe People and Safe Project frameworks be developed by an independent peak body following widespread consultation. These frameworks can provide the basis to credential individual people to access to data at different safe levels.
 -  5 That a national information governance framework be developed that includes the evaluation of Safe People and Safe Projects, which would be undertaken by appropriate prescribed authorities.
 -  6 That the Data Sharing Framework be exposed to international standards-making bodies as the basis to commence international standards development work.
-

¹ Available online at <https://www.acs.org.au/insightsandpublications/publications.html>

Introduction

Data sharing is a hot topic. People have been actively sharing personal data through online platforms for decades. Since the beginning of the internet and the development of HTTP cookies², we have all been generating data about personal interests and preferences through web browsing and online purchases.

More recently, with the rapid expansion in the number and sophistication of mobile devices, data about movement and service quality has been automatically captured *en masse*. The providers then optimise network performance, create location-based services and plan future network infrastructure.

At the same time, social media has provided companies with unprecedented troves of information about locations, relationships, events, plans, personalities and purchases. The internet of things (IoT) is also adding risk: for example, through normal use, a domestic smart light has the potential to generate data on personal habits, sleep patterns and activity. A service provider that aggregates data from multiple homes may use this to optimise energy consumption at a neighbourhood level and will access derived information on the daily lives of every person who uses the smart light service.

How companies use this information has come under intense scrutiny. It was recently revealed, for example, that Cambridge Analytica³ used explicit personal information to target political campaigning, potentially influencing the outcome of elections. At the same time, online browsing and purchasing data is being used to derive information about preferences and create personal profiles of users, while mobile network data has demonstrated it can go well beyond network optimisation to allow customer churn prediction⁴ and even infer relationships to other mobile users.⁵

In Australia, concerns have also been raised about government use of personal information to better target or improve the efficiency or quality of services. High-profile examples relate to social services debt recovery⁶ and the release of the new structure of My Health Record⁷.

This vast increase in data gathering by governments and businesses has become a significant issue worldwide. The three main mechanisms for data sharing – explicit, derived and inferred – each come with concerns about the degree of personal information contained within them and the obligations of the organisation that captures, uses and stores that data.

In both commercial and government examples, other concerns relate to the unanticipated fidelity of data generated, who will access the data, what it will be used for, and what will happen as a consequence of its use. There are questions about the ‘use’ of data by a company or government, and the ‘release’ of data to the wider world. Questions have also been raised as to whether the use of derived information to create highly targeted ‘anonymous identities’ should come with the same restrictions as use of personal information.

2 See https://en.wikipedia.org/wiki/HTTP_cookie

3 <https://cambridgeanalytica.org/>

4 B. Huang, M. TaharKechadi, B. Buckley, ‘Customer churn prediction in telecommunications’, *Expert Systems with Applications*, Elsevier, 39(1), January 2012, pp. 1414-1425. Available online at <https://www.sciencedirect.com/science/article/pii/S0957417411011353>

5 Bao, Yang, Yan, Luo, Jiang, Tapia and Welbourne, ‘CommSense: Identify Social Relationship with Phone Contacts via Mining Communications’, 2015. Available online at http://alumni.media.mit.edu/~emungua/pdf/CommSense_MDM2015.pdf

6 Paul Karp and Christopher Knaus, ‘Centrelink robo-debt program accused of enforcing ‘illegal’ debts’, *The Guardian*, 2018. Available online at <https://www.theguardian.com/australia-news/2018/apr/04/centrelink-robo-debt-program-accused-of-enforcing-illegal-debts>

7 See <https://www.myhealthrecord.gov.au/>

While these issues are yet to be fully addressed, envisaged future ‘smart services’ for homes, factories, cities, and even governments rely on the sharing of large volumes of often personal and sensitive data between individuals and organisations, or between individuals and governments.

The ongoing benefit from sharing data more easily is the ability to improve the efficiency, quality and degree of service personalisation, as well as optimise service delivery across networks. To deliver these benefits, frameworks for data sharing (as opposed to data release) need to be created that preserve the personal privacy of service users while maximising the utility and benefits.

DATA SHARING WITHIN GOVERNMENT

Governments across the world are struggling to meet citizen expectations and ever-increasing demand for services and infrastructure, particularly in response to growing and ageing populations. There is a drive for easier modes of engagement with government agencies, such as a single point of entry for key data and identity authentication. There is also a need to create smarter, data-driven, personally tailored services, and to use data to underpin better policy and resources allocation.

Despite this, many government data custodians are hesitant to share data. Unvoiced concerns include uncertainty and fear about data sharing and the desire of respective agencies to control data about their own activities. Voiced concerns focus on unintended consequences of sharing data through inappropriate use and interpretation, data quality, the possibility of unauthorised release of data in a manner that might lead to reidentification of affected individuals, and adherence to privacy legislation.

Aggregation of individual data is an approach commonly used to reduce the risk of personal information disclosure within a dataset.

A key challenge for data sharing is that **there is currently no way to unambiguously determine if aggregated data contains personal information** or to determine whether multiple disaggregated datasets can be re-combined to identify individuals through mosaic effects.

Consequently, different techniques and different levels of aggregation of data are used across organisations, depending on a perceived risk associated with the data being shared. The implications of this disparity profoundly affect the ways data can be employed to support different use cases.

Concerns are also being raised by privacy advocates as data-analysing capabilities increase. When the number of data sources used to create and deliver a service or address a policy challenge swell into the hundreds or thousands, the complexity of the problem may rapidly exceed the ability of human judgement to determine whether the integrated data (or the insights generated from them) could be analysed to re-identify affected individuals.



WHAT IS ‘PERSONAL INFORMATION’?

Personal information about individuals within datasets potentially covers a very wide field. Privacy and data protection laws use different definitions of personal information, or as it is referred to in other countries, personal data or personally identifying information (PII).

In Australia, the collection, use, storage and disclosure of personal information about individuals is regulated at the federal level under the Privacy Act 1988. Activities of state and territory governments (and in some states and territories, the activities of private sector organisations handling some health sector data) are regulated under state and territory laws or administrative processes that are specific to each state or territory.

Personal information is defined in the Privacy Act as:

... information or an opinion about an identified 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 a material form or not.⁸

In NSW and some other states and territories, there is a similar, but slightly different definition:

... personal information means information or an opinion (including information or an opinion forming part of a database and whether or not recorded in a material form) about an individual whose identity is apparent or can reasonably be ascertained from the information or opinion.⁹

Other countries have their own definitions. In the European Union, personal data means:

... any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.¹⁰

⁸ Section 6 of the Privacy Act 1988 [Cth].

⁹ Section 4 of the NSW Privacy and Personal Information Protection Act 1998, see http://www7.austlii.edu.au/cgi-bin/viewdoc/au/legis/nsw/consol_act/papipa1998464/s4.html

¹⁰ See <https://www.gdpreu.org/the-regulation/key-concepts/personal-data/>

These definitions each acknowledge that the scope of personal information can be very broad, and they are framed in terms of the ability of any organisation accessing that information to identify an individual, not just whether the relevant data itself identifies the individual.

Key aspects for the purpose of identifying data that contains personal information are that either:

1. *the data itself (directly or indirectly) identifies an individual; or*
2. *the data when combined with other reasonably available information makes it possible to identify an individual.¹¹*

Taking a state-based example, guidance from the Queensland Office of the Information Commissioner states that personal information:

... includes information which directly identifies an individual and information that can be compared or cross-referenced with other information to identify an individual. Appropriately de-identified data is no longer linkable to an identifiable individual, which means it is no longer personal information. Once it is no longer personal information, the IP Act does not apply to the data.¹²

In some states and territories, the definition of 'personal information' also covers information that relates to an identifiable individual living, or to a deceased person within 30 years of their death. At the Commonwealth level, personal information only relates to living individuals.

Determining if data contains personal information depends on the circumstances of the use or disclosure of data and can change depending on factors such as who has access to the data and what other datasets are available.

For example, consider the use of National Metering Identifiers within the national energy market. These do not themselves identify an individual and are maintained within a controlled analysis environment, where steps are taken to prevent them being linked to other data. This may be sufficient to prevent a dataset from being classified as containing personal information. However, if the same data were publicly released or able to be linked to other account or address details, the ability to cross-reference the National Metering Identifiers with physical addresses may cause some of the data to become personal information.

The ambiguity about the presence of personal information in sets of data highlights the limitations of the majority of existing privacy regulatory frameworks. The capacity of human judgement to appropriately apply the regulatory test to determine whether there is a 'reasonable' ability to re-identify individuals from datasets is increasingly limited as those datasets grow in complexity and size.

Developing standards around what constitutes 'de-identified' data (or as it is referred to in the European Union and some other jurisdictions, 'anonymised data') would help address the challenges of dealing with privacy. In all parts of the world, there are currently no objective quantitative measures and only high-level normative guidance to determine when data about individuals is de-identified. This leaves organisations to assess what 'de-identified' means on a case-by-case basis, looking at different datasets and how those datasets might reasonably be used or combined with other data.

Technology can potentially play a role in addressing this challenge. However, agreeing and then communicating what an acceptable degree of anonymisation is, and how to achieve it in quantitative terms, would also greatly improve data sharing. This clarification of existing legal frameworks would benefit from including quantified descriptions of acceptable levels of risk.

¹¹ See Office of the Australian Information Commissioner, 'What is Personal Information?', May 2017. Available online at <https://www.oaic.gov.au/agencies-and-organisations/guides/what-is-personal-information>

¹² See Queensland Office of the Information Commissioner, 'Dataset publication and de-identification techniques'. Available online at <https://www.oic.qld.gov.au/guidelines/for-government/access-and-amendment/proactive-disclosure/dataset-publication-and-de-identification-techniques>

Data sharing is used in the context of a data custodian undertaking analytical projects for the discovery phase, while data release is opening up data to the wider world.

SCOPING USES OF DATA

The concerns around data sharing vary depending on the use of data, from the discovery phase, to policy and service design, to service delivery and evaluation (see Figure 1). The type of data used in each of these phases changes, from historical snapshots in the discovery phase, to transactional data in the delivery phase, and outcomes data in the evaluation phase. This paper will primarily focus on the discovery phase, as this represents episodic (rather than continuous) evaluation of data.

This paper will also distinguish between data sharing and data release. Data sharing is used in the context of a data custodian undertaking analytical projects for the discovery phase, while data release is opening up data to the wider world. The key difference is that data sharing occurs within an environment where uses and applications of the data are subject to controls and safeguards that reliably and verifiably effectively prevent misuse of that shared data and/or combination of that data with other datasets.

Data release exposes data to examination, including possible combination or matching of that data with other data, that may enable individuals to be re-identified. The primary focus of this paper is data sharing (not data release).

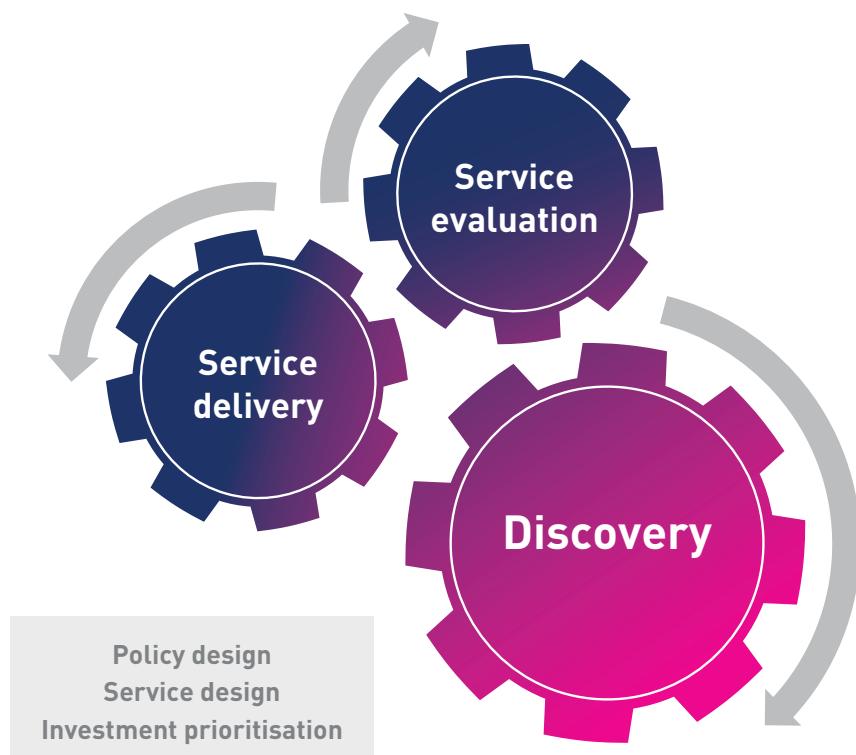


Figure 1. Phases of data use

CONSENT

Consent from individuals to use and share data is an important mechanism in building trust in the design, delivery and evaluation of services. Consent creates awareness of intended use and issues of unintended consequences may be addressed as part of the consent process.

From a personal information context, obtaining the genuine consent of an individual can allow use of datasets containing personal information in accordance with the terms of consent.

However, obtaining genuine consent at an individual level may be challenging, particularly when:

- There are many individuals involved.
- Data has been collected over a number of years.
- Data is collected on the interactions of citizens with government agencies where those interactions are not fully voluntary (for example, obtaining or renewing a business licence). Hence consent may be effectively coerced (and therefore not valid as voluntary) by the requirement for the citizen to deal with government, or by the absence of any reasonably convenient alternative way for the citizen to deal with government.
- Data is collected under government programs with varying stated legislative purposes and those legislative purposes do not match the proposed application.
- The discovery process is targeted at developing completely new services (which were not within contemplation when consent had been obtained) rather than incremental improvements.
- Data is derived or inferred rather than explicitly provided.
- Individuals are vulnerable or in dangerous environments.

The Office of the Australian Information Commissioner has issued guidance on consent to help organisations to interpret the meaning of this term in the context of the *Privacy Act 1988*. The guidance establishes that the four key elements for consent are:

- The individual is adequately informed before giving consent.
- The individual gives consent voluntarily.
- The consent is current and specific.
- The individual has the capacity to understand and communicate their consent.¹³

Valid consent does not need to be expressly given and may be implied by the circumstances, and generally does not require an affirmative action by an individual (such as responding to ‘tick the box’, clicking through via ‘I agree’, or providing a signature), provided that the consent satisfies these conditions.

By contrast, through the GDPR the European Union has introduced an additional requirement that consent be unambiguous. This has generally been interpreted as requiring consent to be signified by an affirmative action of the user. Because of the emphasis placed on genuine consent in the GDPR, significant consideration and effort is involved in obtaining and managing consent processes. In particular, there is an emphasis placed on demonstrating that such consent is both genuine and fully informed. This reflects best practice and should also be adopted in Australia when dealing with datasets that potentially contain personal information.

¹³ Office of the Australian Information Commissioner, APP Guidelines, Chapter B, Key Concepts, discussion of “Consent” at paragraphs [B.34] – [B.58]. Available online at <https://www.oaic.gov.au/agencies-and-organisations/app-guidelines/chapter-b-key-concepts#consent> (accessed 16 September 2018)

To address circumstances where it is impractical to obtain consent but where a particular use of data should be taken to have been within the scope of reasonable application of data, the GDPR¹⁴ introduced a concept of ‘legitimate interests’, where the use of personal data is legitimised through general understanding that a particular use is necessary and the privacy interests of the data subject do not outweigh the interests of the data collector in making a particular use of data about the data subject.¹⁵

The legitimate interests basis for data processing recognises that where information is collected by government or in a service delivery relationship, there is often not a true choice given and it may not be practicable to obtain the genuine and informed consent of an individual. In this circumstance the use should be controlled, confined and not unjustifiably impinge upon the privacy interests of the affected individual. This should be regarded when developing best practice research approaches to data sharing, even where the relevant Australian legislation has not yet reflected these developments.

The focus of this paper is on developing frameworks for data sharing in situations where there is neither express consent nor another clear basis for use of personal information.

OTHER LIMITATIONS

In addition to privacy regulation, it is important to recognise that many statutes impose limitations upon data sharing. In the context of government functions, there are varying legislative restrictions that prevent or restrict the disclosure of information outside a particular government agency (as is inherent in data sharing), including restrictions enacted in the context of:

- (a) National security and limiting disclosure of certain official information.
- (b) Material that has been collected by statistical bodies.
- (c) Material disclosed to a regulator for a specific purpose.
- (d) Welfare, health and social security information obtained for specific service delivery functions.

There may also be additional laws in industry-specific sectors, such as banking and medical services, that restrict the ability to disclose customer information.

This paper focuses on the privacy considerations surrounding personal information about individuals as defined above. When a government is considering the disclosure of datasets (either to the public or as part of a more limited project) it will be necessary to satisfy applicable regulatory restrictions on sharing.

Finally, there are particular additional controls under privacy laws that apply to health data. While the principles set out in this paper apply equally to personal information in a health context, this paper does not consider the additional legislative requirements that apply to such data.

¹⁴ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). A corrigendum to the original text of the GDPR was released in May 2018 and the English language version of the corrected text is available at <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN>

¹⁵ See the U.K. Information Commissioner’s Office Guidance on legitimate interests. Available online at <https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/>

Data sharing frameworks

Sharing data relating to individuals safely, storing it securely and ensuring it is only accessed and used by an approved user is a global challenge.

Sharing data relating to individuals that has been de-identified through replacement of personal identifiers with linkage code and is protected from re-identification through prudent management of a data analytics environment poses particular challenges. The challenges lie in managing a data environment based on de-identification and ensuring the perimeters of that environment are reliably and verifiably effective.

A focused effort is required to achieve safe and efficient machine-readable access and use of data assets across national or state boundaries, or within discrete domains of activity such as education or health. This is necessary to address the technical, privacy, regulatory and information governance challenges of data sharing. There is also growing awareness of the need to develop the trust citizens have in data, or as it is described in the context of the New Zealand's community engagement, a 'social licence' and 'trusted data'.¹⁶

In Australia, many mechanisms can be identified that enable data sharing between parties. However, to date, each has been developed to address specific needs or a particular regulatory framework, without reference to a nationally accepted information governance framework.

Without such an information governance framework, the protocols of acceptable use of data and release of outputs must be determined on a case-by-case basis. This often relies on sector-specific past practice as a precedent to determine what is acceptable, rather than a methodologically clear and justified approach.

Working on a case-by-case basis has the effect of placing undue weight and reliance upon past practice (which is often not a reliable guide of future exposures and risks), requires lengthy negotiation of each new data use type, demands mapping of complex agreements to contracts or memoranda of understanding, and often limits data sharing to highly aggregated or highly perturbed data. Consequently, there is limited data sharing between jurisdictions or between government agencies within jurisdictions.

The Commonwealth government is currently undertaking consultation around future data sharing and release legislation.¹⁷ It is anticipated this legislation will be principles-based, much like existing privacy legislation, and also facilitate further development of detailed rules as to the implementation of those principles.

Sharing large quantities of people-centred data to create smart services requires robust data sharing frameworks that preserve privacy and ensure proper evaluation of outputs before implementation.

¹⁶ The Data Futures Partnership in New Zealand undertook an exploration of social licence, engaging with thousands of New Zealanders. The Partnership defined social licence as follows: 'when people trust that their data will be used as they have agreed, and accept that enough value will be created, they are likely to be more comfortable with its use. This acceptance is referred to as a social licence.' The Partnership summarised its key conclusions in 'A Path to Social Licence: Guidelines for Trusted Data Use'. Available online at <https://trusteddata.co.nz/>. The Guidelines focus on eight key questions that organisations can answer to explain how they collect and use data, to better build trust with clients and the wider community.

¹⁷ See <https://www.pmc.gov.au/resource-centre/public-data/issues-paper-data-sharing-release-legislation>

Such data frameworks will need to provide the necessary guidance and direction to enable all actors in a data-driven process – data custodians, analysts, data governance staff, managers and service providers – to understand how to meet their obligations and respect limits of acceptable use.

These frameworks, the manner of their implementation and management and the safeguards for quarantining of outputs for human evaluation, must be sufficiently transparent and understood by citizens to mitigate risk. The views of government and its agencies as to their good intentions are likely to be contested by at least some citizens. An effective data sharing framework must contain controls and safeguards that can be demonstrated to citizens as reliable and effective.

An efficient authorising environment could be managed consistently through a nationally accepted information governance framework, designed to guide the regulators, data owners and data custodians in a practical way. This framework could help clarify the risks at each stage of the data analysis process and provide appropriate transparency to citizens.

For the authorising environment to be truly effective, it needs to comply with an appropriate information governance framework that demonstrates transparency, trust, efficacy and value.

A MODIFIED FIVE SAFES FRAMEWORK

In September 2017, ACS released a technical whitepaper, *Data Sharing Frameworks*, that explored the challenges of data sharing.¹⁸ The paper highlighted that one fundamental challenge for the creation of smart services is addressing the question of whether a dataset contains personal information. Determining the answer to this question is further complicated as the act of combining datasets creates information. The paper proposed a modified version of the Five Safes Framework¹⁹ for data sharing that attempts to quantify different thresholds for 'Safe.'

The 2017 whitepaper introduced several conceptual frameworks for practical data sharing, including an adapted version of the Five Safes Framework. Many organisations around the world, including the Australian Bureau of Statistics, use the Five Safes Framework to help make decisions about effective use of data which 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 this context, 'appropriately' means 'in accordance with the required standards of behaviour', rather than level of statistical skill. 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 increase 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.

18 Available online at <https://www.acs.org.au/insightsandpublications/publications.html>

19 T. Desai, F. Ritchie, R. Welpton, 'Five Safes: designing data access for research', October 2016. Available online at <http://eprints.uwe.ac.uk/28124/1/1601.pdf>

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 Settings encompass 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, the quality of the data (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 publishing sensitive data.

The Five Safes Framework is relatively easy to conceptualise when considering the idea of 'Extremely Safe', although it does not unambiguously define this. 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 each involves judgement.

The adapted model explores different, quantifiable levels of Safe for each dimension of People, Projects, Setting, Data and Outputs and how these different Safe levels could interact in different situations. Figure 2 shows the dimensions of the adapted Five Safes Framework taken from the 2017 ACS *Data Sharing Frameworks* whitepaper.

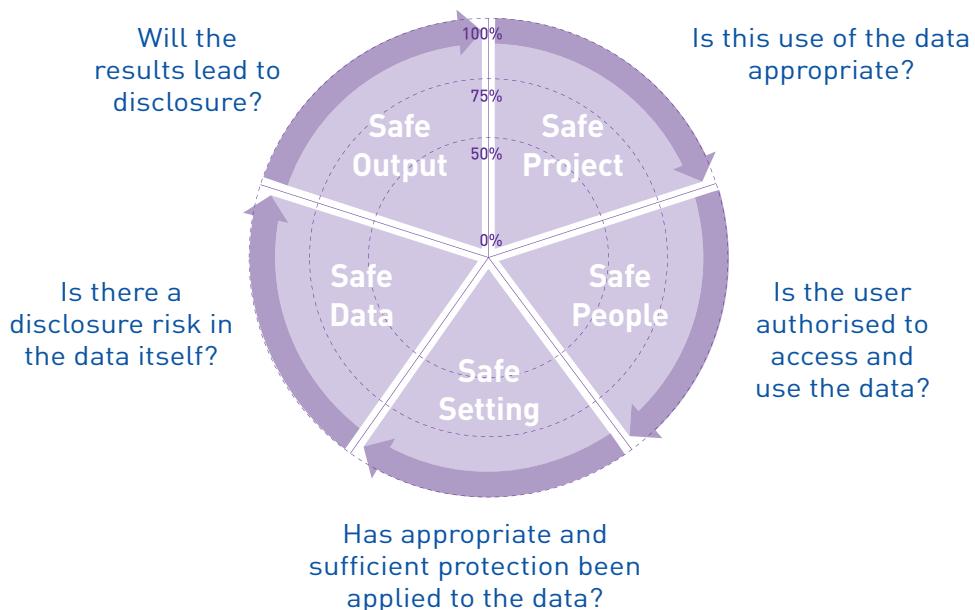


Figure 2. Modified Five Safes Framework

PUTTING THE FIVE SAFES FRAMEWORK INTO A LARGER CONTEXT

The Five Safes Framework provides a useful conceptual model for an individual project but falls short of describing the ongoing context within which a project typically takes place and does not adequately address what happens with outputs once a project is complete. Further, it seems better suited to episodic (case-by-case) projects rather than continuous analysis, where outputs would drive new projects.

One of the main challenges posed repeatedly during the development of this whitepaper was the distinction between outputs and outcomes. When an analytics project delivers a result (an output), what safeguards are there for the initial use of that output and/or how that output may subsequently be used to deliver an outcome that has a relevant effect upon how citizens are treated? Are limitations in the outputs that affect their reliability to guide or effect outcomes understood by the decision-makers who will apply those outputs to deliver outcomes? Will an outcome be intermediated and evaluated by humans? Will an outcome be ongoing (continuous) or discrete (singular)? Are envisaged outcomes reliably fair, reasonable and unbiased?

Secondly, an individual researcher (or team) may be considered to have a certain Safe Level; however the way the data, settings and outputs interact with those people for the project are bound by the systems, processes and governance of the organisations within which the project takes place. In this context, there is a large degree of correlation and feedback between the dimensions of Safe People and Safe Setting. A Safe Organisation has demonstrable controls, processes and culture of adherence to quality, security and safety.

Finally, there is a need to consider the varying value of an output over time. The value of data is a complex issue in its own right because of the multitude of purposes for which a dataset may be used.²⁰

In many cases, the value of the project output is related to the risk associated with uncontrolled release. Different levels of governance and controls may be employed as the value (and therefore risk) changes over time, depending on whether it decreases, increases or remains constant. Time-sensitive information identified in an output may initially have high value and decrease in value as it becomes public through other means; for example, information on land rezoning.

Information identified in an output may also increase in value over time if it relates to rare events or resources, such as the location of rare natural resources. Information identified in an output may remain constant over time if, for example, it relates to something which is in constant, steady demand, such as an analytically derived international standard.

Accordingly, three additional dimensions have been proposed for the Five Safes Framework to reflect the organisational, outcome-focused and time-varying nature of the value of project outputs:

SAFE ORGANISATION

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

²⁰ The issue of value was tackled in the 2017 ACS *Data Sharing Frameworks* technical whitepaper but remains a topic for further investigation, in part because of the multiple uses of data and the differing ways of describing value. For an example framework that addresses value in a commercial context, see D. Laney, 'Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage', 2018.

SAFE OUTCOMES

Refers to the ultimate uses of the project outputs. A variety of outcomes frameworks have been developed that can be informed by the outputs of individual data linkage and analysis projects. An example framework is the Human Services Outcomes Framework developed by NSW Government. Figure 3 provides a high-level summary of that framework, which is underpinned by a range of quantitative and qualitative parameters.²¹

SAFE LIFECYCLE

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

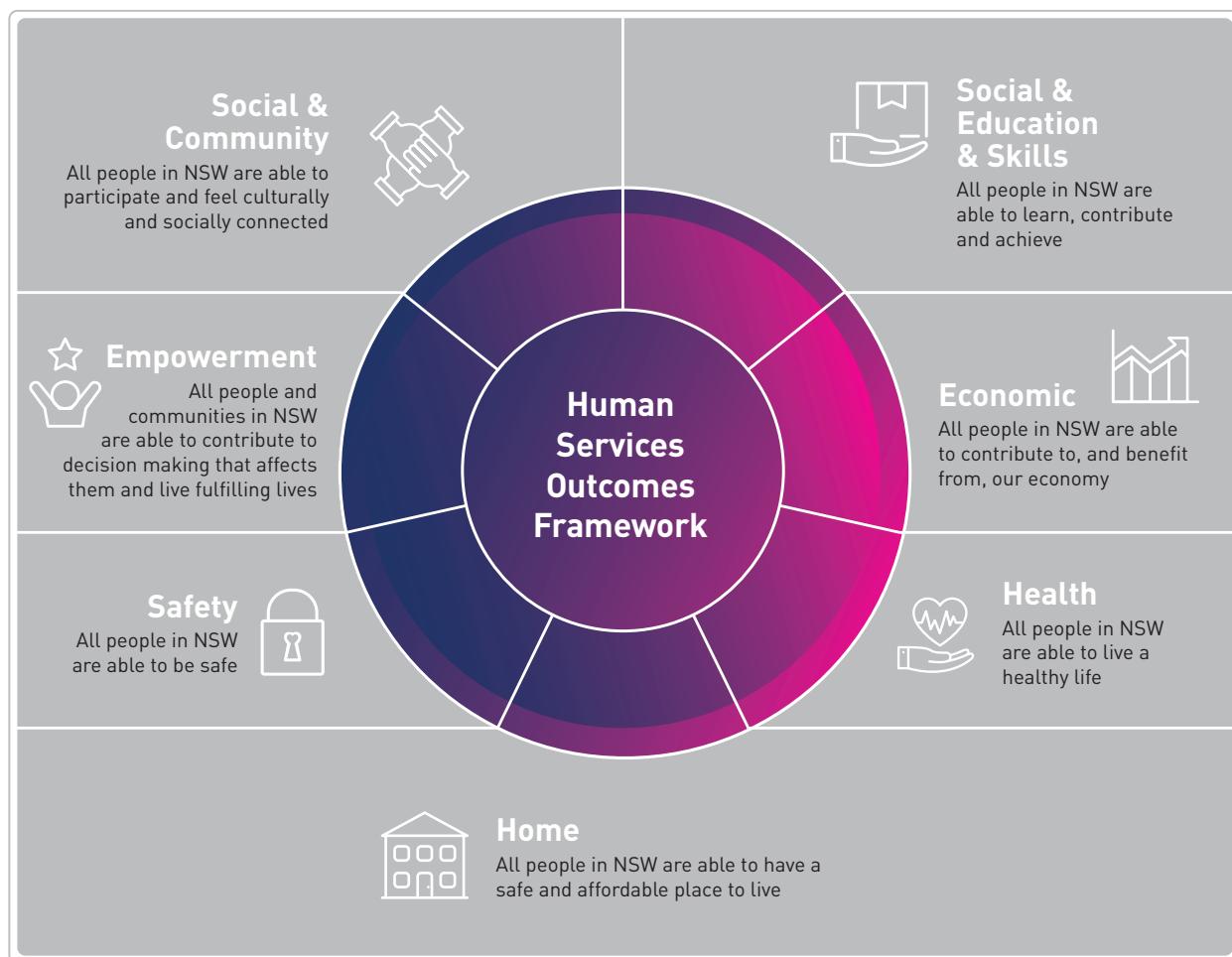


Figure 3. Example of a Safe Outcome Framework – the NSW Human Services Outcomes Framework

21 See NSW Innovation website, https://www.finance.nsw.gov.au/human_services

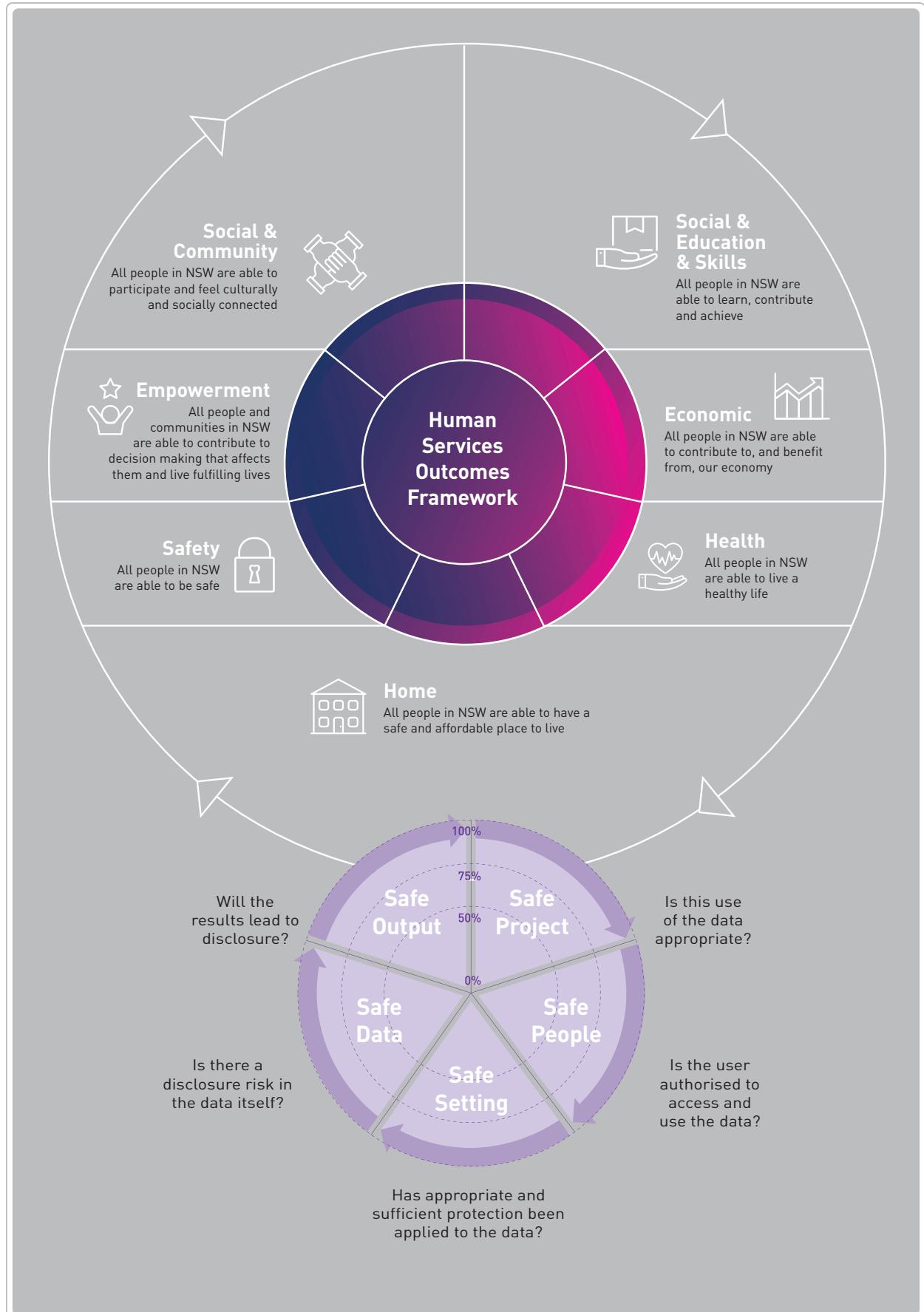


Figure 4. Example of the Modified Five Safes Framework interacting with a Safe Outcomes Framework

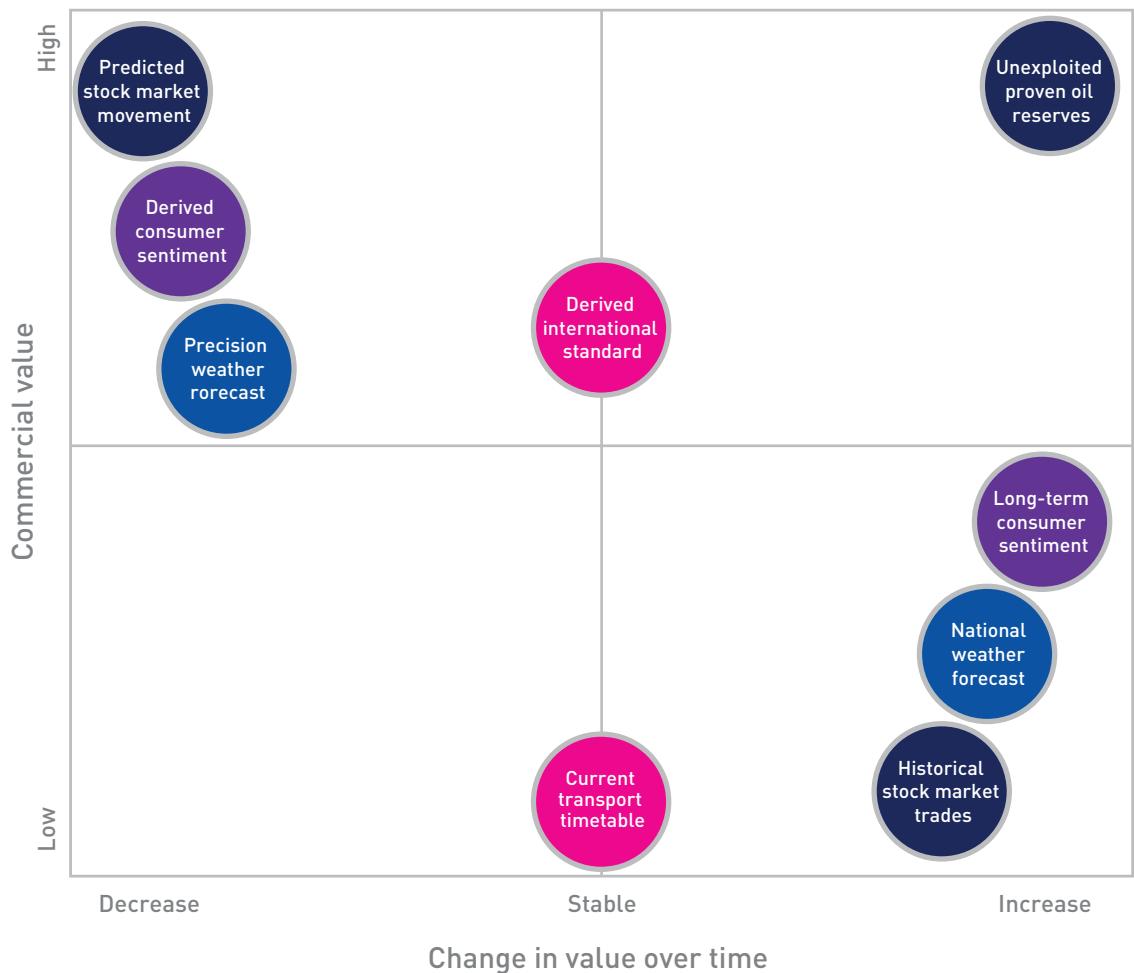


Figure 5. Examples of different outputs with changing value over time

Figure 4 shows one possible relationship between the Five Safes Framework, Safe Organisations and Safe Outcomes. Safe Outcomes interact with Safe Projects by providing justification, specifications and mandates. Safe Outputs are then used within the Framework to inform policies and strategy, help refine service offerings and provide direction for future projects.

Safe Organisation is shown in Figure 4 covering all aspects of the Fives Safe Framework.

Figure 5 highlights different outputs from analytics projects with different possible commercial values (to the data holder) and how these will trend over time. Each of the examples in Figure 5 could be argued to be more or less commercially valuable depending on exact content of the output (such as prediction of extreme weather conditions); hence, these examples should be considered as merely illustrative. The 2017 whitepaper explored the wider relationship of value of data and outputs, identifying up to eight elements of value for companies, individuals and government.



EXTERNAL FACTORS AFFECTING WILLINGNESS TO SHARE DATA

The variable value of data and outputs over time implies that different levels of control and governance are required at different stages of the data lifecycle. Without external factors, it would be expected that controls and governance would track with the increased or decreased assessed value (or risk) over time.

The willingness to authorise the release of potentially risky data is impacted by a number of factors. While the data itself does not change, the treatment and controls applied through the use of the Five Safes, as well as external events and behaviours, will impact the willingness of individuals and data custodians to release data and related outputs.

This behaviour, and differing appetites for risk versus reward, is further complicated when comparing personal benefits with commercial gains from the release and use of data and outputs. For example, in an emergency such as natural disaster, there is an increased willingness to share sensitive, high-value data, even where there has been no change in the underlying risk or the personal information within the data. In fact, criticism is often made of response agencies that there is insufficient high-quality data sharing to coordinate an effective response. A notable example was the response to Hurricane Katrina in the USA in 2005, where the lack of coordination by response agencies was underpinned by a lack of effective data sharing between agencies.²²

The willingness to release data and outputs (and assume associated risk) also arises from perceived short-term opportunities where the perceived benefit outweighs the perceived risk. Many people have received a spam email claiming to be a Nigerian Prince or foreign government official with an offer of millions of dollars in exchange for agreeing to an urgent business engagement. This engagement typically requires the recipient to provide sensitive, high-value data in the form of bank details, date of birth and full name, in order to receive funds.

Whilst a reasonably transparent scam scenario, it is estimated that in Australia alone more than \$89 million²⁵ was lost from 8,000 individuals deceived by this type of scam in the period between January and September 2018. On the more positive side, individuals or companies may re-examine their risk appetite for release of valuable data or outputs for genuine short-term opportunities, such as investment and building trusted relationships.

22 A. Chua, S. Kaynak and S. Foo. 'An analysis of the delayed response to Hurricane Katrina through the lens of Knowledge Management', Journal of the American Society for Information Science and Technology, 58(3), 2007, pp. 391–403. Available online at http://www.ntu.edu.sg/home/sfoo/publications/2007/2007JASIST_fmt.pdf

23 Current statistics can be found at <https://www.scamwatch.gov.au/about-scamwatch/scam-statistics>

BUILDING AND MAINTAINING TRUST

One of the key remaining challenges is the broader acceptance by the community of the public value of an analytical project's outputs and outcomes. This includes the willingness of the broader community to accept the integration and use of different types of data, the acceptance of the judgment of the level of Safeness of projects, the acceptability of the use of outputs, and the appropriateness of outcomes frameworks. This process has of late been referred to as social licence, or social capital. In practice these terms are misleading as ultimately the issue is building and maintaining trust related to appropriate and safe use of data, and appropriate and safe use of outputs of that data.²⁴

Many organisations use proxies for consideration of data trust in the form of statutorily appointed privacy commissioners, internal or external ethical review committees, risk and audit committees or even external governance groups. None of these represent a forum for evaluating data trust or a mechanism for engaging the broader community to understand their needs, expectations and acceptance of data sharing.

24 For a functional definition on social licence, see <https://sociallicense.com/definition.html>

There is cost to organisations in building and maintaining data trust of citizens and consumers. Traditionally the healthcare sector has been a leader in consumer engagement and consent-driven participation in health trials. Health information is subject to additional protections under legislation including the Privacy Act 1988 (Cth) and under separate State and Territory based legislation such as the Health Records and Information Privacy Act 2002 (NSW). While placing additional burdens on the Healthcare sector to appropriately manage personal information, and generally preventing the use of personal information of a health nature in research without consent, this legislation has driven the sector to codify appropriate use of data and outputs through specific processes.

Health and other sectors have also successfully sought and gained authorisation to share people's personal data by seeking specific and informed consent from participants. Through the use of informed consent to share and use people's data, an individual is making an informed choice on the personal benefit versus the risk, as well as the potential social good of sharing their personal data. The ability of organisations to gain and maintain trust is an essential consideration for their ongoing ability to operate, as people choose whether or not to share their information with an organisation..

Advances in affordable high-performance computing technology and developments in machine learning and artificial intelligence (AI) are driving greater demand for data. With greater reliance on algorithmic driven decision making, the need to maintain social acceptance and support faces greater challenges in engaging and educating the community.

In light of ethical and privacy considerations, building trust and confidence in the data being accessed, integrated and used requires an understanding that the data itself is value-laden.

It can also be argued that ethical frameworks are context-laden. In light of the capabilities of big data analytics and the associated acceleration of data access and use, a national data governance framework is required to address data quality and information quality derived from integrating people-centred data.

INDIGENOUS DATA SOVEREIGNTY

Ways to develop trust and confidence in data sharing is not uniform across the whole community. As a result of different communities' requirements, and in light of the potential social good, the question of trust and support must be considered from different communities' perspectives and needs.

The following section considers the perspective of indigenous peoples, and the growing sensitivity of data sovereignty principles as they apply to other communities within diverse multicultural societies.

In Australia and many other nations, there is the essential need to acknowledge and consider in a thoughtful and respectful manner the effects, impact and possible bias in the data resulting from colonisation of indigenous peoples and the resulting cultural impact.

Developing trust and support to share data must consider the community's priorities, perspectives, and expectations in all stages of the data lifecycle, including collection, sharing, use and outputs. This is true for both indigenous and non-indigenous communities. The respectful collection and use of data on indigenous peoples and communities needs to consider issues arising from colonisation and dispossession and focus on empowerment of indigenous communities through data. This may be achieved by working in partnership and seeking reconciliation, based on the principles of indigenous data sovereignty. Internationally there is a growing and important need for data sharing to be respectful of indigenous peoples, with clarity provided by evolving indigenous data sovereignty principles.

INDIGENOUS DATA SOVEREIGNTY IN AUSTRALIA

The Maiam nayri Wingara Indigenous Data Sovereignty Collective and the Australian Indigenous Governance Institute met in Canberra in June 2018²⁵ and confirmed that data is a cultural, strategic and economic asset for Indigenous peoples, stating that Indigenous Australians have the right to:

- Control in the data ecosystem, including creation, development, stewardship, analysis, dissemination and infrastructure.
- Data that is contextual and disaggregated (available and accessible at individual, community and First Nations levels).
- Data that is relevant and empowers sustainable self-determination and effective self-governance.
- Data structures that are accountable to indigenous and First Nations peoples.
- Data that is protective and respects individual and collective interests.

²⁵ Maiam nayri Wingara Indigenous Data Sovereignty Principles are available online at [http://www.aigi.com.au/wp-content/uploads/2018/07/Commrique-Indigenous-Data-Sovereignty-Summit.pdf](http://www.aigi.com.au/wp-content/uploads/2018/07/Communique-Indigenous-Data-Sovereignty-Summit.pdf)

ENGAGING WITH INDIGENOUS COMMUNITIES IN ALBERTA

A practical example of indigenous engagement is the Alberta First Nations Information Governance Centre (AFNIGC), which promotes and advances indigenous data sovereignty in research and information management.²⁶

The power of data from the perspective of Bigstone Cree Nation Chief Gordon T. Auger in Alberta, Canada states, 'quality information is often a key catalyst for change in First Nations communities... Nobody will give you anything without information.'²⁷

The ability of research analysts and organisations globally to be respectful when working in partnership with indigenous communities and individuals is important. Seeking priority setting from the community perspective is essential, combined with the ability to understand and focus available resources on addressing the community-led priorities and initiatives.

Consequently, effective relationships with indigenous communities and indigenous leaders is critical, along with the need for the community to be actively involved in safe project design and project outputs. Trust is critical when sharing and using indigenous data. In light of the indigenous data sovereignty principles, there is a requirement to feed project outputs back to communities.

Engaging with indigenous people requires targeted and culturally sensitive processes for communicating and explaining project outputs. Importantly, respect for indigenous data sovereignty means first liaising with a community to discuss which information is to be accessed, what the community wish to be released or shared, how it will be presented and the words to be used in its presentation.

Nurturing a lasting relationship between the research analyst and the members of the community is an important part of building and maintaining trust. The establishment of such an engagement capacity would potentially have long-term impacts for future data collection, analytical and research activities.

DATA LINKAGE INFRASTRUCTURE IN AUSTRALIA TO SUPPORT TRUSTED DATA SHARING FRAMEWORKS

Targeted engagement, communication and education activities can be designed and implemented relatively easily within single organisations or jurisdictions. When expanding across jurisdictions or addressing communities with a wider range of interests, conveying the social benefit of greater sharing and use of people-centred data becomes substantially more difficult. This has been shown to be a significant challenge in Australia.

A key limitation to the existing data linkage infrastructure in Australia is the time required to identify, cleanse and link data, and the inability to access high-quality linked data in near real-time. Significant resources are spent and significant delays experienced in reviewing, matching and validating poor quality records. Different approaches with greater automation are needed.

In Australia, data sharing for research and analysis has been generally restricted to discrete jurisdictions or domains of activity. At present, Australia does not have an effective, nationwide, federated data sharing framework. Consequently, there has been an inability to easily share, access and integrate data across jurisdictions.

26 The Alberta First Nations Information Governance Centre (AFNIGC), <http://afnigc.ca/main/index.php?id=home&content=home>

27 Bigstone Cree Chief Gordon T. Auger's perspective on power of data can be found online at http://afnigc.ca/main/includes/media/pdf/news/FNIGC_PoD_Series-Bigstone_FINAL_SCREEN.pdf

AUSTRALIAN EXAMPLES OF DATA LINKING

The Population Health Research Network

Population-based data linkage infrastructure is seen as an important strategic asset for Australia, enabling data access for research and ethical analysis while protecting people's identity and privacy. The Population Health Research Network (PHRN)²⁸ received Australian Government funding under the National Collaborative Research Infrastructure Strategy (NCRIS) from 2009.²⁹

The NCRIS strives for research excellence and application. Through the PHRN, NCRIS has co-funded nationwide collaboration between the university and government sectors, resulting in data sharing infrastructure servicing all states, territories and the Commonwealth.

The PHRN funding decision by the Australian Government was driven by decades of pioneering data linkage work in Western Australia led by Michael Hobbs, Fiona Stanley, D'Arcy Holman, Di Rosman, and John Bass. The Western Australian work demonstrated the feasibility and best practice protocols for bringing disparate datasets together for analysis, and the resultant value to the community from safely sharing data across the research and government sectors.³⁰

SA NT Datalink

In 2009, SA NT DataLink was established using the separation principle pioneered in Australia by the work in Western Australia.³¹ SA NT DataLink operates as a trusted third party, facilitating data access and use and ensuring the necessary approvals and controls are in place to access integrated people-centred data.

Legal agreements with data owners are an important component of the data linkage infrastructure, providing the authority to add data into an enduring master linkage file, along with the commitment by data providers to consider, on a project-by-project basis, proposals for linking data for policy analysis, research and program evaluation and monitoring. The master linkage file is a population spine which has the links to individuals' records existing across multiple data sources.

The decision to establish SA NT DataLink was based on a pilot data linkage project in South Australia, considering the 'Clients in Common' from the people-centred data from hospitals, mental health, public housing, disability, youth and family services records. From the success of this pilot and the support of NCRIS funding, SA NT DataLink facilitated safe access to data from Health, Education, Social Housing, Youth Justice and Child Protection government departments.

In 2009

SA NT DATALINK WAS
ESTABLISHED USING THE
SEPARATION PRINCIPLE
PIONEERED IN AUSTRALIA BY THE
WORK IN WESTERN AUSTRALIA.

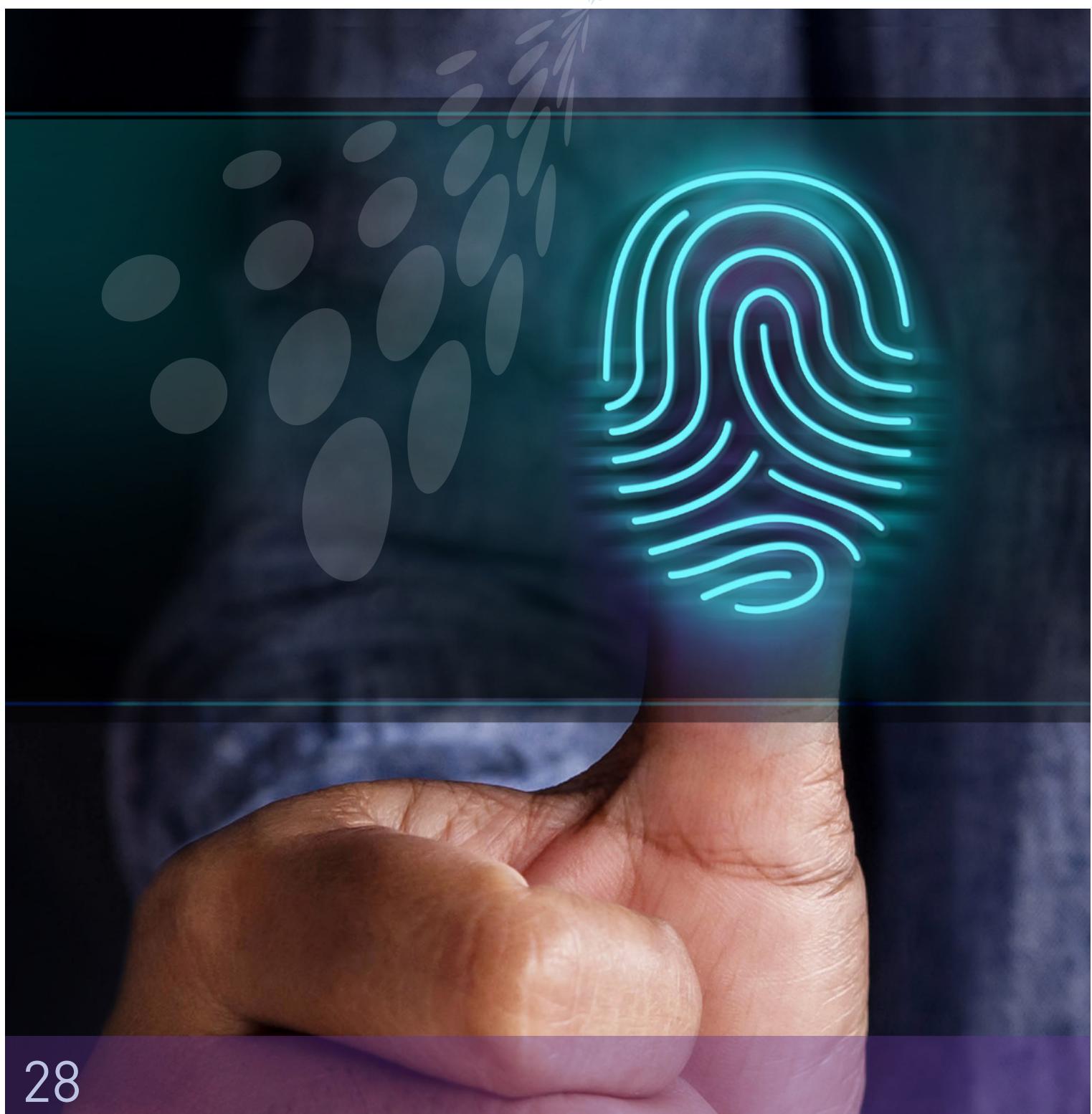
28 The Population Health Data Linkage Network, <https://www.phrn.org.au/>

29 The Australian Government National Collaborative Research Infrastructure Strategy (NCRIS), <https://www.education.gov.au/national-collaborative-research-infrastructure-strategy-ncri>

30 C. Kelman, A. Bass and D. Holman, 'Research use of linked health data - a best practice protocol', Aust N Z J Public Health 200, 26(3), pp. 251–5. See also the Report of the WA Data Linkage Expert Advisory Group, 'Developing a whole-of-Government data linkage model: a review of Western Australia's data linkage capabilities', December 2016. Available online at <https://www.itsi.wa.gov.au/docs/default-source/default-document-library/a-review-of-western-australia-s-data-linkage-capabilities---developing-a-whole-of-government-model---december-2016.pdf>

31 A definition of the separation principle can be found at https://www.santdatalink.org.au/Privacy_Protecting_Model

02



Identifying personal information in datasets

IS PERSONAL INFORMATION PRESENT IN DATA?

As outlined in Chapter 1, datasets that do not identify particular individuals may still be used to create personally identifiable information if other datasets are accessible that reasonably enable identification of the individuals to whom the shared datasets relate. This other information might be available either:

- Internally – for example, by looking up another dataset and cross-matching transaction data sorted by transactor key or device identifier; or
- Externally, such as re-identification of individuals through matching of datasets in searchable databases such as ASIC records, Land Titles Office property records or through search engines.

It may also be that another entity might hold the same datasets, but:

- Without other internal datasets which would enable identifying lookups; or
- Subject to safeguards and controls that are effective to prevent access to external identifying information.

Such an entity would not hold personal information about identifiable individuals. However, if that entity elected to share or release (disclose) that data in circumstances where recipients could reasonably re-identify an individual within that released dataset, the entity would have disclosed personal information about individuals, even though they have shared or released a dataset that appears to be de-identified.

Accordingly, determining whether datasets relating to individuals that are not expressly identified contain personal information requires a context-specific inquiry as to who holds the relevant information and the nature of relevant identification reasonably available to that entity.

An entity releasing information in purportedly de-identified form must consider the nature and extent of other information available and potentially useable by reasonably anticipated future recipients.

Such an inquiry must undergo two stages:

1. Is personal information about individuals present in the dataset, having regard to other potentially identifying information reasonably available; and
2. Is personal information about individuals present in the dataset, having regard to other potentially identifying information reasonably available to any future anticipated recipients of that dataset.

A PERSONAL INFORMATION FACTOR

For the purposes of this document we will use a hypothetical parameter, the Personal Information Factor (PIF) to examine the likelihood of the dataset containing personal information. The PIF considers the:

- Personal information content of each of the individual datasets used to create a ‘service’ (the simplest service may be data sharing).

- Functions that operate on the datasets (such as logical operations or other processing) to produce insights and models.
- Individual knowledge of the observer/user of the data of the insights or models.
- Additional information available to the observer that the observer could bring to the insights or models.

The personal information content of each of the individual datasets and the PIF remain to be defined as discussed below.

Figure 6 shows the context for evaluating the degree of personal information as part of assessing the PIF in a closed system, taking into consideration only the first two factors outlined above for the PIF.

As an example, consider an information service that determines the number of people who arrive at each train station in NSW, for each hour of the day, for different passenger types (student, pensioner, adult). Using de-identified input datasets, such a service may deliver the insight that on certain days, at one regional station, there is only a single pensioner who alights between 6:00pm and 7:00pm.

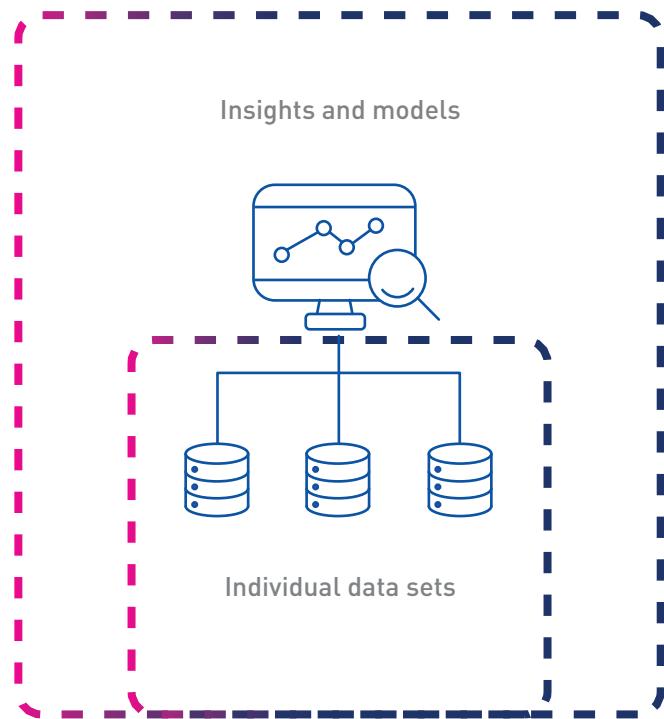


Figure 6. Closed system context for evaluating PIF

Figure 7 shows the context for evaluating the degree of personal information when considering the observer's own knowledge of the world. Extending the example above, if the observer has personal knowledge of the regional station identified and knows several pensioners who live nearby and travel by train, the PIF associated with insight produced by this service is increased.

Figure 8 shows the framework for considering PIF when additional information can be brought into the context of information/data that has been shared. Extending the example above, if the observer has personal knowledge of the regional station identified and knows several pensioners who live nearby and

travel by train and this observer waits at the station on the days the individual is known to travel, then the PIF associated with insight produced by this service is increased to the point where the individual travelling pensioner can be identified. Specifically, the PIF can be brought to 1 (100% personally identifiable).

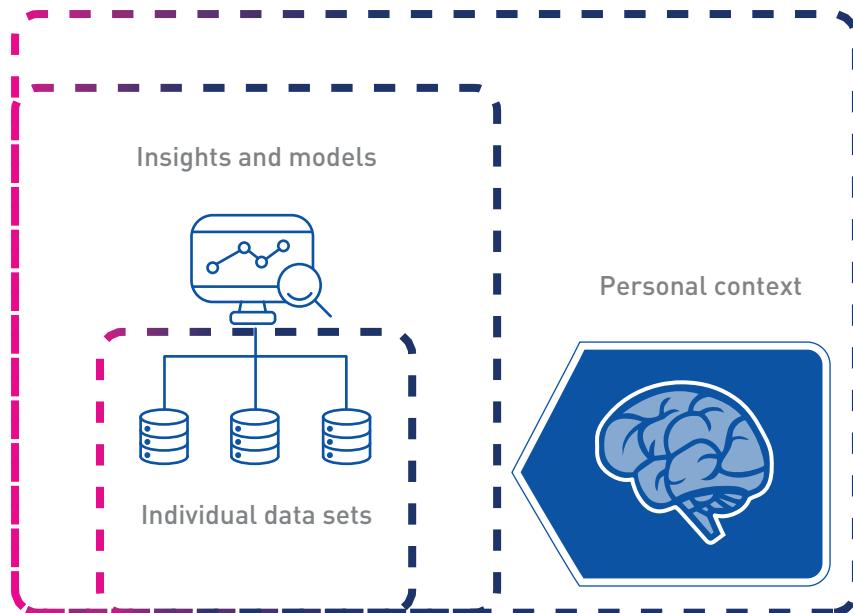


Figure 7. Human context for evaluating PIF

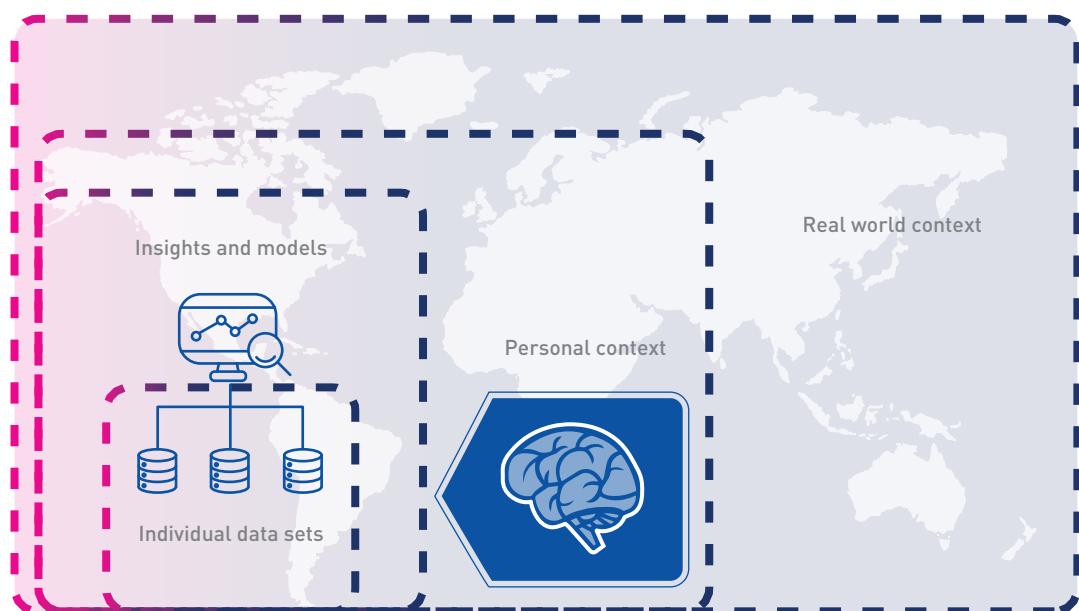


Figure 8. Real world context for evaluating PIF

03



Safe data and personal information factors

HOW SAFE IS A DATASET?

The aspects of Safe Data that were described earlier primarily focus on the risk of re-identification but include aspects of the quality of the data (accuracy), the conditions under which it was collected, the percentage of a population covered (completeness), the number of features included in the data (richness), and the sensitivity of the data. Figure 9 illustrates the different aspects that will be considered in this paper to determine the Safe level of data.

The legal tests for personal information generally relate to the situation where an individual identity can ‘reasonably be ascertained’. The 2017 ACS *Data Sharing Frameworks* technical whitepaper uses a concept of Personal Information Factor (PIF) to describe the level of personal information in a dataset or outcome as shown in Figure 10. A PIF of 1 means personal information exists, a value of 0 means there is no personal information.

Personal information that includes health information is excluded from the scope of this paper. It is important to note the PIF method described is not a technique for anonymisation: rather, it is a heuristic measure of potential risk of re-identification.

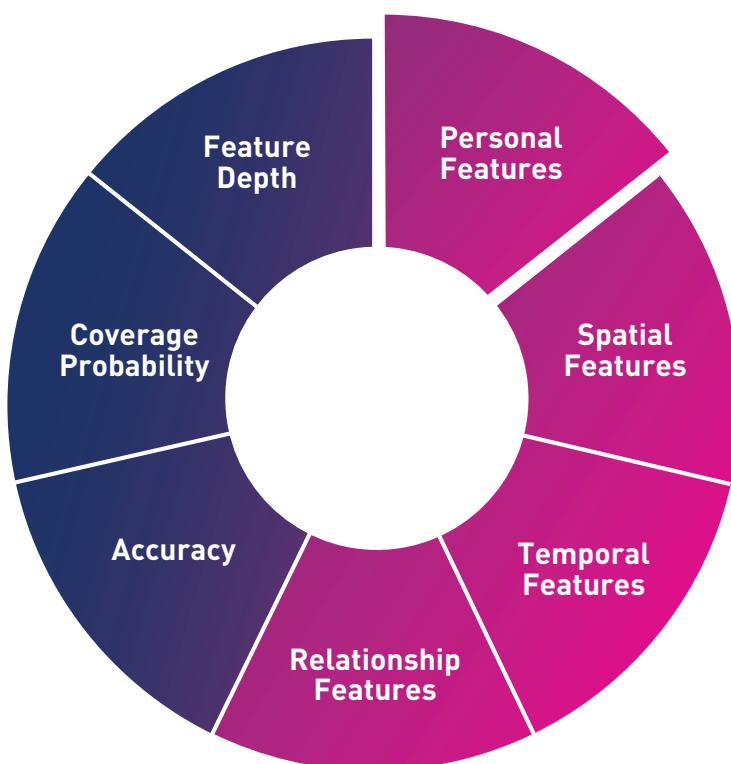


Figure 9. Aspects of Safe Data including Personal Information Factors

In this paper:

- **Feature depth** is the number of independent features in the dataset. For example, in the binary valued feature set: *eye_colour_is_brown*, *individual_is_adult*, *gender_is_female*, the feature depth is 3. If one of these features is dependent on another, or can be derived from a combination of features, the feature depth would be 2 (for example, a pregnancy feature may also enable the gender feature to be derived). An implicit assumption is that all features carry equal information for an individual. Also, the sensitivity of each feature is not considered.
- **Coverage probability** is the probability that an individual is in the population included in the dataset. In a closed analytical environment, with randomly selected samples and no other information available, this is taken to be the percentage of the entire population covered by the sample dataset. For example, a sample dataset of ten men with beards taken from a known population of 1,000 men with beards has a coverage probability of 1/100. If an individual is known to be in a dataset, either because the data was not selected randomly or the sample set covers the entire population, the coverage probability is 1.
- **Accuracy** refers to the ratio of the number of correct values in all features in the dataset to the number of all values for all features in the dataset. For a sample population of eight individuals with ten features each, of which 20% of the values were known to be wrong for one feature, the accuracy is 0.98. If a second feature was known to have 20% of incorrect values, the accuracy would drop to 0.96. No consideration is given to values which are almost correct. This is discussed further in a subsequent section.

Personal, spatial, temporal and relationship features will be discussed in greater detail in a subsequent section.

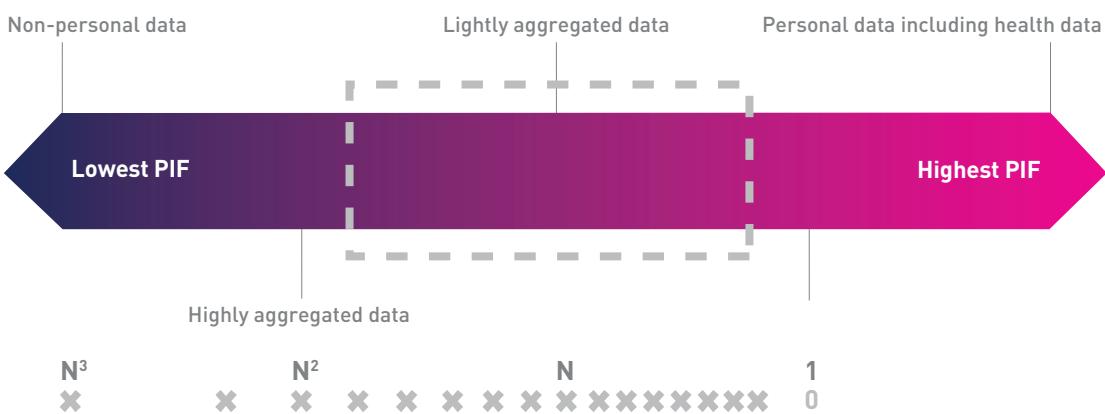


Figure 10. Personal Information Factor and aggregation level

DEFINING A PERSONAL INFORMATION FACTOR

Aggregation is often used to protect individual identity, ensuring outputs are not released for cohorts smaller than 'N'. The value of N depends on the risk appetite of the organisation and the perceived sensitivity of the data itself.

In principle, for any value of N selected, if $(N-1)$ other datasets can be found that relate to the cohort of interest, then the cohort of size N can be decomposed into identifiable individuals. As the aggregation levels increase (cohort sizes of N, N^2 , N^3 and so on for $N \rightarrow 1$), the level of protection increases, as more related datasets are needed to identify an individual within the cohort. The fundamental weakness nonetheless remains that determining N is dependent on the risk appetite.

The definition of PIF is still to be robustly determined; however, the working definition is upper-bound and defined within a closed, linked, de-identified dataset as:

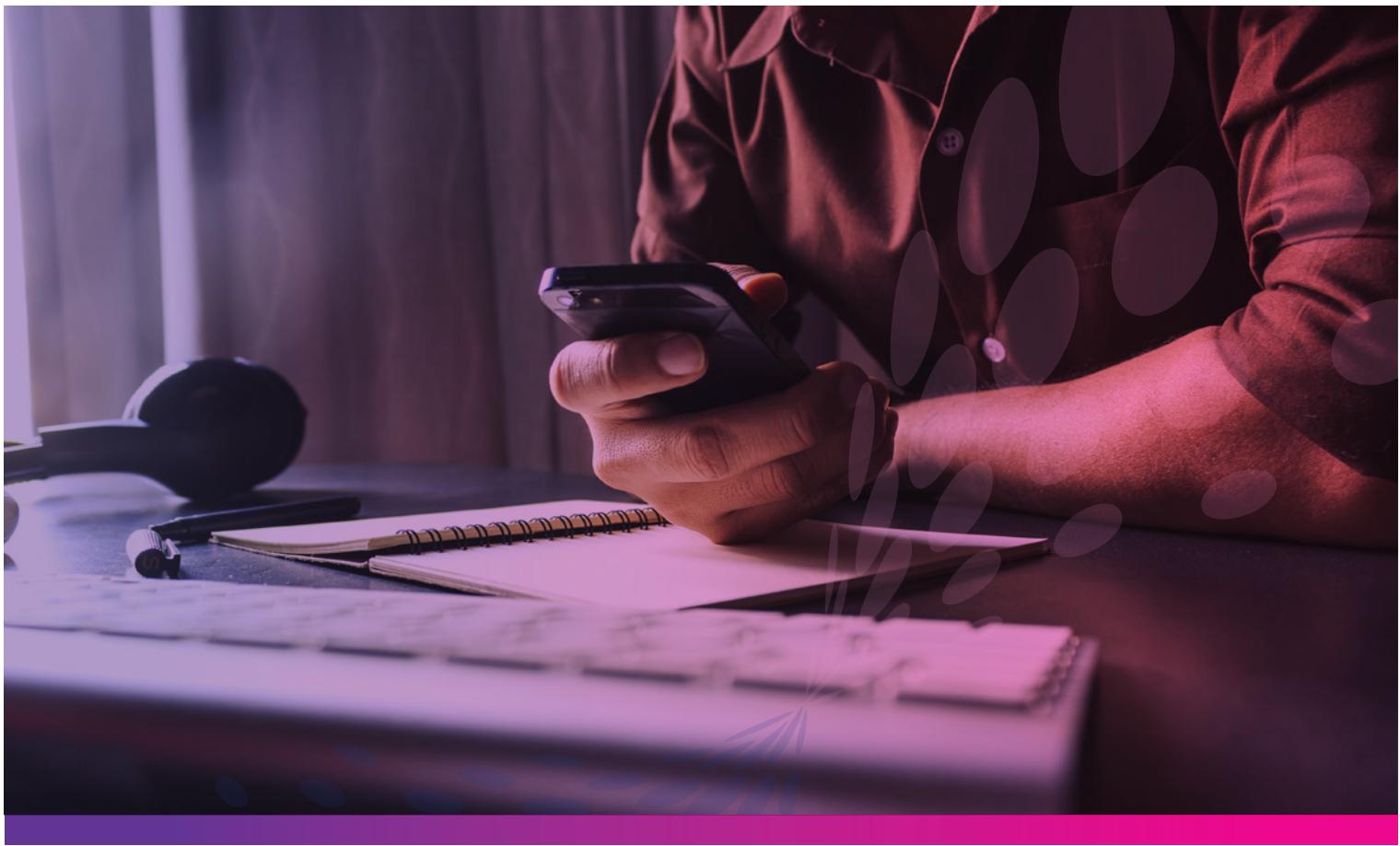
$$PIF < 10^{-\log_{10}(\text{Minimum Identifiable Cohort Size}) - \varepsilon}$$

The minimum identifiable cohort size (MICS) is the smallest group within a dataset that can be created from the available features.

For example, in one dataset there may be 100 males without beards born in NSW. If an additional feature is included (those under 18), this number may reduce to 10. In this example, the MICS is at most 10. The 'at most' is important, as it specifies there cannot be a cohort smaller than this. The strict condition of the MICS being determined within a closed, linked, de-identified dataset is required to satisfy the condition that no additional data can be introduced to this set.

FOR A MICS OF:

1	2	5	10	100
the PIF is less than 1.0	the PIF is less than 0.5	the PIF is less than 0.2	the PIF is less than 0.1	the PIF is less than 0.01



As new datasets are added to an existing closed linked dataset, new features are potentially identified. As a consequence, the MICS will potentially reduce, leading to higher PIF values.

The notion of bound is important, as having a cohort size of 1 in a deidentified dataset is not the same as having personal information (when the MICS is 1, the PIF is still strictly less than 1). Some additional data or feature is needed to identify the actual individual.³²

The term ϵ or 'epsilon' in the PIF calculation is intended to reflect the fact that with de-identified data, even at MICS of 1, at least one additional data field is required to map to the identifiable individual.

In the example above of a defined de-identified cohort, knowing there is only one male member does not provide sufficient information to identify the male as a named individual. Depending on the exact circumstances, it is possible to imagine additional data (an additional feature) which would allow identification. Similarly, if there were two males in the cohort, it is possible to imagine several additional datasets (features) that would allow individual identification.

The approach continues for five or ten males in a defined cohort. The PIF is therefore treated as upper-bound rather than an exact value. The additional information required to link the individual described by their feature set in the data may include a unique personal feature, a unique name, a unique address or a unique relationship.

Figure 11 shows a simple example of a closed, linked, de-identified dataset with a population of size 16 ($P=16$), with eight features ($F=8$) and four equal-sized cohorts ($MICS=4$). The PIF for each of these cohorts is strictly less than 0.25. In this simplistic example, the first four features (f_1, f_2, f_3, f_4) define the cohorts, and the addition of features 5 through 8 do not impact the cohort sizes.

³² The members of this data set may be reasonably identifiable in this circumstance, just not actually identified.

	f1	f2	f3	f4	f5	f6	f7	f8
p1	1	0	0	0	0	0	0	0
p2	1	0	0	0	0	0	0	0
p3	1	0	0	0	0	0	0	0
p4	1	0	0	0	0	0	0	0
p5	0	1	0	0	0	0	0	0
p6	0	1	0	0	0	0	0	0
p7	0	1	0	0	0	0	0	0
p8	0	1	0	0	0	0	0	0
p9	0	0	1	0	0	0	0	0
p10	0	0	1	0	0	0	0	0
p11	0	0	1	0	0	0	0	0
p12	0	0	1	0	0	0	0	0
p13	0	0	0	1	0	0	0	0
p14	0	0	0	1	0	0	0	0
p15	0	0	0	1	0	0	0	0
p16	0	0	0	1	0	0	0	0

Figure 11. Population of 16, with eight features and four equal-sized cohorts

The quantification of epsilon is still to be finally determined and will be contextual. It relates to the uniqueness of the minimum identifiable cohort and is currently defined for the purpose of this paper as:

$$\epsilon = \sum_{i=1}^F \frac{1}{d^2(i) * Gp(i)}$$

Where:

- $d(i)$ is the Hamming Distance³³ (the count of features that do not match) between the minimum identifiable cohort and all cohorts of size Gp at distance i .
- $Gp(i)$ is non-zero.
- F is the number of features (for example, hair colour) in the closed, linked, de-identified dataset.

As illustrated in the example population shown in Figure 11 and conceptualised in Figure 12, there may be more than one cohort at any given distance from the minimum identifiable cohort and there may be more than one cohort with the MICS.

³³ For an explanation of Hamming Distance, see <http://www.oxfordmathcenter.com/drupal7/node/525>. While many features will have non-binary values – hair colour may be a range of values, age may be recorded in number of years – each feature can be mapped to one of a finite number of values as a categorical variable without loss of information. The use of Hamming Distance as a measure of similarity relies on counting the number of features which differ, not considering how much they differ.

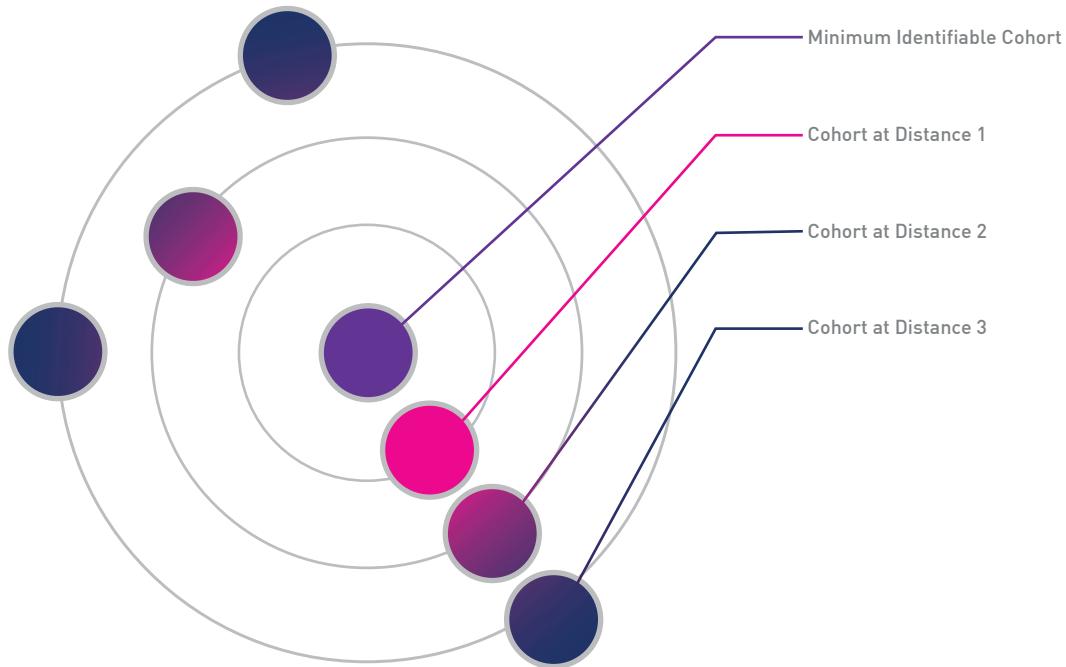


Figure 12. Illustration of the relationship between the minimum identifiable cohort and other cohorts

The more unique a cohort is, the smaller the epsilon. In the population example of Figure 11, the Hamming Distance (the count of the number of features which are different) between each cohort is 2, and the epsilon value for each cohort is approximately 0.02. The larger the value of epsilon, the more similar the cohort is to other cohorts, and so the larger the number of additional features required to identify a unique member of the population. In the special case that the MICS is the entire population (P), then epsilon is 0 and the PIF is bounded by $1/P$.

AS AN EXAMPLE, FOR AN EPSILON OF 0.01, AND A MICS OF:

1	2	5	10	100
the PIF is less than 0.98	the PIF is less than 0.49	the PIF is less than 0.20	the PIF is less than 0.10	the PIF is less than 0.01

As a point of note, every new feature introduced to a population can increase the Hamming Distance between the minimum identifiable cohort and other cohorts by at most distance 1. As a consequence, even if addition of a new feature splits a cohort to create a MICS of 1, the Hamming Distance to the remainder of the previous minimum identifiable cohort is at most distance 1.



BREAKING CONTEXT: SPATIAL, TEMPORAL AND RELATIONSHIP INFORMATION FACTORS

In a closed, linked, de-identified dataset, it is assumed that each feature is independent. That is, no information can be gained about one feature by examining another. Knowledge of context can, however, allow information to be inferred, decreasing the feature depth. Separating features that provide context from features that describe a person or object helps ensure the independence of features.

In an exact analogy to the Personal Information Factor, this paper introduces a Contextual Information Factor (CIF) which is combination of a Spatial Information Factor (SIF), Temporal Information Factor (TIF) and Relationship Information Factor (RIF).

Data is a record of events that have occurred in the past, and all data relates to events which happen somewhere. By separating out features from the dataset that describe location (for example, street, suburb, latitude) and time (for example, hour, day, date), we are able to isolate features that describe a person or object without spatial or temporal context.

In a closed, linked, de-identified dataset, relationships connect members of that population and create additional links beyond correlations between features. By separating out features from the dataset that describe relationship (for example, kinship, ownership, partnership), we are able to isolate features that describe a person or object without relationship context.

In an exact analogy to the PIF, we define each of:

$$PIF_nocontext < 10^{-\log_{10}(\text{MICS Personal}) - \xi_p}$$

$$SIF < 10^{-\log_{10}(\text{MICS Spatial}) - \xi_s}$$

$$TIF < 10^{-\log_{10}(\text{MICS Temporal}) - \xi_t}$$

$$RIF < 10^{-\log_{10}(\text{MICS Relationship}) - \xi_r}$$

where the MICS in each context feature set is the smallest cohort formed exclusively from personal, spatial, temporal or relationship features. The value of epsilon in each case is an exact analogy to the value calculated for PIF. It describes the uniqueness of a cohort in a population using only those features that describe spatial/temporal/relationship attributes.

When calculating the overall Contextual Information Factor (CIF) for a population described by a set of spatial/temporal/relationship attributes:

$$CIF < 10^{-\log_{10}(\text{MICS Spatial} \cap \text{Temporal} \cap \text{Relationship}) - \xi_c}$$

where \cap is the intersection operator acting on cohorts defined by spatial/temporal/relationship features and ξ_c is defined as before, but across all cohorts formed by spatial/temporal/relationship features.

Figure 13 illustrates how cohorts based on different contextual feature sets intersect to create a minimum identifiable cohort. For clarity, the MICS spatial/temporal/relationship cannot be larger than the cohorts formed by any one context feature set and the MICS must be at least 1. If the MICS for the context feature set is the whole population, then the CIF is 1.

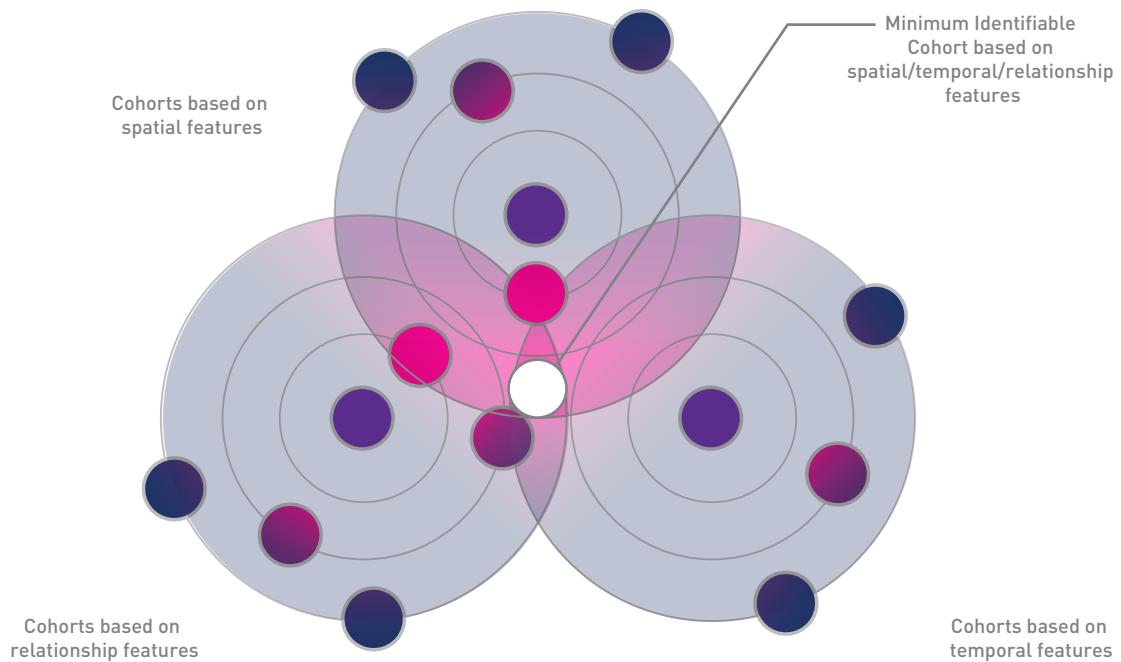


Figure 13. Illustration of cohorts based on spatial, temporal and relationship features

The PIF is then updated to be defined by:

$$PIF < 10^{-\log_{10}(\text{MICS Personal}) - \xi_p} \times 10^{-\log_{10}(\text{MICS Spatial} \cap \text{Temporal} \cap \text{Relationship}) - \xi_c}$$

As the value of **PIF_nocontext** and **CIF** are strictly less than 1, the product will be smaller. For a **PIF_nocontext** less than 0.5, and a CIF of:



The relationship between **PIF_nocontext** and **CIF** means that a PIF can potentially be reduced whilst maintaining MICS Personal. This is useful if population sizes are relatively small or if the number of personal features is relatively large.

EXPLORING PERSONAL INFORMATION FACTOR(S), K-ANONYMITY AND THE IMPORTANCE OF CONTEXT

A common approach to protecting personal information in a dataset is to reduce the risk of re-identification by use of k-anonymity (or l-diversity)³⁴. These techniques represent ways of minimising risk of re-identification, rather than measures of personal information in the dataset.

A dataset is said to have the k-anonymity property if the information for each individual contained in the release cannot be distinguished from at least $k-1$ individuals whose information also appears in the dataset. There are two commonly employed approaches for achieving k-anonymity (for a given value of ' k '):

- **Generalisation** – where values of selected attributes are replaced by a broader category. For example, age may be replaced by bands of 0–5 years, 5–10 years and so on.
- **Suppression** – where certain values of the attributes are replaced by a null value before release. This is often used for values such as a person's religion.

Because k-anonymisation does not include any randomisation, someone attempting to re-identify an individual can still make inferences by linking other datasets to the k-anonymised set. It has also been shown that using k-anonymity can skew the statistical characteristics of a dataset if it disproportionately suppresses and generalises data points with unrepresentative values.

If a person is known to be in a dataset and can be identified to be in the minimum identifiable cohort using a subset of features (for example on eye colour, gender, age), then any additional features not used to identify them can be learned. For this reason, k-anonymity is not considered a good technique for protection of privacy of individuals in relation to high-dimension datasets.

³⁴ See, for example, L. Sweeney, 'k-anonymity: a model for protecting privacy', International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002, pp. 557-570. Available online at https://epic.org/privacy/reidentification/Sweeney_Article.pdf.

The approach of separating selected context features in a dataset has been explored in mobile communications systems³⁵ to protect location information of mobile users. These approaches are referred to as 'spatial cloaking' and are employed in circumstances where aggregation techniques such as k-anonymity are used to reduce privacy threats resulting from uncontrolled usage of location-based services. Extending the contextual separation process to include spatial, temporal and relationship features potentially further increases the effectiveness of spatial cloaking-style protection.

The PIF described is not a technique for anonymisation. Rather, it is a heuristic measure of the potential risk of re-identification of an individual based in a given dataset based on the smallest identifiable cohort. At its simplest, the PIF described reduces to $1/k$ if k is the MICS and there are no other cohorts identified in the population. If there are other cohorts, the PIF is less than $1/k$.

To illustrate, Figure 14 shows the evolution of the PIF data in Figure 11 as the features in the first row change value (individual feature values change from 0 to 1). The dataset initially has four cohorts of size 4 and all cohorts are equidistant from each other. From a value of approximately 0.23, the PIF rises quickly as a minimum cohort of 1 is created with the first change of feature value (f2 changes from 0 to 1).

This smallest cohort now has distance 1 from a cohort of size 3 and a cohort of size 4 and has distance 2 from two cohorts of size 4. The distance to, and size of, these other cohorts means the PIF does not reach 1. As the number of features that change value in the first row increases, the PIF moves closer to 1. The cohort of size 1 becomes more unique as the distance increases from all other members of the population.

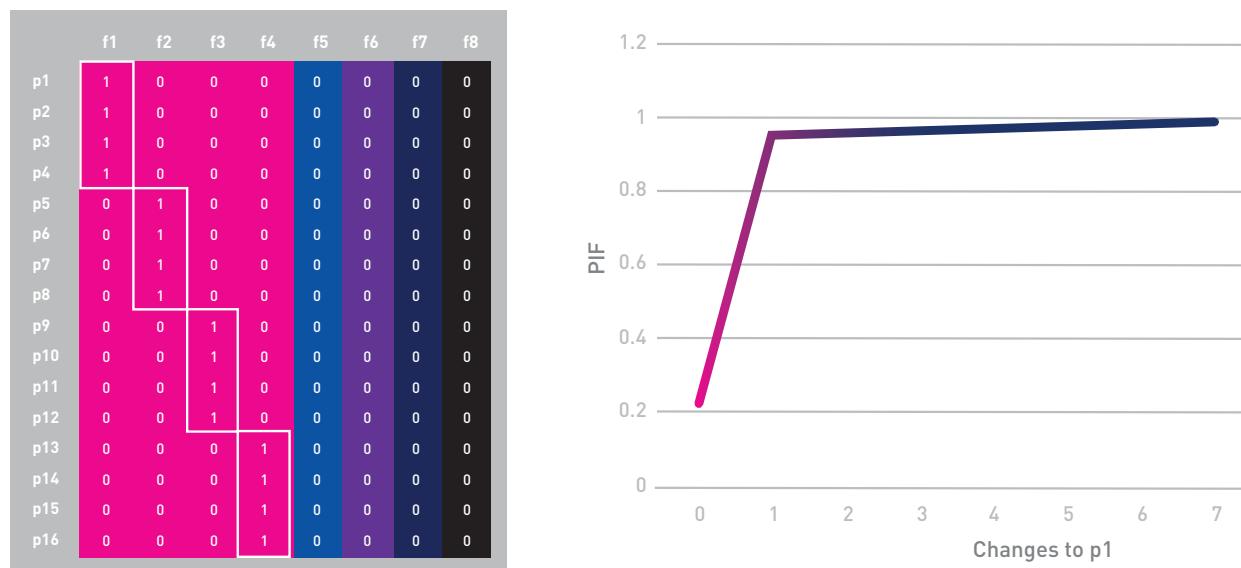
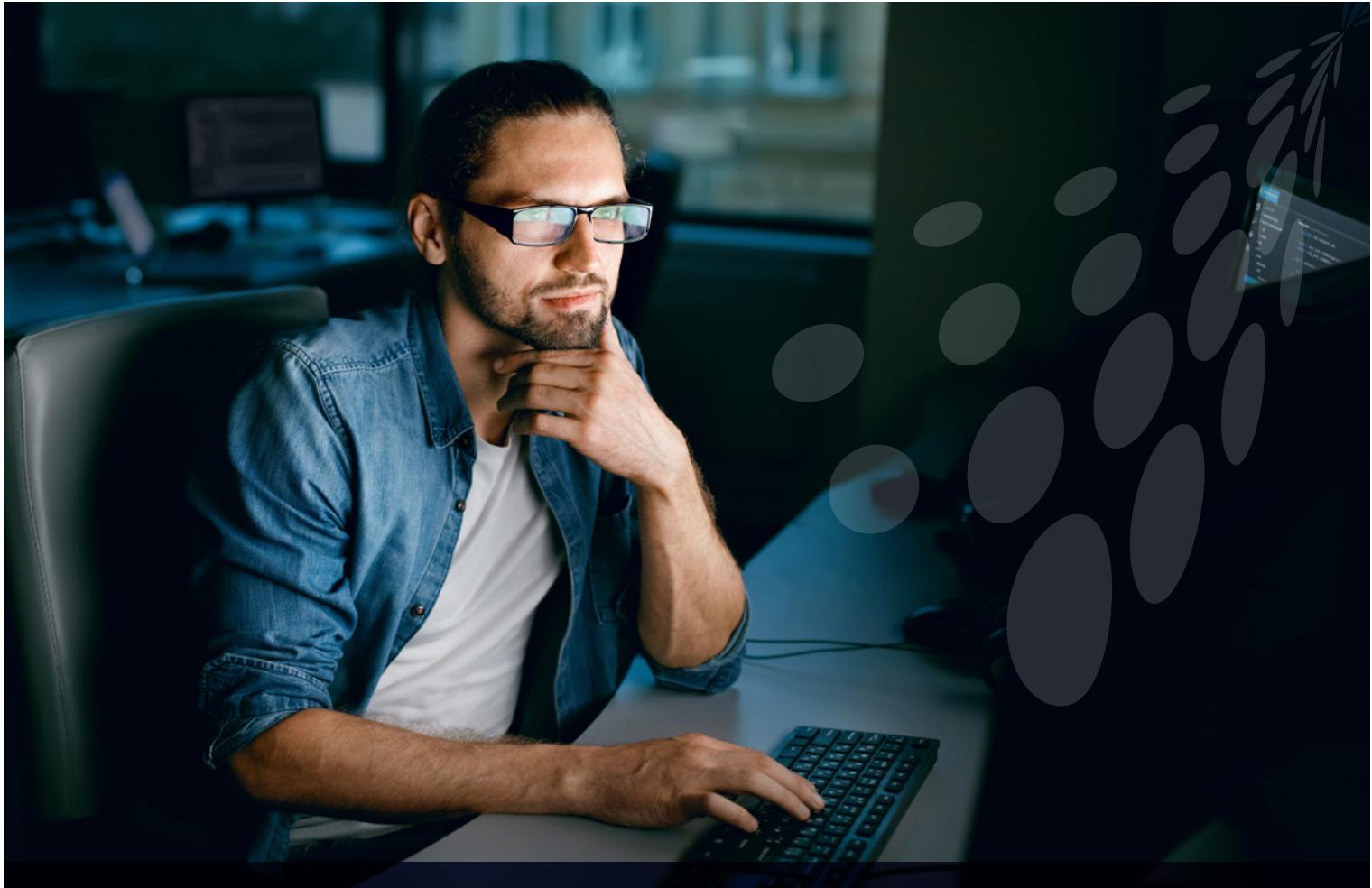


Figure 14. Population example with MICS of 1 and cohorts of size 3 and size 4

³⁵ See B. Gedik, L. Liu, 'Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms', IEEE Transactions on Mobile Computing 7(1), January 2008. Available online at <https://ieeexplore.ieee.org/abstract/document/4359010/>



A DATA SAFETY FACTOR (DSF)

Referring again to Figure 9, the Safeness of data included a PIF built on a range of features, including feature depth, coverage probability and accuracy. This paper proposes a heuristic for a DSF:

$$\text{Data Safety Factor} \geq \frac{1}{\text{PIF}} * \frac{1}{\text{Feature Depth}} * 10^{(-\text{Coverage Probability} * 2)} * \frac{1}{1 + 10^{(0.7 - \text{Accuracy})}} * 10$$

Feature Depth is weighted as an inverse multiple, reflecting the significance of the additional information that would be revealed about an individual for each additional feature included in the dataset.

Coverage Probability is weighted as an inverse exponential (squared), reflecting that small reductions in total population inclusion lead to increased uncertainty that a known individual will be present in a sample dataset.

Accuracy is weighted as a sigmoid function, reflecting that small reductions in accuracy produce significantly less safe data and outputs. The sigmoid function has a value of 0.5 at 70% accuracy, reflecting the significant reduction in data safety as data accuracy reduces.

These factors interact to ensure that as the value of PIF increases towards 1 or the accuracy decreases, the DSF reduces rapidly towards 1.

As the number of independent features increases, the DSF decreases. As coverage probability approaches 100%, the DSF reduces rapidly. Figure 15 shows the scaling factors associated with coverage and accuracy.

If accuracy and coverage are unknown, they are assumed to have no effect. The Data Safety Factor simplifies to the combination of the inverse of the PIF and the inverse of the Feature Depth. If the Feature Depth is not known, the DSF reduces to simply the inverse of the PIF.

Figure 16 shows how the DSF changes with coverage probability (from 10% to 100%) for a linked, de-identified dataset with ten independent features and an accuracy of 100% for differing PIF values. It is again emphasised that this Data Safety Factor is a heuristic measure. Figure 17 shows the change in values when accuracy falls to 80%.

The level of data safety to be made available will depend on the other Safe settings. This will be discussed further in a later section.

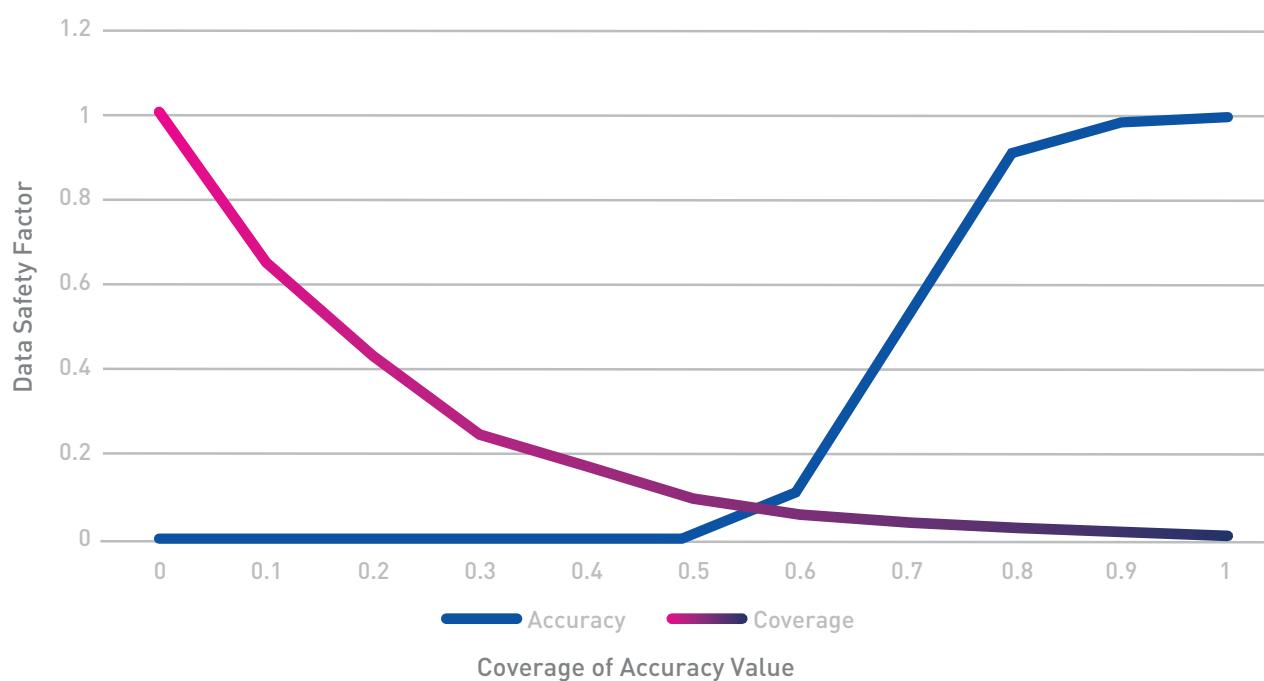


Figure 15. Scaling associated with coverage and accuracy parameters

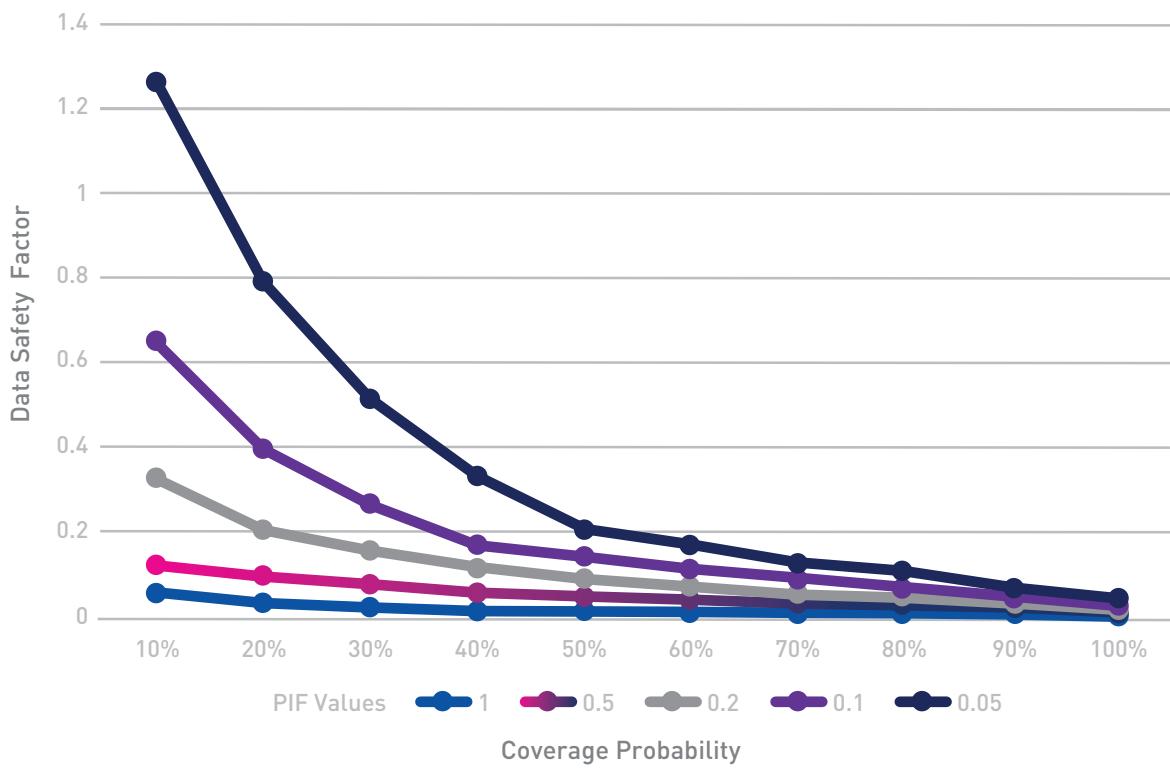


Figure 16. Data Safety Factor versus Coverage Probability for a Feature Depth of 10 and Accuracy of 100%

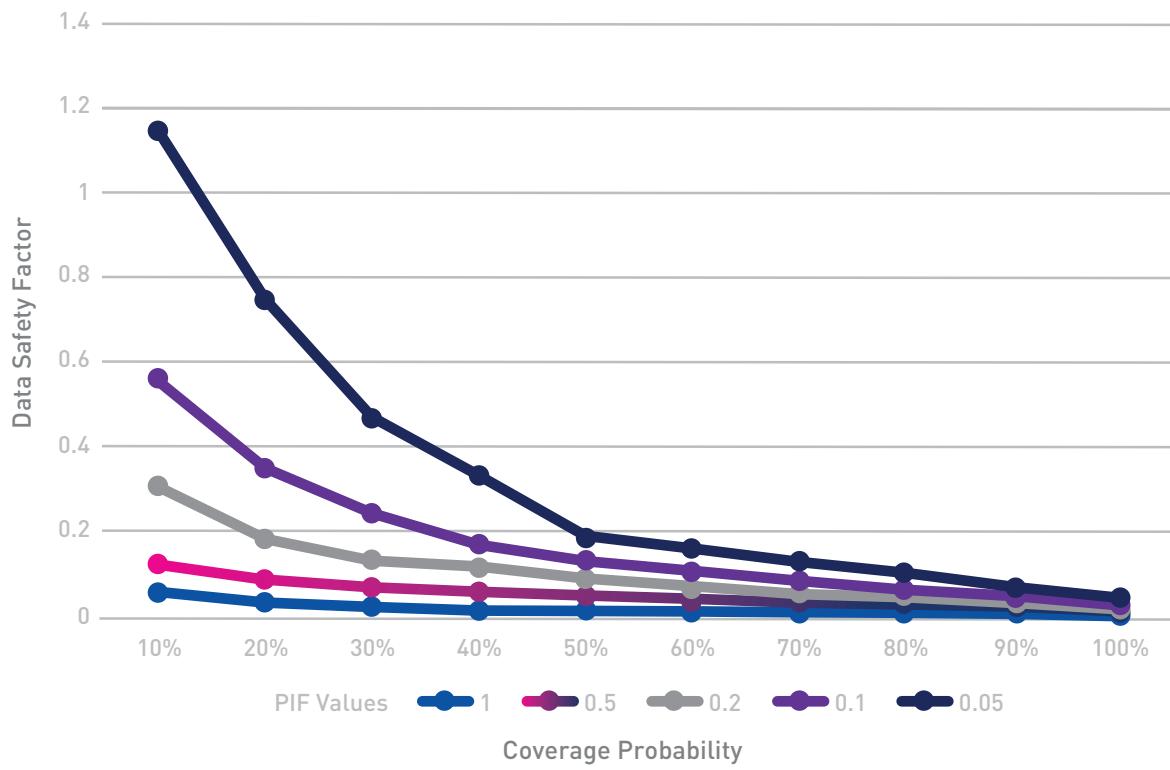
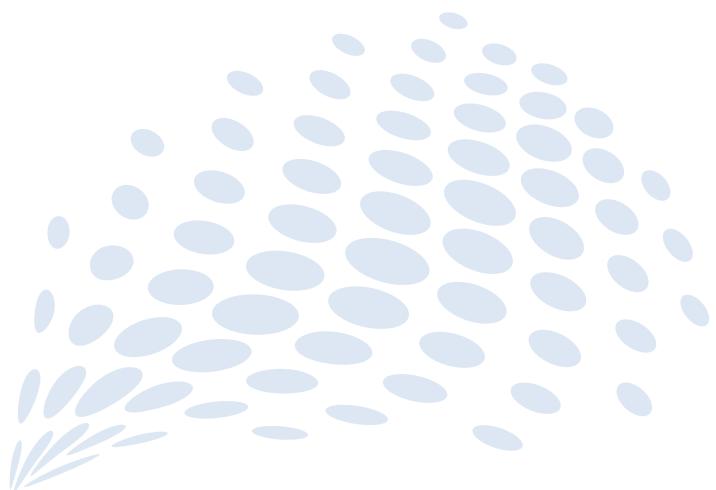


Figure 17. Data Safety Factor versus Coverage Probability for a Feature Depth of 10 and Accuracy of 80%

04



Addressing the ‘reasonable’ challenge

The reliance under privacy laws on a test for ‘reasonable likelihood’ when determining if personal information is present in a dataset has a fundamental limitation centred on the human ability to decide if data contains personal information.

While human ability might suffice for a modest dataset, if many tens or even hundreds of datasets are to be linked for analysis, the ability to make this assessment becomes extremely difficult. In these circumstances, it would be appropriate for an organisation to adopt a more formal (and repeatable) process of risk assessment before a decision is taken.

SAFE OUTPUT – WHEN IS PERSONAL INFORMATION REVEALED?

It is important to be clear when personal information has the potential to be revealed in a project. In this paper, a distinction is made between the level of personal information (see Figure 18):

- When linked and analysed in an analytical environment (Insights and Models level);
- When considering outputs at different stages in a project which are seen by an observer (personal context level); and
- When outputs are made available to the wider world and may be linked to datasets in the wider world (real world context level).

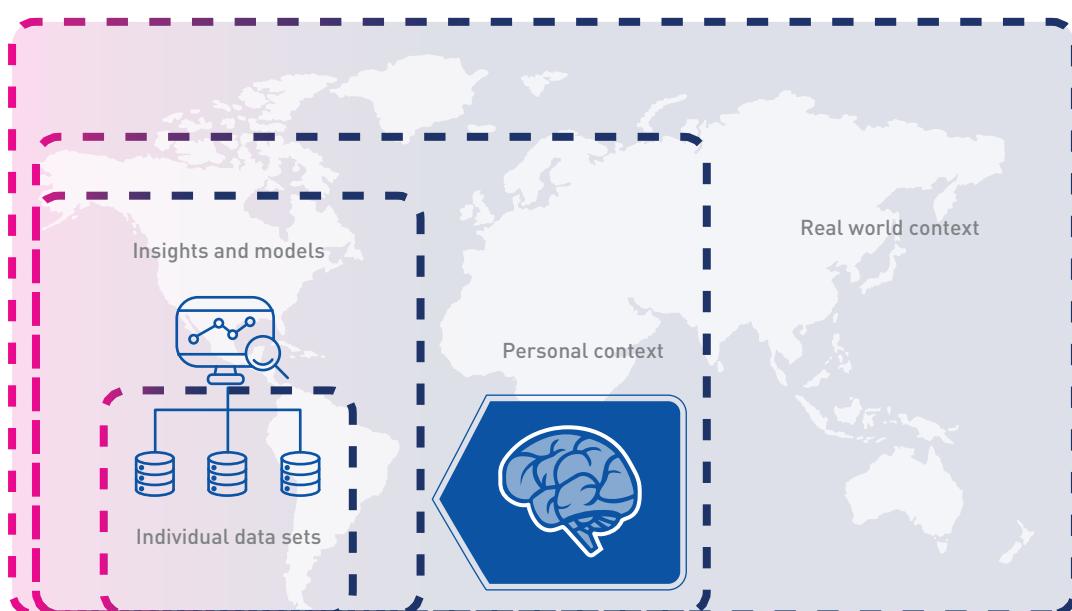


Figure 18. Context for determining the degree of personal information

In the lowest level in Figure 18 (Insights and models level), it is possible to link de-identified datasets and ensure the PIF does not reach 1 by mathematically exploring the feature sets that describe the MICS. If the smallest identifiable cohort is $N > 1$, then the PIF is strictly less than 1. This means more independent data (features) are needed to reach a PIF of 1.

When working with de-identified data, a minimum identified cohort of 1 does not explicitly imply a PIF of 1 (i.e. personal identification). As discussed above, a de-identified dataset with a MICS of 1 still requires additional data to map to an individual. In the closed analytical environment (Insights and models level), this additional data is not available.

In the next level of this model (Personal context), any observer who views results will bring their own experience, knowledge and perspective to that observation. At this point the 'reasonable' test is truly applied. It is impossible to know the total range of interactions between the PIF developed by the linking of analytical processes and the additional information brought by personal context of the observer. The risk mitigation required when revealing outputs at different stages of the project depends on the level of safety of the observer in context of the other Safe dimensions of the project (Safe Setting, Safe Data and Safe Output). This is discussed further in the next section.

In the final level of the model (Real world context), any observer who views the results not only brings their own knowledge and experience, but also has access to a wide range of other datasets to potentially link to the project outputs. The level of protection via MICS becomes increasingly important.

SAFE PEOPLE AND SAFE PROJECTS – WHO CAN ACCESS DATA AND WITH WHOM CAN OUTPUTS BE SAFELY SHARED?

In this paper, a distinction is made between concerns about the sensitivity of project findings and privacy. A project may produce results that are challenging; however, unless there is an issue of privacy, these concerns are not considered here. This paper acknowledges that outputs are produced at multiple stages in a project rather than just at completion. This section therefore deals with Safe People and Safe Projects at different stages of a project lifecycle.

The level of Safeness of people relates to the level of pre-qualification for inclusion in the project – from deep involvement to no vetting at all. The level of Safeness of a project relates to the level of PIF involved in the project – from 'very safe', with a PIF of 0, to 'not safe', with a PIF at a level close to 1.

The term 'not safe' is used simply to reflect a scale which has 'very safe' at one end.

As results from different stages of an analytics project are produced, they potentially increase in PIF as a consequence of linking and analysing datasets, and so greater risk is associated with sharing.

Figure 19 shows an example of how Safe settings may be established for combinations of different levels of safety for people and projects. In this example, people considered to be Not Safe (or not evaluated) only gain access to data that is publicly available. If open data is the only data used, it is impossible to overlay governance on a project. Projects that are evaluated as Not Safe (PIF of equal to or greater than 1) are excluded from this example as they require individual evaluation.

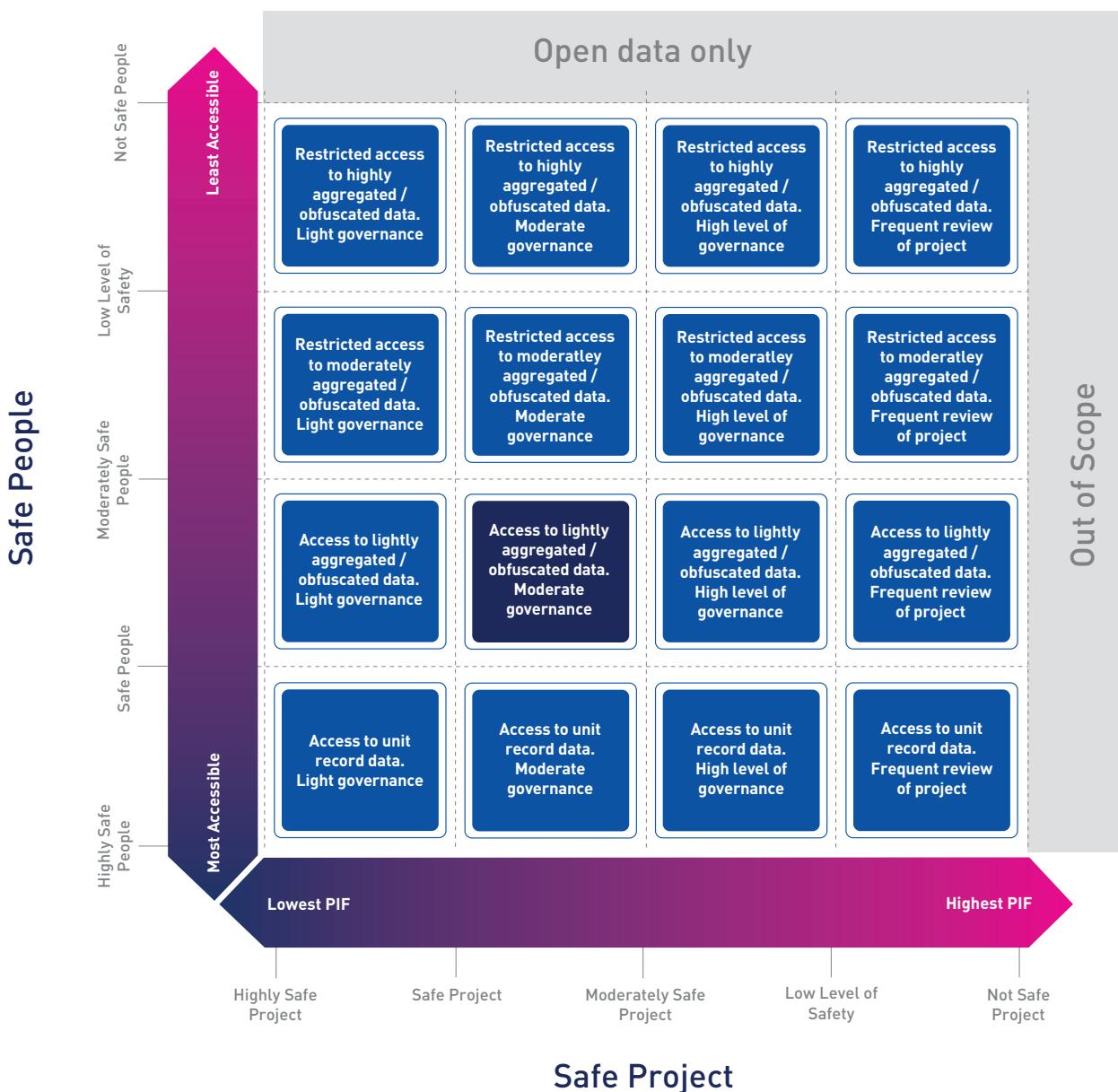


Figure 19. Safe settings for a combination of projects and people

While technology cannot be considered to be the complete answer to Safe Setting, it can help mitigate risks for different levels of safe. Examples of systems that provide Safe Setting at different levels already exist. The challenge with many of these current frameworks is that they are not particularly well suited to widespread, automated data sharing.

As an example, the Secure Unified Research Environment (SURE)³⁶ is a long-established framework that enables researchers to access sensitive data. Authorised researchers working on approved projects operate on data within a constrained environment. Researchers perform operations over unit-record level data, and are prevented from adding identified data and on-sharing data. While addressing the needs of individual researchers, the system is not well suited to wide ranging collaboration in its current form.

³⁶ For more information see <https://www.saxinstitute.org.au/our-work/sure/>



At the other extreme, systems such as data.gov.au³⁷ provide examples of data sharing mechanisms for open data. While appropriate for the release of raw data, particularly from government agencies, it remains limited from the perspective of wide-ranging collaboration.

Technology under active development allows computational operations to be performed that return the answer to a query without providing access to the underlying stored data. The de-identified computations can be distributed, performing calculations over multiple data sources at multiple sites and returning just the computed outcome.

These developments are well advanced, and while there will be a significant additional information and communications technology (ICT) burden associated with this approach, it may significantly lower privacy and legal concerns associated with use of data and so reduce governance requirements.

DEALING WITH MIXED SAFE LEVELS

One of the fundamental principles underpinning the challenge of data sharing is addressing the challenge of value, risk and trust in data sharing. This can change as a data analysis project (the simplest case being data sharing) develops through the major phases of:

- Project scoping (including identification of people)
- Data collection, organisation and curation
- Data analysis
- Results interpretation

³⁷ See <https://data.gov.au/>

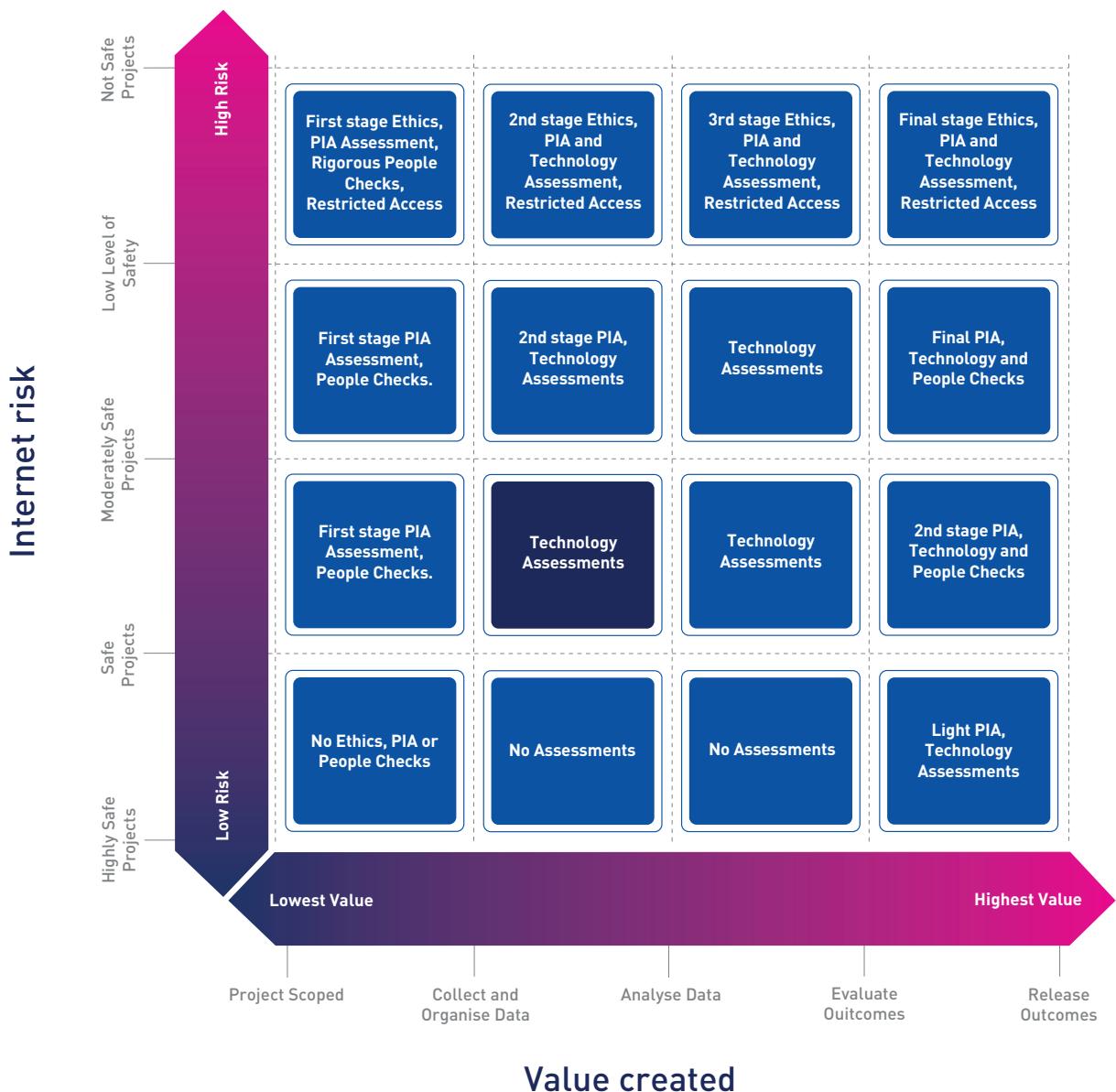


Figure 20. Ethics, privacy impact, technology and people assessments for different risk levels

As each of these phases progresses, the value of the outputs increases and potential risks to privacy may also increase. An important consideration is that projects involving any element of discovery need periodic review depending on the level of risk assessed at each of the major project phases. Identification of the impact on privacy or the ethical considerations of a project will depend on what is identified, and this may not be known at the outset.

A more flexible approach to data analysis projects may allow light-touch up-front assessment of privacy impact, people and technology, and increase the frequency or intensity of these assessments as the project continues.

A summary of possible guidelines is given in Figure 20. Figure 21 attempts to map the major data analysis project phases to the risk mitigation focus for each dimension in the Five Safes Framework. The benefit of a multi-stage assessment for privacy and ethics is that it is no longer necessary to preconceive at the outset of the project all of the issues or risks which may arise during analysis. In these figures, PIA refers to a Privacy Impact Assessment.

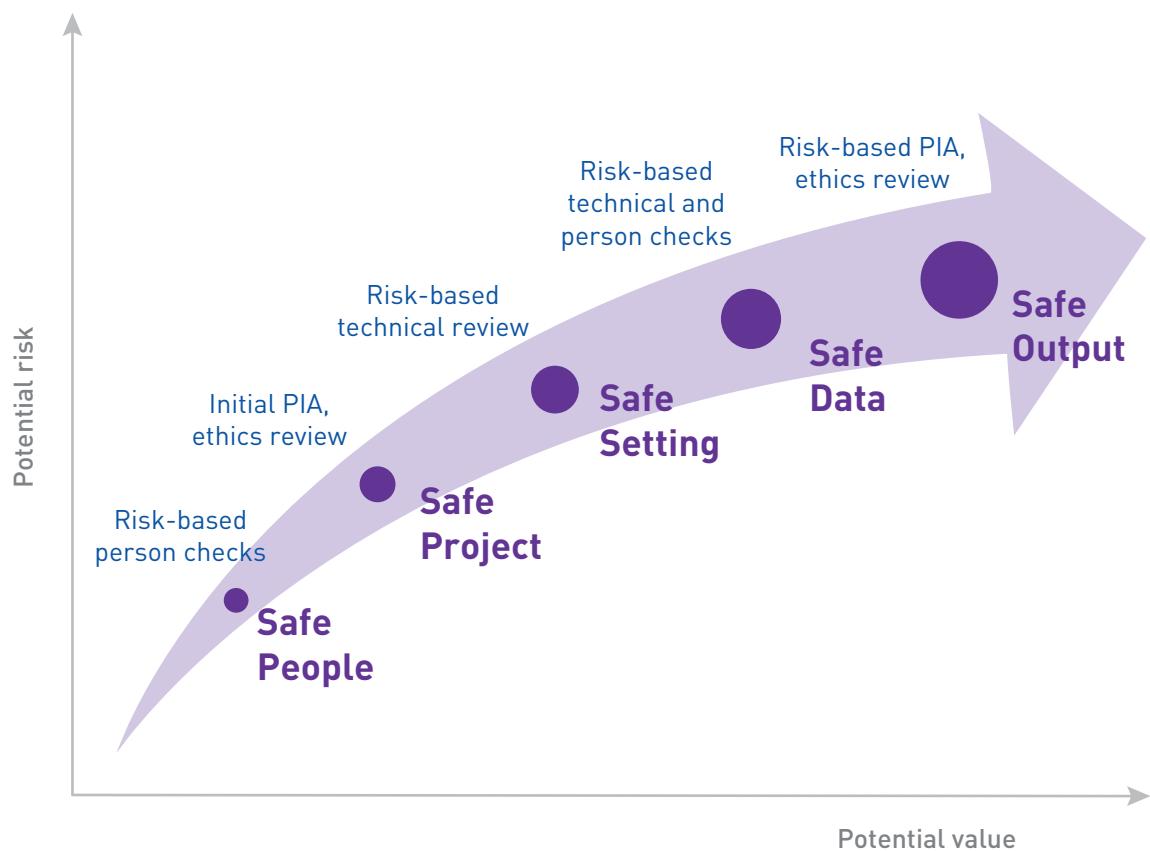


Figure 21. Mapping to the Five Safes Framework



05



Quantifying the Five Safes framework

This section will seek to provide more quantified measures of the modified Five Safes Framework without addressing the added dimensions of the Safe Organisation, Safe Outcomes or Safe Lifecycle.

SAFE PEOPLE AND SAFE PROJECTS

Evaluating Safe People requires an evaluation of intention and judgement of the character of individual participants. This may be assisted by identification of conflicts of interest, reference checks or specialised checks such as police checks, working with children checks or national security checks. The outcome of such an evaluation will establish the level of access that can be provided to an individual participant, including the sensitivity of the data and which sort of projects they may be involved in.

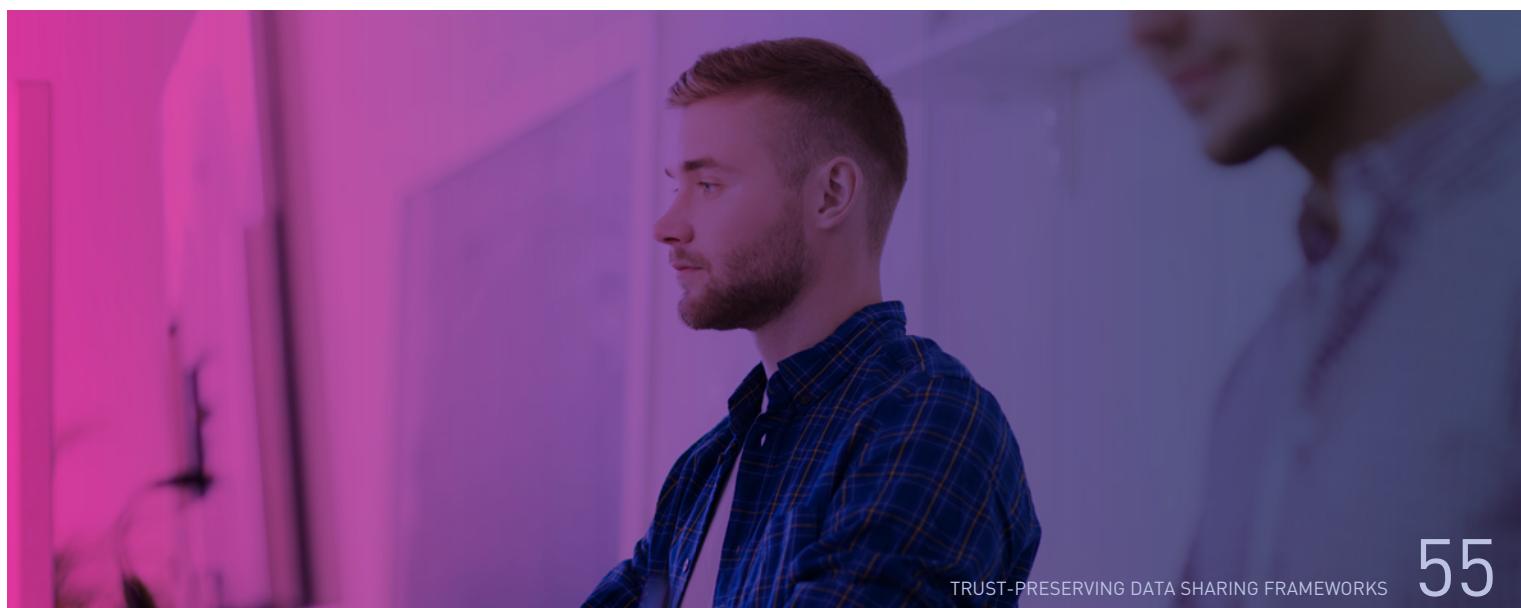
None of these checks provide a definitive indication of the intention of the person involved in the project or the likelihood they will breach an aspect of the Safes framework. Rather, identification of possible motive and evaluation of past performance are used as predictors of future actions.

Evaluating Safe Projects requires judgement of the purpose of the project from a risk and ethical perspective. Formally convened ethics committees exist in most countries to evaluate research projects and to provide guidelines for conduct when carrying out projects. As an example, the UK's Social and Economic Research Council (SERC) provides a framework for research ethics principles, procedures and minimum requirements.³⁸

Within the scope of a project, the major factors to consider are:

- The (potentially) increasing PIF at each stage of the project.
- The people who can access the outputs at each stage of the project and at what level of PIF.

38 Available online at http://www.gla.ac.uk/media/media_326706_en.pdf



Taking an example project that involves linkage and analysis of data on children and families, the characteristics of different levels of Safe for people may include the following:

Safe Level 5 – Highly Safe People:

Example: researcher or research supervisor

- Security check such as Police Check or Working with Children Check.³⁹
- Formally verified data analytics skills.
- Higher technical degree or working under supervision of higher technical degree.
- Named access on relevant data sharing agreements (such as an MoU).
- Access to de-identified, linked, unit-record data.
- Access to results at de-identified, linked, unit-record level.
- Entered into a specific confidentiality undertaking.

Safe Level 3 – Moderately Safe People:

Example: agency partner

- Working with Children Check.
- Named access on relevant data sharing agreement.
- Knowledge of data at dictionary level (i.e. which features are used).
- No access to de-identified, linked, unit-record data.
- Access to aggregated results at cohort level (further increased size of MICS).
- Made aware of confidentiality of the data and requirements to protect privacy.

Safe Level 4 – Safe People:

Example: partner agency project reviewer

- Police Check and Working with Children Check.
- Verified data analytics skills.
- Named access on relevant data sharing agreement.
- Knowledge of data at dictionary level (i.e. which features are used).
- No access to de-identified, linked, unit-record data.
- Access to aggregated results at cohort level (increased size of MICS).
- Some legal obligations to maintain confidentiality at a general level, such as through employment relationships or professional duties.

Safe Level 2 – Low-level Safe People:

Example: unrelated agency

- Not named access on relevant data sharing agreement.
- No access to de-identified, linked, unit-record data.
- Access to more highly aggregated results (further increased size of MICS).

Safe Level 1 – Not Safe People:

Example: general audience

- No security checks.
- Not named on relevant data sharing agreement.
- No access to de-identified, linked, unit-record data.
- Access to aggregated results at trend level (largest MICS).

³⁹ See Office of the Children’s Guardian, <https://www.kidsguardian.nsw.gov.au/child-safe-organisations/working-with-children-check>

When applied to the Five Safes Framework, example threshold tests for Safe Projects may include:

Safe Level 5 – Highly Safe Project:

- Having no identified ethical aspects or not using data involving people.

Safe Level 4 – Safe Project:

- Having minor ethical risks that can be mitigated, or using highly aggregated or obfuscated data that has little residual personal information (large MICS and low PIF).

Safe Level 3 - Moderately Safe Project:

- Having ethical risks that require monitoring, or using lightly aggregated or obfuscated data with a possible risk of re-identification of individual information (smaller MICS).

Safe Level 2 - Low-level Safe Project:

- Having identifiable ethical risks that require significant attention, or using lightly aggregated or obfuscated data with a plausible risk of re-identification of individual information (smaller MICS).

Safe Level 1 – Not Safe Project:

- Having clear ethical risks or using personal information.



Following the flow of logic in Figure 19, Figure 22 shows the relevant squares highlighted for different levels of safe for participants or observers of the outputs:

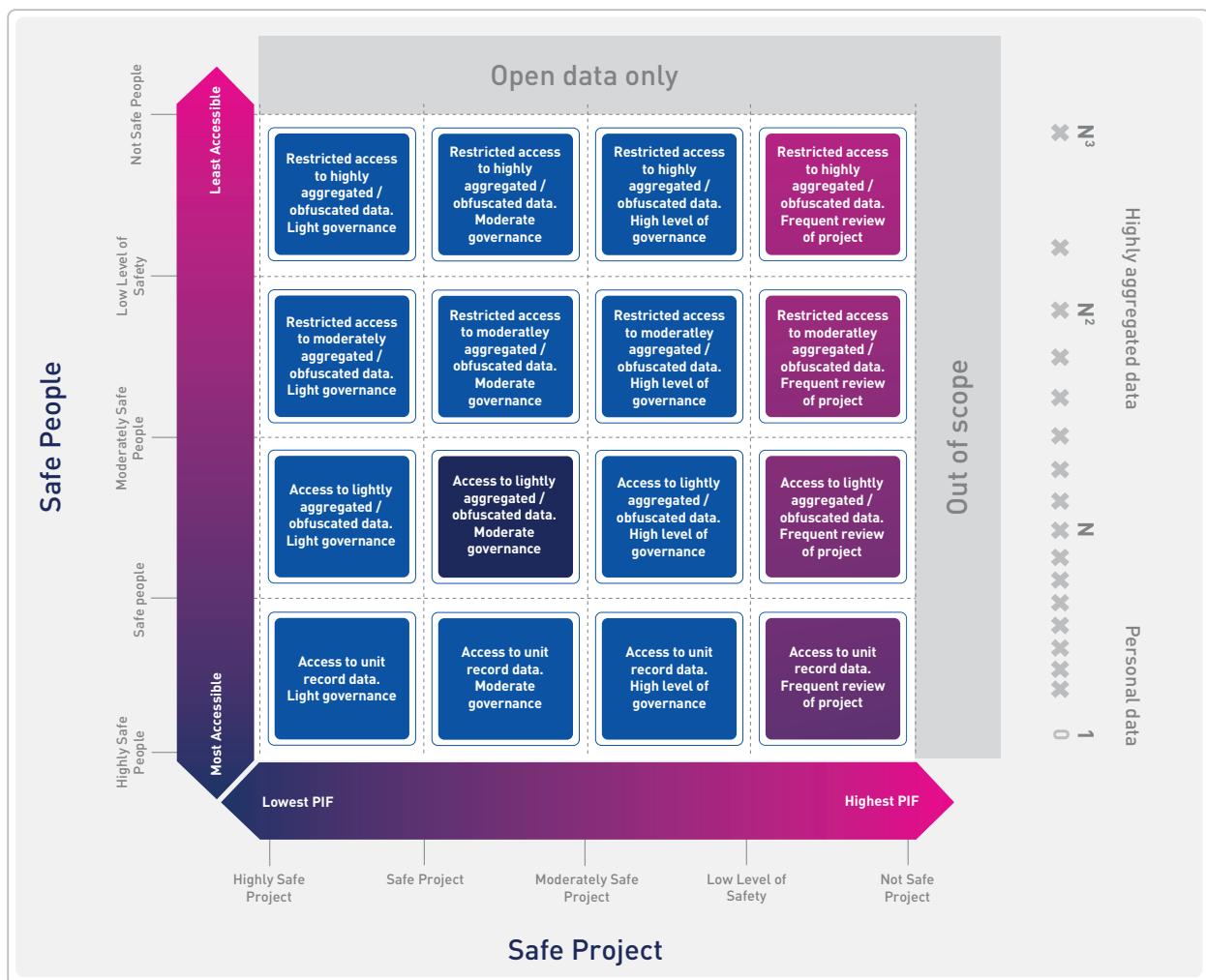


Figure 22. Access and supervision for Safe People and Safe Projects

SAFE DATA AND SAFE OUTPUTS – THRESHOLDS BASED ON DATA SAFETY FACTOR

The Safe Data dimension of the Five Safes Framework is illustrated in Figure 23.

The similarities between a PIF based on MICS and k-anonymity as a means of preserving privacy have been discussed. Whilst the PIF calculation depends on more than just the smallest cohort, the weakness of the approach becomes most significant when the minimum cohort size is small or when additional information is known about the population. For example, if an individual is known to be in a sample dataset, then one or more of the features of that dataset are likely to be associated with that individual. The known presence of the individual provides the basis for strong inference of personal attributes.

The concept of the Data Safety Factor (DSF) takes into consideration the likelihood that an individual known to exist in the wider population is present in the sample dataset. It also considers the information that would be revealed if an individual were identified in the sample dataset by assuming each feature reveals an amount of information about the individual. Finally, it considers accuracy of data with the assumption that increasing accuracy means increasing safety.

In all cases, however, the DSF relies on a determination of acceptable Safe Data Levels. Taking some threshold values as examples:

SAFE LEVEL	SAFE LEVEL	SAFE LEVEL	SAFE LEVEL	SAFE LEVEL
5	4	3	2	1
$1.0 \leq DSF$	$0.5 \leq DSF < 1.0$	$0.2 \leq DSF < 0.5$	$0.1 \leq DSF < 0.2$	$DSF < 0.1$

It is proposed that Safe Level 5 be used as the threshold for open data.

Working through an example, consider the pharmaceutical benefits and medical benefits (PBS/MBS) dataset released by the Commonwealth Government, which covers approximately 10% of Australia's population.⁴⁰

⁴⁰ See OIC web page, <https://www.oaic.gov.au/privacy-law/commissioner-initiated-investigation-reports/publication-of-mbs-pbs-data> [Accessed 15th September 2018]

DATA SAFETY FACTOR IN PRACTICE

In August 2016, the Department of Health published a collection of Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Schedule (PBS) related data on data.gov.au. The data consisted of claims information for a 10% sample of people who had made a claim for payment of Medicare Benefits since 1984 or for payment of Pharmaceutical Benefits since 2003. The dataset was quickly removed when it was discovered that cohorts of size 1 could be isolated in the de-identified dataset.

In this case:

- MICS is 1.
- Assume accuracy is 100%.
- Coverage probability is 10%.
- Assume PIF cannot be broken down by context (space/time/relationship).

If the feature depth is 10 and the PIF is less than:

1 then DSF is less than 0.06
(Safe Level 1)

0.5 then DSF is less than 0.13
(Safe Level 2)

0.2 then DSF is less than 0.32
(Safe Level 3)

0.1 then DSF is less than 0.63
(Safe Level 4)

0.05 then DSF is less than 1.26
(Safe Level 5).

If the feature depth reduces to 5 and the PIF is less than:

1 then DSF is less than 0.13
(Safe Level 2)

0.5 then DSF is less than 0.32
(Safe Level 3)

0.2 then DSF is less than 0.63
(Safe Level 4)

0.1 then DSF is less than 1.26
(Safe Level 5)

0.05 then DSF is less than 2.52
(Safe Level 5).

Reducing the risk of re-identification (and so increasing the level of safe for data and outputs) relies on forcing an increase in the PIF, reducing the feature depth or reducing the coverage probability before access to data or release of outputs. It is unlikely that data accuracy can be significantly improved as ways of increasing data safety.

While health information is not the primary focus of this paper and additional considerations may also apply, the MBS/PBS example shows that, in order to maintain a given level of data safety, either the PIF needs to decrease or the feature depth needs to reduce.

In this example, the safest data level (Safe Level 5) would require a PIF of approximately 0.05 for a feature depth of 10. Taking the most conservative perspective, a PIF of 0.05 means the MICS is 20 or higher, so that not less than 20 records have exactly the same characteristics and are indistinguishable from one another.

These records are also de-identified. The PIF for the safest data level increases to approximately 0.1 if the feature depth is reduced to 5. Taking the most conservative perspective, a PIF of 0.1 means that the MICS is 10 or higher.

In this example, with a PIF of 1, there is no way to reduce the feature depth to a level which supports Safe Data Level 5. Consequently, if this safe level was the threshold for open data, there is no way this data could be manipulated so as to be appropriately safe for open release. Instead, access to the data can be restricted to people and projects that are safer (screened, accessed and qualified).

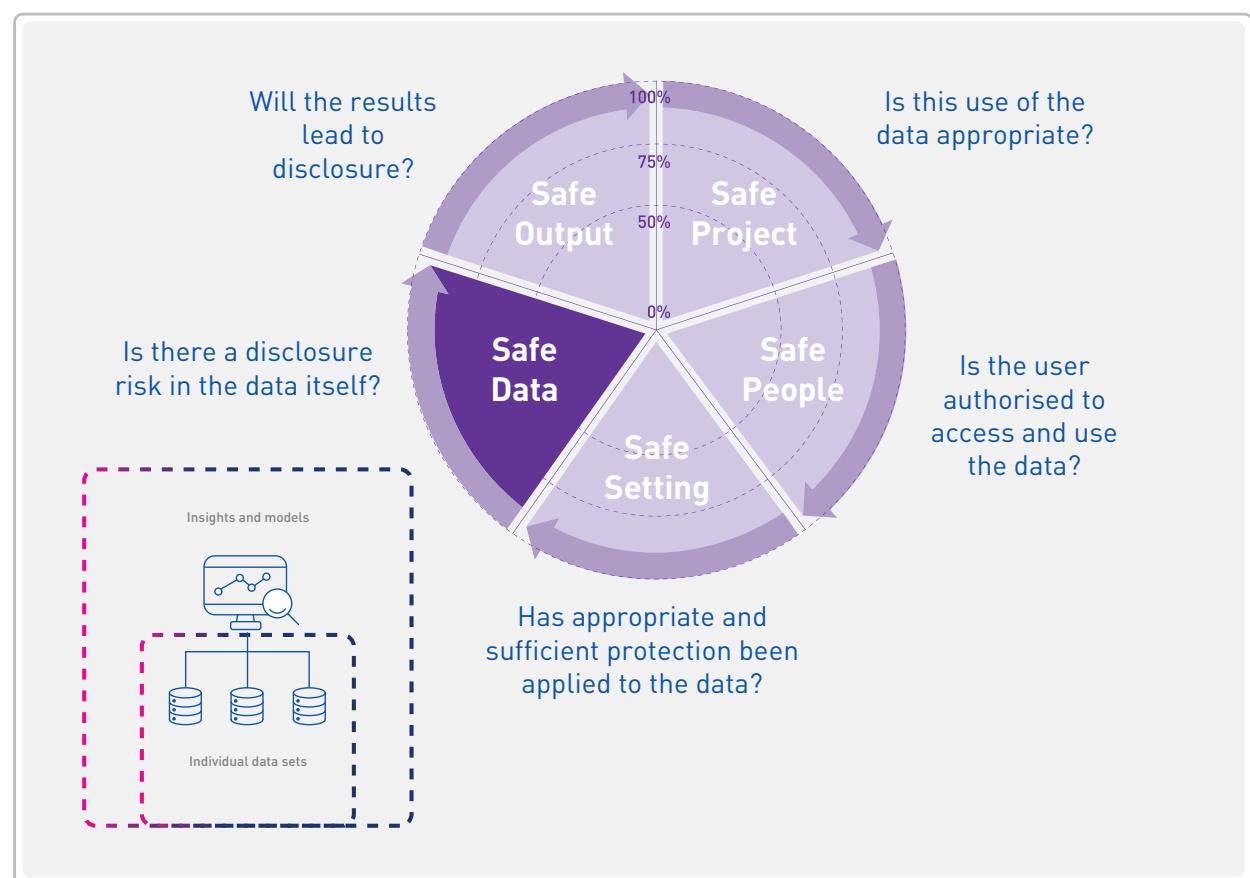


Figure 23. Data Sharing Framework with quantified Safe Data

Figure 24 shows an example of how the DSF data may be modified depending on the context of sharing with different levels of Safe People and Safe Projects. In this figure, data at Safe Level 5 is effectively open data available to users of any Safe level for projects of any Safe level. Clearly, the ability of very safe people to work on not safe projects is a matter of governance and oversight rather than rigid policy, as highlighted in Figure 22.

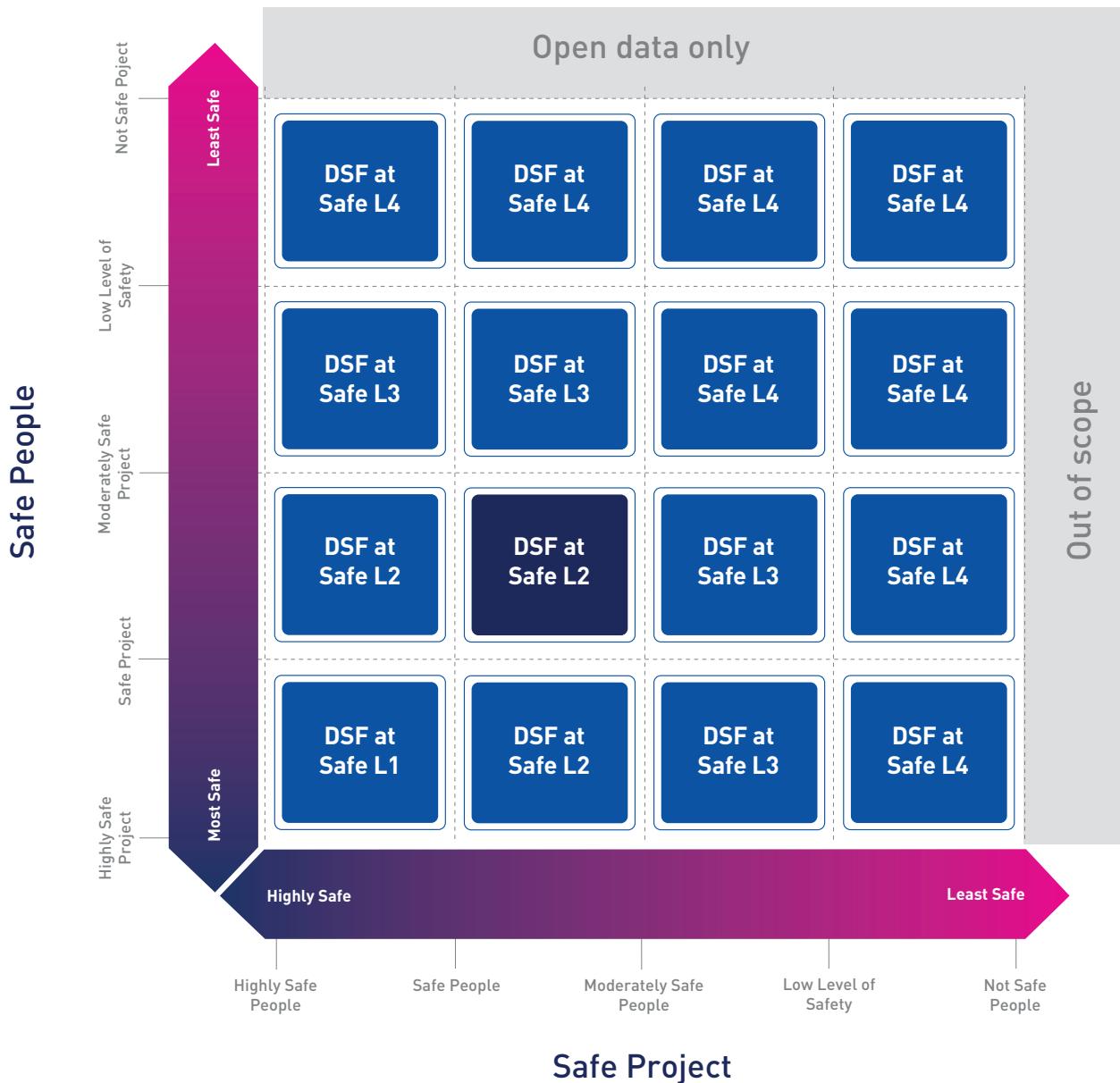


Figure 24. Example DSF settings in context of Safe People and Safe Projects

CHANGING THE SAFE LEVEL AS A PROJECT DEVELOPS

The challenge of determining the appropriate level of Safe remains the challenge of trusting the recipient of a project outcome will use the insights as expected. This is complicated by the circumstances of the recipients, as their own knowledge and personal context may reveal personal information in the results.

Added to this, a recipient's ability to find additional data in the wider world to combine with the outputs of the project increases the potential for re-identification of an individual. The challenge is again risk management. The major factors of risk explored in this section relate to the value of the data and the level of safety of the project. Figure 25 highlights the relevant Safe Outputs dimension of the Five Safes Framework.

Within the scope of a project, the major factors to consider are:

- If the PIF approaches 1 for any of the outputs of the different stages of the project.
- If the reduction in CIF can be used to preserve PIF at acceptable levels.
- If the other elements of the DSF need to be adjusted to increase the safe level of outputs.

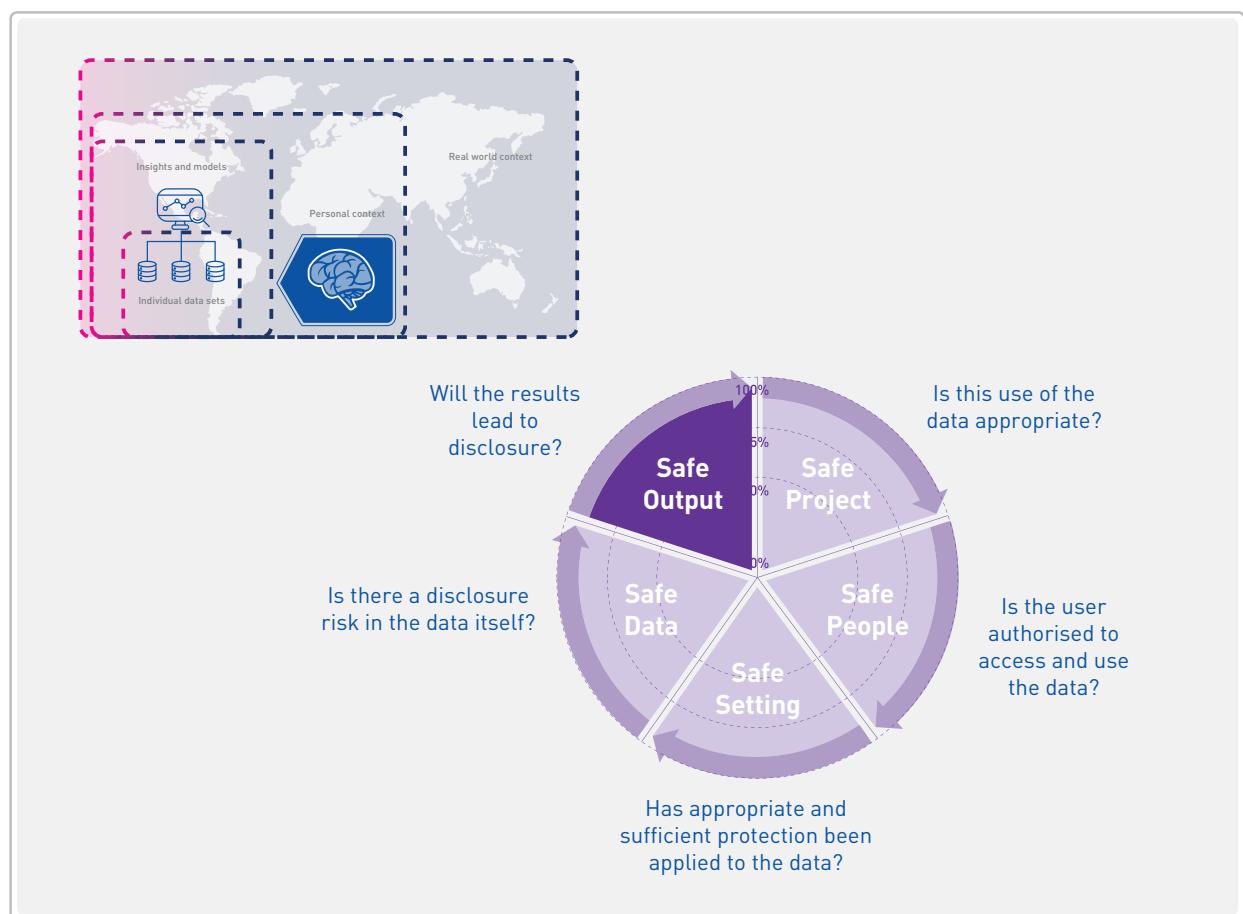


Figure 25. Adapted Five Safes Framework

Following the flow of logic in Figure 20, Figure 26 shows the relevant squares highlighted for different stages of the project if the project operates with, and reports on, de-identified data with a MICS of 1 ($PIF < 1$). To reduce the risk at each stage of the project, the DSF can be increased before outputs are released, as shown in Figure 27.

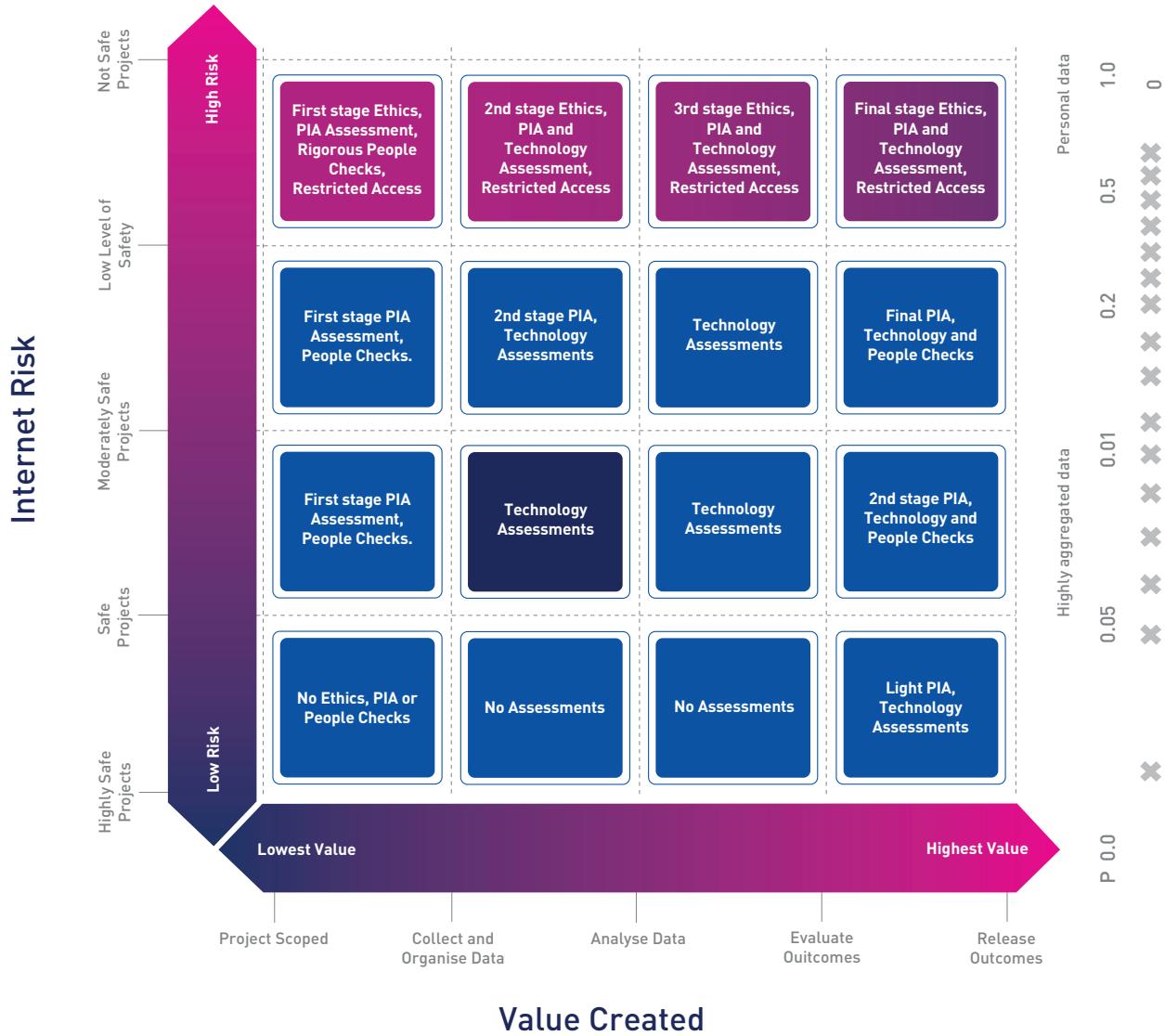


Figure 26. Project risk profile for differing levels of risk

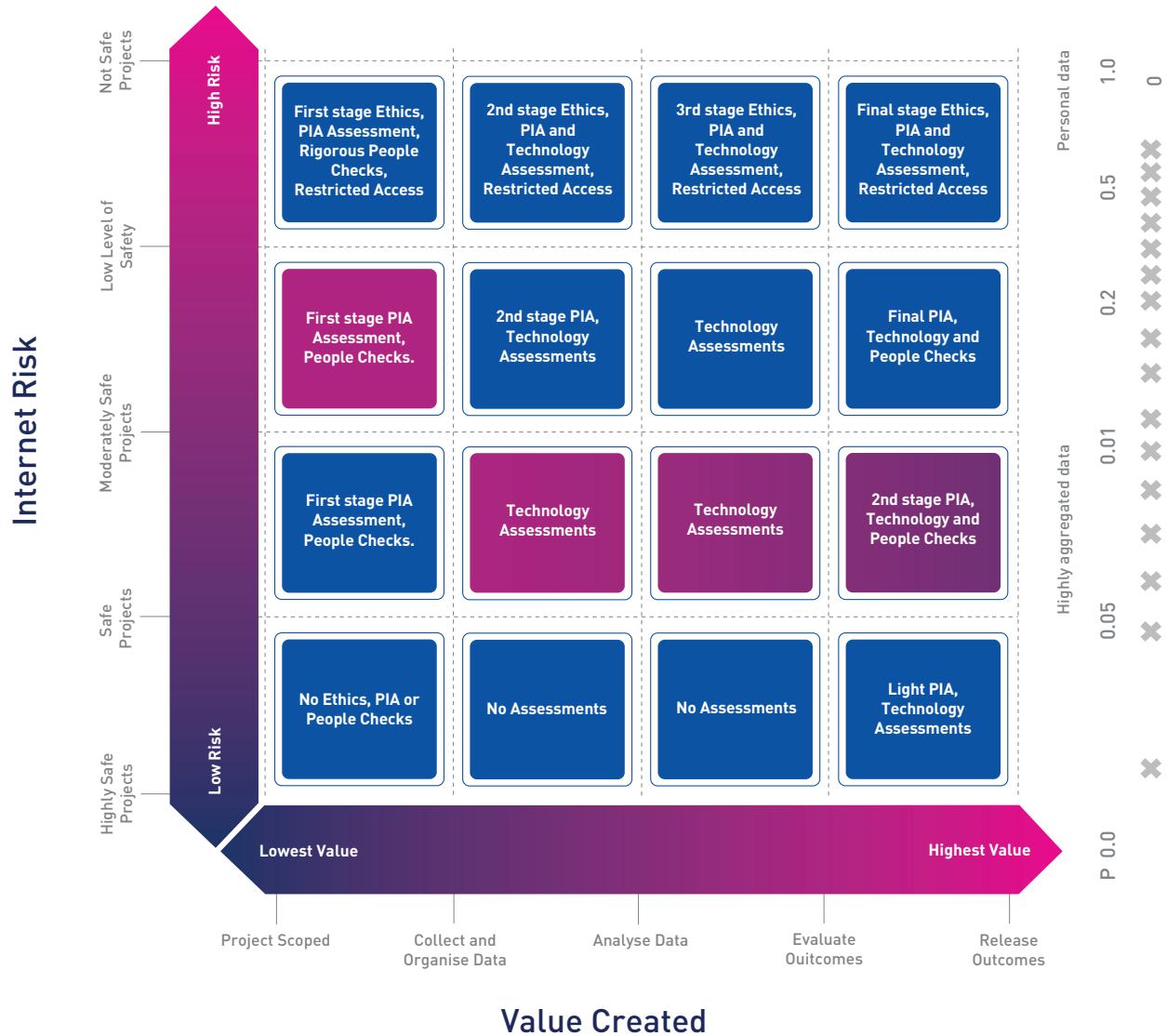
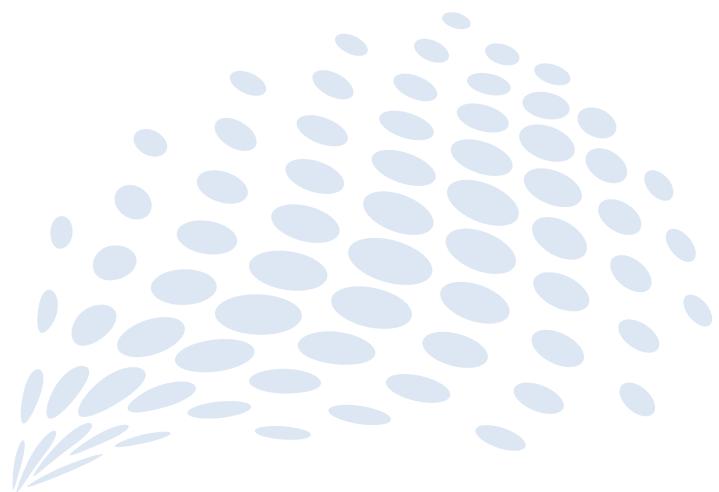


Figure 27. Risk reduction in project based on increasing DSF

06



Dealing with AI – What happens when the ‘People’ are ‘Algorithms’?

Artificial Intelligence (AI) has had a long gestation and seen surprising advances over the decades in which it has been formally recognised as a field of study. From the challenge of developing chess-playing robots in the 1970s and 1980s, AI has developed rapidly and is now deployed in various forms in online retail, recommendation systems, personal assistants, unmanned aircraft and driverless cars.

The current level of AI is still far from the level of sentient machines framed in popular culture that self-learn, self-replicate, self-analyse and ultimately challenge human beings for supremacy. However, the ability to legally place an AI program in control of a passenger vehicle being used in the real world does create the very real dilemma of who to choose to prioritise in an inevitable crash situation (passenger, pedestrian, bystander).

In the past years, AI has been introduced into financial markets and provided robo-advice in a range of professional service areas. Self-driving vehicles have gained experience and public trust and the first robo-lawyer was brought into a US legal firm to assist with bankruptcy cases.⁴¹

These questions and the increasing ability to ask powerful questions of the world through learning machines based on analysis of potentially thousands of datasets has led many to ask the question, “even if we can, should we?”

The ability for everyday objects to become autonomous leads to the question, “who is responsible when something goes wrong?” We currently lack robust frameworks to answer these questions or even understand how far AI will develop. We also lack the ability to definitively say how we will keep human beings and human judgement in the supervisory loop for these technologies.

Technology giants including Alphabet, Amazon, Facebook, IBM and Microsoft, as well as high-profile individuals like Stephen Hawking and Elon Musk, have stated that it is time to talk about the landscape of artificial intelligence.⁴² In many ways, this is just as much a new frontier for ethics and risk assessment as it is for emerging technology. Great fear exists around the possible impacts of AI, ranging from the future of work to ethical questions of machine decisions, to concerns of loss of control of our human destiny.

41 See, for example, https://www.washingtonpost.com/news/innovations/wp/2016/05/16/meet-ross-the-newly-hired-legal-robot/?utm_term=.cc5a10c1fc22

42 See World Economic Forum ‘Top Nine Ethical Issues of Artificial Intelligence’. Available online at <https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/>

ADAPTING THE FIVE SAFES FRAMEWORK FOR AI

In the world of AI, the Safe People dimension may be replaced with algorithms that process data supplied for analytical purposes (such as clustering or classification) or for the purpose of delivering smart services (such as smart lighting or smart message routing).

The environment an algorithm operates in may be very different to a human researcher and the restrictions and scrutiny placed on an algorithm may be far more intrusive than those that can be applied to a human researcher. Consequently, some of the implicit assumptions in the Five Safes Framework need to be re-examined.

The Five Safes Framework is a system model and is intended to be considered in the context of all the elements. The answer to whether a researcher (or algorithm) is permitted to access a dataset assumes that all other necessary conditions are in place. If secure facilities do not exist, this does not seem like an appropriate way to use the data.

However, this does not mean the question of whether a researcher should have access to the data changes, only that the proposed solution as a whole is not acceptable – in this case because of a failure of the Safe Setting dimension.

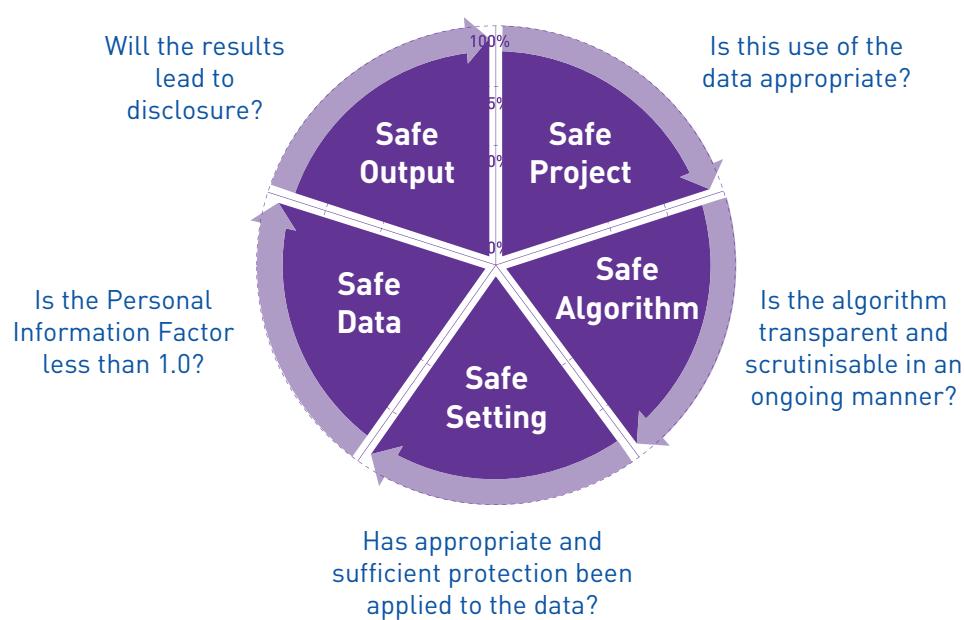


Figure 28. Five Safes Framework for algorithms

SAFE ALGORITHMS – for Safe People, this refers to the knowledge, skills and incentives of the users to store and use the data appropriately. For Safe People, ‘appropriately’ means ‘in accordance with the required standards of behaviour’, rather than level of statistical skill. For an artificially intelligent algorithm, the behaviours and associated access conditions can be enforced under many circumstances more easily than for a person, but will need supervision if adapting over time. Any biases that develop also need to be monitored.

SAFE PROJECTS – still refers to the legal, moral, and ethical considerations surrounding use of the data. Grey areas might exist when exploitation of data may be acceptable if an overall public good is realised, or with consent from the person who is provided the project outcome (knowledge), or who benefits from the AI-driven service. The safeness of the project that an algorithm undertakes should be known before application of the algorithm to the data. The challenge, however, is in discovery as the project progresses or if the project is a continuous operation rather than a discrete event.

SAFE SETTING – refers to the practical controls over data access. At one extreme, researchers (or algorithms) may be restricted to using the data in a supervised physical location or environment. At the other extreme, there are no restrictions on data accessed and linked from external sources. Safe Setting encompasses both the physical environment (such as network access) but also procedural arrangements such as the supervision and auditing regimes. When the researcher is an algorithm, the operating environment can be locked, disconnecting the algorithm from other sources of input. This does not, however, allow for any biases in the algorithm itself being evaluated or the implications of these being understood.

SAFE DATA – for Safe People, this refers primarily to the potential for identification in the data. It could also refer to the sensitivity of the data itself. When the observer is an algorithm, the context which the algorithm brings to the data can be limited through limiting access to other datasets, strictly limiting the Personal Information Factor to be less than 1.

SAFE OUTPUTS – refers to the residual risk in publications from sensitive data. There is a distinct difference to be further examined as to a single discrete output from an algorithm and something that feeds an operational loop (such as a steering algorithm or cruise control algorithm).

The underpinning concepts of the Five Safes Framework are significantly stretched when ‘person’ or ‘researcher’ is extended to an artificially intelligent algorithm. However, the basic considerations of the risk framework remain, including the Safe People and Safe Projects dimensions.

Safe Algorithms may be peer reviewed to detect bias and constantly monitored as they develop. Safe Projects may be extended to consider the real-world implications of a steering or braking decision of a self-driving vehicle.

One fundamental difference when considering the operation of an algorithm is that it may train on a set of data and then continually adapt or learn post-training during the operational phase. The Five Safes Framework implies distinct, discrete discovery-oriented analytics projects rather than a continuous operational loop; a discrete project carried out by a person who releases results which inform those who operationalise a service or system. If the Framework was a continuous process where outputs fed directly into a next loop of projects, the evaluation of Safe People, Safe Projects and Safe Data would need to be automated.

The potential for continuous learning by algorithms introduces distinct challenges. It has been cited numerous times that AI is prone to amplify sexist and racist biases from the real world^{43 44} and potentially evolve to positions well beyond those intended by developers. A Safe Algorithm must be constantly monitored for their safe level, which may change over time or be recalibrated.

In practice, the project undertaken by AI may be very small compared to the scope undertaken by a human researcher. Consider, for example, the use of Monte Carlo analysis⁴⁵, which consists of repeated evaluations of an environment under different sampled values of random variables. Each project is small, however, the results of thousands of small projects may be merged to create a deeper understanding of a process or system.

The framing questions to be considered include:

- Is it possible to apply the Five Safes Framework when the researcher is an algorithm?
- Is it possible to determine 75%, 50% or 25% safe levels for aspects of the model (see Figure 29) for an algorithm?
- Could, for example, a 100% safe level for an algorithm be described and combined with a 25% safe setting?

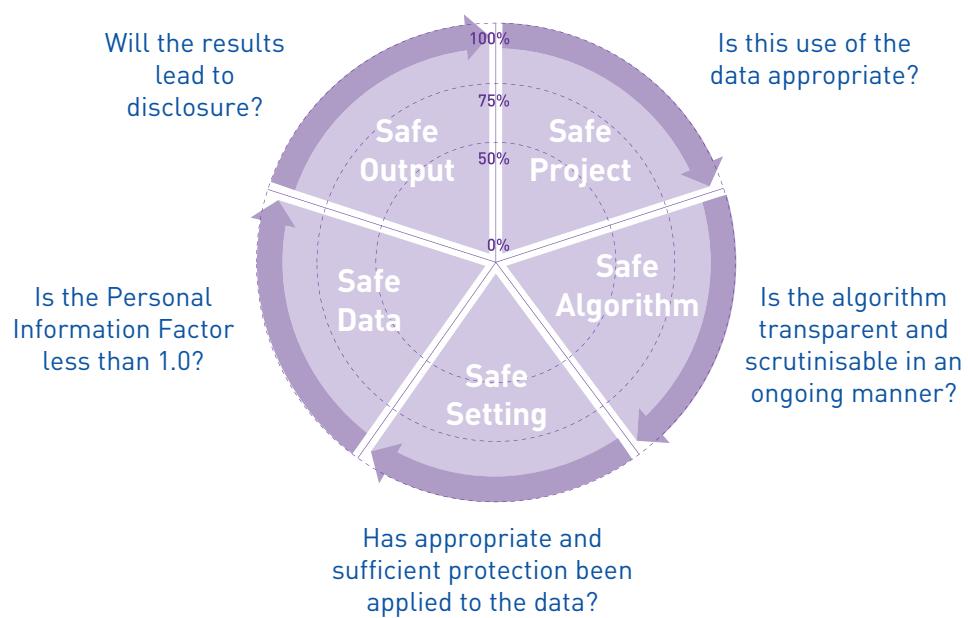


Figure 29. Quantified Five Safes Framework for AI

43 D. Cossins, 'Discriminating algorithms: 5 times AI showed prejudice', New Scientist, April 2018. Available online at <https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/>

44 H. Reese, 'Why Microsoft's 'Tay' AI bot went wrong', TechRepublic, March 2016. Available online at <https://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/>

45 See https://en.wikipedia.org/wiki/Monte_Carlo_method.

The Five Safes Framework was developed for research projects and implies sharing of data in a controlled environment, performing analytical operations on the data and then sharing the results of analysis. The simplest version of a project being simply passing through the data, with no linkage or analytical work being performed, and aggregating or anonymising the data before sharing. Many devices undertake this function today: intelligent wireless routers and mobile phone base stations, for example.

A person may interact with a mobile device, generating data that reveals personal information on preference, usage, relationships, activity and location. The mobile device's interaction with the mobile network – and the AI behind message-routing algorithms that direct traffic across networks based on some of this personal information – should be considered two different projects, and so have two different versions of Safe Setting within the Framework.

One of the implications that can be drawn from the discussion of the Framework is that several of the dimensions are highly dependent on judgement. Safe Projects are particularly dependent on a judgement-based evaluation of risk. While frameworks may be developed to help decision making in these areas, there is no unambiguous way to determine quantified levels of safe for this dimension.

Safe Setting is largely depended on restrictions applied at a technology and governance level.

The Safe Outputs dimension brings us back to the heart of the challenge of data-driven analytics: the human context of the recipients of the results of the data analysis project or the AI-driven service. For a project outcome, the challenge relates to the ability of any human (or algorithmic) recipient of these outputs to find additional data in the wider world to combine with the outputs of the data analysis project.

For the recipient of the AI-driven service, the challenge relates to the responsibility of the real-world outputs of the service.



EVALUATING SAFE ALGORITHMS

In many respects, evaluating Safe Algorithms is easier than evaluating Safe People. Evaluating Safe Algorithms requires consideration of the ways an algorithm may access or use data over time. Hard restrictions may be applied during the training of a learning or adapting algorithm, however over time, the context of these restrictions may change. Ongoing supervision is required to ensure the principles of access to data (and outputs) are maintained. Following are the minimum requirements for evaluating Safe Algorithms:

- Algorithms must be open to independent evaluation.
- An algorithm's trained or learned state must be open to periodic, independent review and evaluation.
- Access to data outside that required for the project must be defined.
- The evaluators themselves must be evaluated (in terms of Safe People) as highly ethical and be able to identify the different levels of ethical issues. They must also be highly sensitive to all forms of potential bias as any residual ethical concerns or bias may be amplified by the algorithms over time.

EVALUATING SAFE PROJECTS FOR ALGORITHMS

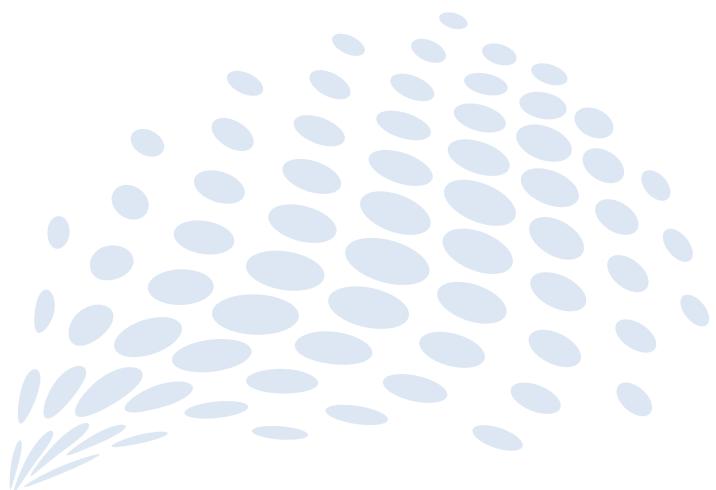
Evaluating Safe Projects still requires judgement of the purpose of the project from a risk and ethical perspective. Formally convened ethics committees accustomed to considering human researcher interactions must now consider ethics issues given that an algorithm will not exercise human judgement and will instead operate according to pre-programmed rules or according to rules learned after training. As highlighted earlier, the rules learned after training can potentially present the greatest challenge to static ethical evaluations. Following are the minimum requirements for evaluating Safe Projects for algorithms:

- Research should be designed, reviewed, and undertaken to ensure integrity, quality and transparency.
- The implications of the much faster rate at which algorithms can produce results (or partial results) must be considered.
- The implications of the limited contexts that algorithms operate within should be considered.

When evaluating Safe Algorithms and Safe Projects, it is not implied that human judgement or biases should be reintroduced to the algorithmic researcher, rather that the issues highlighted be considered.



07



Making it practical

For data sharing to work in an automated environment, the anonymisation and enforcement of Minimum Identifiable Cohort Size must operate in an automated fashion.

As shown in Figure 30, a possible architecture for the realisation of the modified Safes Frameworks relies on a three-stage process:

STEP 1: The data custodian enters information on data accuracy and coverage probability and then makes the data available via an application programming interface [API]. This API hides personal information fields according to a schema and calculates the PIF and the DSF for the dataset. Data is assumed to be hosted by the data custodian in static form rather than being streamed.

STEP 2: Users (people or algorithms) seek access to the data based on their validated Safe People and Safe Project levels.

STEP 3: Data is made available based on the Safe Data levels established by the PIF calculation.

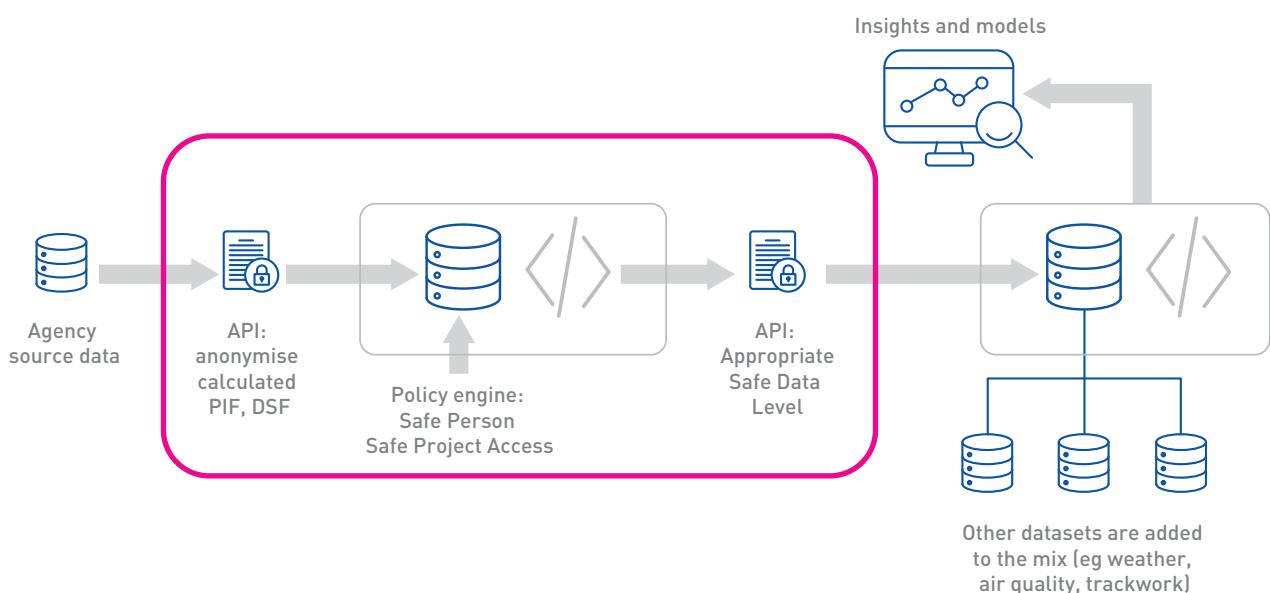


Figure 30. Possible architecture for PIF processing

When a data custodian provides access to data, it is important to identify the DSF and the associated safe level. Subsequent to the DSF assessment by the data custodian, data can be made available to different users, transformed through de-identification, obfuscation or perturbation to reduce the PIF. Figure 30 shows a possible API-based architecture for PIF processing.

EXAMPLES OF DSF USE

This section provides working examples of how the PIF and DSF approach can be applied to minimise the risk of re-identification. In all cases, the Safe Data levels are assumed to be:



The first example is based on increasing the MICS for a large dataset with small number of features. Whilst this example inherently relies on the defences of k-anonymity to reduce the risk of re-identification, the level of aggregation changes with Safe Data levels.

The second example covers increasing MICS size to increase the Safe Data level, breaking context and reducing coverage probability. The combination of factors allows greater defences to be engaged to reduce the risk of re-identification for a large, feature-rich dataset.

The third example explores the use of lower accuracy data and a small number of features. The protection mechanism takes the form of breaking context with the one spatial feature.

EXAMPLE 1: INCREASING MICS

 **AMIT** is a data custodian for a transport department and holds data on train passenger journeys, which is to be made available via API. The data covers passenger journeys over the course of three months, is de-identified and recorded at unit-record level. The data has five independent features: origin station, destination station, start time, end time and journey date (the feature depth is 5).

Amit plans to make the data of a randomly selected 10% of the total journeys available (coverage probability is 10%) and assesses the data to be approximately 98% accurate.

The MICS is determined to be 1 and the PIF is calculated to 0.90 (due to the presence of other cohorts of size 1). The DSF for this dataset is calculated to be greater than 0.14 (Safe Level 2) in its de-identified form.

 **CHARLIE** wants access to train journey data for personal curiosity and is uncredentialled. He is therefore only able to access data at Safe Level 5. The API makes the data available at this higher safe level by increasing the MICS based on the full set of five features.

In this example, the MICS is increased to 8, creating a PIF 0.1 (due to the presence of other cohorts of size 8). The DSF for this dataset is calculated to be greater than 1.26 (Safe Level 5) in its de-identified form.

 **BARBARA** wants to access train journey data for a project on urban development. In order to access the data in its most granular form, Barbara would need to be credentialled to access data at Safe Level 2. As Barbara has the appropriate credentials for Safe Level 4 but not Safe Level 2, the API makes the data available at this higher safe level by increasing the MICS based on the full set of five features.

IN THIS EXAMPLE, THE
MICS IS INCREASED TO

4

CREATING A PIF

0.2

(DUE TO THE PRESENCE OF OTHER
COHORTS OF SIZE 4)

THE DSF FOR THIS
DATASET IS CALCULATED
TO BE GREATER THAN

0.63

(SAFE LEVEL 4) IN ITS
DE-IDENTIFIED FORM

EXAMPLE 2: INCREASING THE MICS, BREAKING CONTEXT AND REDUCING COVERAGE PROBABILITY

 **ALICE** is a data custodian with data on patient access to Pharmaceutical Benefits, which is to be made available via API. The data covers 12 months for all of Australia, is de-identified and recorded at unit-record level. The data has 50 independent features, including pharmacy name, prescriptions presented, prescriptions filled, dosage, method of payment, time of visit, date of visit and location of pharmacy.

Alice plans to make the data of a randomly selected 10% of the total population available (coverage probability is 10%) and assesses the data to be approximately 95% accurate.

The MICS is determined to be 1 and the PIF is calculated to 0.85 (due to the presence of other cohorts of size 1). The DSF for this dataset is calculated to be greater than 0.01 (Safe Level 1) in its de-identified form.

 **BRIAN** wants to access the data for a project on social policy. In order to access the data in its most granular form, Brian would need to be credentialled to access data at Safe Level 1. As Brian only has credentials for Safe Level 3, the API makes the data available at this higher safe level by increasing the MICS based on the full set of 50 features.

In this example, the MICS is increased to 15, creating a PIF 0.06 (due to the presence of other cohorts of size 15). The DSF for this dataset is calculated to be greater than 0.21 (Safe Level 3) in its de-identified form.

 **CHEATNA** wants access to the data for a high school project and is uncredentialled. She is only able to access data at Safe Level 5. The API makes the data available at this higher safe level by:

REDUCING THE DATA AVAILABLE TO

1%

1% COVERAGE PROBABILITY
(PROVIDING A RANDOM SUB-SAMPLE
OF THE WHOLE DATASET)

BREAKING THE FEATURES LINKS TO CREATE

TWO

SEPARATE SETS OF CONTEXTUAL AND
PERSONAL FEATURES

INCREASING THE MICS
OF THE PERSONAL
FEATURES TO 9, CREATING
A PIF_NOCONTEXT OF

0.09

INCREASING THE
MICS OF THE CONTEXT
FEATURES TO 4,
CREATING A CIF OF

0.21

The overall PIF becomes 0.019. The DSF for this dataset is calculated to be greater than 1.01 (Safe Level 5) in its de-identified form. The equivalent MICS for the de-contextualised dataset is approximately 50.

EXAMPLE 3: LOWERING ACCURACY AND REDUCING FEATURES

❶ **ANGELA** is a data custodian and has data for electricity consumption, which is to be made available via API. The data covers household level consumption over 12 months, is de-identified and recorded at unit-record level. The data has three independent features: suburb, power meter reading and off-peak power consumption. Household address, meter identifier and suburb are not linked in the dataset.

Angela plans to make the data of a randomly selected 1% of the total readings available (coverage probability is 1%) and assesses the data to be approximately 99% accurate as it is automatically recorded.

THE MICS IS DETERMINED TO BE

1

AND THE PIF IS CALCULATED TO

0.99

(DUE TO THE PRESENCE OF OTHER COHORTS OF SIZE 1, BUT WIDE RANGE OF VALUES OF EACH FIELD)

THE DSF FOR THIS DATASET IS CALCULATED TO BE GREATER THAN

0.32

(SAFE LEVEL 3) IN ITS DE-IDENTIFIED FORM

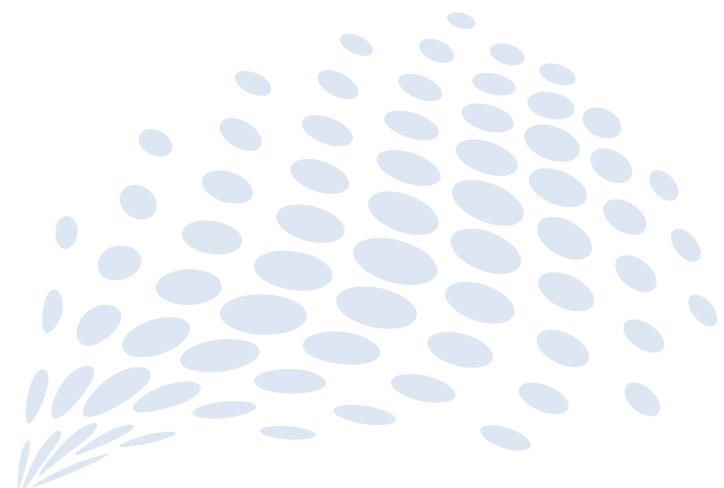
❷ **BETHANY** wants to access data for a population study. In order to access the data in its most granular form, Bethany would need to be credentialled to access data at Safe Level 3. As Bethany has the appropriate credentials for Safe Level 3, she gains access to the data in unit record level form.

❸ **CONNOR** wants data access to compare to his own home consumption. He is only able to access data at Safe Level 5. The API makes the data available at this higher safe level by:

- Breaking the features links to create two separate sets of contextual (suburb) and personal features (meter reading and off-peak consumption).
- Increasing the MICS of the context feature to 3, creating a CIF of 0.3 (due to the presence of other cohorts of size 3).
- Leaving the MICS of the personal feature unchanged at 1 so *PIF_nocontext* remains 0.99.

In this example, the PIF is 0.30. The DSF for this dataset is calculated to be greater than 1.07 (Safe Level 5) in its de-identified form.

08



Limitations of the approach

While many of the challenges of data sharing have been broadly identified as factors associated with trust, the approach explored in this paper has been focused on privacy-preserving data sharing.

This paper has also attempted to quantify a range of concepts that are not well defined, including the impact of data quality, the risk that a known individual is included in the dataset and the risks associated with the re-identification of an individual based on the number of features in the dataset. The measures described are all heuristic approximations of risk leading to a heuristic measure of data safety.

To be of benefit, even heuristics metrics need to be practical, measurable and finite. The underlying assumptions and limitations also need to be understood.

LIMITATIONS ON ACCURACY, FEATURE DEPTH AND COVERAGE PROBABILITY PARAMETERS

Limitations of the Accuracy parameter

In science and engineering, the accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's true value. The precision of a measurement system, related to reproducibility and repeatability, is the degree to which repeated measurements under unchanged conditions show the same results. Although the two words are commonly used interchangeably, they are different concepts in the context of the scientific method.⁴⁶

For the purposes of this paper, we have defined Accuracy as the ratio of correct values to total values for all features in the sample population represented in the dataset. This simple definition assumes that a value is either correct or not correct.

For binary-valued features, this is a good measure of Accuracy. For values such as 'age in years' or 'height in centimetres', a difference of 1 unit may still allow an inference to be made through probabilistic matching. By considering features such as 'height in centimetres' to be correct within plus or minus five centimetres or 'age in years' to be within plus or minus one year, the Accuracy of the dataset can be increased significantly.

This approach changes the resolution of the data by reducing the values that can be distinctly represented (for example, 199cm, 200cm, 201cm and 202cm become 200 +/- 1cm and 202 +/- 1cm). It may also have the effect of increasing the minimum identifiable cohort size, as fewer distinct values exist for each feature.

Determining the Accuracy of a sample dataset is a non-trivial exercise. It requires careful mapping of held data with verified exact values. This can often be a laborious process that is approximated by taking random samples of data and manually validating the correct value to determine Accuracy. Accuracy may also change over time as systems vary or different processes are used to collect data. In practical systems, a snapshot in time or a crude estimate may be all that can be expected for a measure of Accuracy.

⁴⁶ See for example https://en.wikipedia.org/wiki/Accuracy_and_precision

Limitations of Feature Depth

The Data Safety Factor considers what would be revealed if an individual were identified in the sample dataset by considering each unique Feature to reveal an equal amount of information about the individual. This assumes that all Features carry equal information and that Features are unrelated. The sensitivity of features is not considered.

Each Feature carries information. In a binary-valued feature set, it would be yes/no value for each Feature, so one bit of information would be known for each Feature.⁴⁷ If each Feature was multi-valued, with, for example, one of 16 possible severity values, all of which are assumed to be equally likely, then four bits of information would be gained from each Feature.

For multi-valued Features, Feature richness may be better expressed as the sum of all the information (in bits) contained in all the Features. The challenges of this extra level of specificity are many; however two of the main challenges relate to the ability to determine the exact information content of each Feature, and introducing units of bits to a heuristic measure that is otherwise without units.

Secondly, consider the inter-relatedness of each Feature. At first glance, it may seem that 'age in years' and 'height in centimetres' are unrelated. To date, there has never been known a baby of age one which is 201cm tall. Whilst some age and height ranges will not provide mutual information⁴⁸, what we know of human physiology can be used to code rules which can infer age ranges from height.

Similarly, if we can, with high probability, infer or directly deduce one Feature from several others, that Feature carries little or no information. For example, if we know 'born in Australia', 'mother is Australian' and 'father is Australian', we can infer 'is Australian'. Similarly, if we know 'age at December 2018', we can deduce 'age at December 2017'.

Mutual dependencies are rarely as straightforward as the simplistic examples above, so calculating Feature Depth is challenging without an in-depth analysis of the Feature dependencies. The simplistic approach of counting Features provides a crude but conservative measure.

Limitations of Coverage Probability

The concept of Coverage Probability takes into consideration the likelihood of the presence of an individual known to exist in the wider population to be present in the sample dataset. If the sample dataset is chosen randomly, the probability of the individual being in the dataset is `sample_size / population_size`. Datasets are, however, rarely ever assembled from random data and the wider population is sometimes difficult to describe.

Consider the Australian national census, which takes place on a five-year basis, and estimates population on a single day selected each time the census is run. The coverage of the national population is often reported as high, but the ability to know the actual population is limited by the lack of an objective measure.

47 For a closed system with equally likely values of all events, the information carried (in bits) in one event or record is $\log_2[1/P(x)]$ where $P(x)$ is the probability of the event or record occurring. See, for example, a discussion on Information Theory and Entropy, https://en.wikipedia.org/wiki/Information_theory

48 In information theory, the mutual information of two random variables is a measure of the mutual dependence between the two variables. It quantifies the amount of information (in bits) obtained about one random variable, through the other random variable. See, for example, https://en.wikipedia.org/wiki/Mutual_information

The estimate of the national population is based on the census itself. While the rate of return of census forms gives one measure of Coverage Probability, knowing if people were omitted completely is impossible unless other measures of the population are considered. Secondly, if data is not collected randomly, the measure of likelihood of an individual being in the sample dataset is no longer based on the simple ratios outlined above. The total population is then described by the set of factors that describe the dataset (for example, age, presence of disease), which datasets are able to be collected (for example, individual hospital records) and the conditions under which the data was collected (was the individual a patient in the hospital from which the data was collected).

These factors make Coverage Probability difficult to determine in many cases. In the example of the MBS/PBS dataset release described earlier, the sample dataset was approximately 10% of the total population represented in the total dataset. Whilst the total dataset may not have been the entire national population, the estimate of 10% of such a large dataset may be a sufficient approximation for determining risk.

The final consideration is the time-varying nature of a population. For a five-year census, what is meant by 'population' is reasonably clear. When considering people travelling on public transport, the total population changes significantly each day, with weekends having a very different population of travellers compared to weekdays. The total unique number of travellers during a year will also be different to the number of travellers averaged by day over the course of the year. The question of definition of what is meant by population must be considered.

Limitations of a PIF-based approach to Safe Data and Safe Outputs

Both the PIF and the DSF are heuristic measures rather than robustly quantifiable terms. Heuristics are open to the challenge that they may contain assumptions and approximations. This is true in the case of the PIF and DSF described in this paper and some of these assumptions and limitations are described above. Despite these limitations, it is proposed that the DSF based on the PIF be used as the basis for determining safe levels of data for different levels of Safe Projects and Safe People.

It is important to note that the Safe Levels themselves are subjective. Currently, decisions are made about the level of Safeness for release or use of data based on subjective criteria. The difference with the heuristic approach is that it provides a framework to addresses many of the major factors of concern when subjectivity is considered.

It is also relevant to highlight the need for a conservative bias when determining how to use the PIF or DSF for Safe Data and Safe Outputs. The harm caused by releasing personal information to the public will be different to the harm if a single researcher in a laboratory environment is able to identify a person they know from the data they are reviewing. Privacy legislation makes a distinction between the likely harm caused by misuse of sensitive data (such as religion or sexuality) and non-sensitive data (such as names). This should also factor into the assessment of risk and risk appetite in terms of Safe People and other parameters of the Five Safes Framework.



Conclusions

Our world is relentlessly digitising, hyper-connected and personalised. The ability to effectively harness data and analytics is essential for Australia to maintain its place in the world and address issues of an ageing population and weak productivity growth as well as some of the greatest challenges we will face in health, our natural environment and the future of work. Embracing data-driven technologies for smart cities, smart factories, smart homes and smart government all require us to understand, frame and then address the challenges of data sharing.

Better information governance is required in Australia to provide a foundation that can influence a national regulatory response to the accelerating rise of data analytics and data-driven disruptive developments. Our future must include a data regulatory framework informed by well-considered governance practices and solid understandings of the risks and benefits of data analytics.

A critical aspect of trusted data sharing is the effectiveness of transfer of knowledge and insights, which is in turn dependent on growing data and algorithm literacy. Transparency is necessary but not sufficient to ensure efficacy of data insights and outputs. With ever more decisions being made based on machine learning – including insurance policies, credit rating, loan applications and bail conditions – ever greater accountability and transparency in data and associated algorithms are needed.

Legislation and regulation by governments is highly likely and necessitates industry practice guidelines and information governance frameworks to ensure what is being done is truly in the community's best interest.

Data sharing extends across government and non-government sectors, so any effort to improve data sharing must be able to address the challenges for government-funded service delivery by the for-profit and not-for-profit sectors.

Efforts to progress data sharing must include active considerations of appropriate use and clear demonstrations of transparency to ensure community and public trust is honoured and maintained. Demonstrations of trust maintenance are essential from the individual's perspective and will ensure corporate and government interests are identified and seen to be clearly managed.

Systemising and standardising algorithmic calculations of safe data sharing, with independent verification demonstrating trust, efficacy and benefit to the community, will be required for Australia to benefit from evolving digitally driven developments. As Australia and many other nations grapple with the need to efficiently deliver personally tailored smart services, the demand is growing for safe data sharing at scale, in real or near-real time.

This paper has presented frameworks that address major issues of data sharing, considering issues of re-identification risk, data quality and outcomes frameworks. Whilst often heuristic in approach, the frameworks presented demonstrate useful ways to consider the challenges of data sharing and hopefully provide a basis to anchor principles-based data sharing and governance frameworks.

Trusted data sharing frameworks are essential for Australians to address the challenges of our future and maximise the return on investment in smart systems and data-driven technologies. The ability to deliver benefits across the community using the algorithmically generated insights and outcomes from data sharing come with a responsibility for efficacy and ethical consideration of those outcomes, including the increased capability to harvest and analyse people's data. Practical and independently verifiable trusted frameworks with the critical ability to automate the safe sharing of data, as presented in this whitepaper, will play a critical role in ensuring benefits from safe data sharing are realised.

Appendix

Appendix A – International examples

GLOBAL ALLIANCE FOR GENETICS AND HEALTH AND THE GENERAL DATA PROTECTION REGULATION

The development of precision medicine requires effective safe data sharing at scale. International data sharing frameworks have been developed for sharing genetic data through the Global Alliance for Genetics and Health (GA4GH).⁴⁹ The GA4GH charter is designed to enable responsible genomic data sharing for the benefit of human health. The GA4GH is a policy-framing and technical standards-setting organisation with international membership that is seeking to enable responsible genomic data sharing within a human rights framework.

The General Data Protection Regulation (GDPR) took full legal effect across the European Union (EU) on 25 May 2018. The GDPR has a number of implications for data sharing and international research involving the collection, use and cross-border sharing of people's personal data. The GA4GH has published a useful guide highlighting areas where the GDPR affects international data sharing and health research.⁵⁰

The GDPR seeks to change the ways in which organisations both within and outside Europe collect, use and share personal data. The GDPR recognises rapid developments in digital technology have increased the scale, scope and speed at which personal data are collected, used, analysed and distributed, thereby necessitating a stronger legal framework that enhances the rights of data subjects.

The GDPR regulates the processing activities of two key actors: (i) data controllers, meaning persons or entities that determine the purposes and means of processing personal data, such as companies, researchers, universities, and (ii) data processors, referring to persons or entities that process personal data on behalf of a data controller, such as cloud providers and research collaborators. The GDPR defends the data protection rights of data subjects, who are most likely to be research participants in the health research context.

INTEGRATED DATA INFRASTRUCTURE (IDI) NEW ZEALAND AND SCOTTISH HEALTH INFORMATION SYSTEM (SHIP)

The New Zealand Government has demonstrated the feasibility and utility for policy evaluation, planning and research from integrating people-centred data from a range of government agencies. This has been achieved through the proactive leadership and vision for using data for better decision making in government.

The IDI is led and managed by Statistics NZ, working effectively with a clear authorising environment and mandate within this single jurisdiction. This data integration infrastructure and analytical work in harvesting, cleaning and integrating data from multiple sources provides safe and secure access for approved public good projects.⁵¹ The IDI in New Zealand compares favourably with other international examples from Scotland.⁵²

49 The Global Alliance for Genomics and Health, <https://www.ga4gh.org/>

50 Data Sharing implications from the EU General Data Protection Regulation (GDPR), <https://www.ga4gh.org/news/EcitNA0tSxyA1Ley50WpkQ.article>

51 The New Zealand Government's Integrated Data Infrastructure, <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/>

52 The Scottish Health Information System (SHIP), <https://www.nhsresearchscotland.org.uk/research-in-scotland/data/health-informatics>

The Chief Scientist, government and universities have invested in developing sophisticated data integration and analysis capability for five million people in Scotland. The Scottish Health Information System (SHIP) has pioneered a nationwide approach to data sharing, and, similar to New Zealand, they successfully and proactively engaged their community on the processes and practices for safe data access and use. This has enabled data to be considered a community asset, with Safe Data and Safe Project processes accepted by the community. This approach supports a data-driven evidence base being available for more efficient government decisions across a range of human services, including education, justice, welfare and other health and social services.

The difference between Scotland and New Zealand is that, while both jurisdictions are fully committed to safe and secure data access and use, the Scottish example has invested heavily in enabling the source operational systems to provide the necessary data feeds, with automated processes developed for harvesting and transforming the data for analytical use and safe projects.

THE HEALTH DATA RESEARCH UK (HDR UK)

The HDR UK is the national institute for data science in healthcare. Building on the work of the Farr Institute, the HDR UK was established in late 2017 to transform health research by applying cutting-edge data, science and computational techniques to answer questions on dynamic, multidimensional health and wellbeing data. Based in London, the HDR UK is a collaborative network of a further six sites across the UK, spread across Wales, Northern Ireland, the Midlands, Cambridge, Oxford and Scotland. As a UK-wide network, each of the nodes harvests data off the source-operational government and non-government systems using advanced computing and privacy-protecting extract, transform and load (ETL) processes and infrastructure to enable integrated data to be made available for approved use across organisational silos.⁵³

ESTONIA'S DIGITAL SERVICES ECONOMY

Over the past 20 years, Estonia has built a digitally enabled society with the most technologically advanced government in the world. This has supported a growing and prosperous society for its 1.3 million residents, in which practically every government service is paperless and performed electronically. As a result, the Estonian government and society benefit from digitally enabled services but are highly dependent on the underpinning data and information systems. The recognised vulnerability and risk of cyber-security and military attack resulted in an agreement with Luxembourg, signed in June 2017, that established an out-of-country secure data storage facility, referred to as the 'data embassy'.⁵⁴

Historical events have taught Estonians valuable lessons regarding the limits of privacy protection, principally in conditions where those in positions of power have little or no respect for the rule of law and the fundamental right to privacy might not be respected or have any relevant meaning at all. As a result, Estonians take privacy very seriously and have maintained it as a key topic in all discussions surrounding the development of Estonian e-government services, as well as dialogues involving economic stability, resource maximisation and the improvement and preservation of quality of life for Estonian citizens.

53 The Health Data Research UK, the peak UK health data science institute, <https://www.hdruk.ac.uk/>

54 Estonia, the most advanced digitally enabled society in the world, <https://e-estonia.com/>

Before any digitally enabled government could be established or considered in Estonia, the Estonian Government had to establish trust with its citizens to ensure confidence in entrusting private data to government. It was essential that citizens could faultlessly rely on government systems to protect their privacy in all instances. The success of Estonia's digitally-enabled government is based on three pillars: citizen ID, digital signature for citizens, and appropriately designed and engineered secure and privacy protecting information systems, named X-Road.⁵⁵

FDA SENTINEL – MONITORING MEDICAL PRODUCT SAFETY ACROSS THE WHOLE OF THE USA

The United States Department of Health and Human Services, Food and Drug Administration (FDA), has a legislated mandate through the Food and Drug Administration Amendments Act of 2007 (FDAAA), requiring the FDA to collaborate with public, academic and private entities to develop methods for obtaining access to disparate data sources and to validate means of linking and analysing safety data from multiple sources. The FDA chose to expand its existing post-market safety surveillance system to actively gather information using a distributed system, allowing data to be retained in local systems, with queries able to be run on the nationwide data holdings. A key benefit from this distributed approach is patients' sensitive and personally identifying details are kept behind an organisation's local firewalls in their existing protected environments, protecting privacy while liberating the ability to analyse the data by better enabling a data owner's involvement.⁵⁶

The FDA Sentinel has access to hundreds of data sources and is a multipurpose federated data network. It is arguably the largest robustly curated database in the world. FDA Sentinel has feeds from 17 health insurance companies and one national hospital system, with 67 million individual's records available for aggregated responses to queries. It operates from a combination of automated and semi-automated feeds from the distributed nodes, each of which requires holes through the firewalls. The FDA Sentinel initiative has been running for seven years, with the last two or so years moving from pilot to a live and fully operational status. The ETL is loaded every quarter, with rigorous comparison and checking of ETL to ETL occurring to have confidence in data completeness and quality.

The FDA has contracted not-for-profit organisation Harvard Pilgrim Health Care Institute, based in Boston, to run the FDA Sentinel Coordinating Centre with 70 staff and a budget of US\$12 million per annum. Conformance checking of the data queries is a critical and ongoing task, with over 1,400 data checks prior to release of the outputs. Five full-time employees are dedicated to data checking the outputs, rates and counts. Ten key staff members are responsible for the preparation and release of the queries run on the federated data system under the authority of the FDA.⁵⁷

The FDA Sentinel Coordinating Centre has a strong control and coordination function, with the legislated mandate to direct and compel the data owners to respond. The legislation is necessary but not sufficient for success. The success of FDA Sentinel is demonstrated by the quality, completeness and consistency of the data results and queries, which is built on trust and collaboration of partners. Partner organisations can use FDA Sentinel for their own projects.

55 J. Priisalu and R. Ottis, 'Personal control of privacy and data: Estonian experience', *Health and Technology*, 2017, 7(4), pp. 441–51. Available online at <https://doi.org/10.1007/s12553-017-0195-1>

56 'Report to Congress, The Sentinel Initiative – A National Strategy for Monitoring Medical Product Safety', August 2011.

57 Associate Professor Jeffery Brown, Associate Director FDA Sentinel, speaking at the International Population Data Linkage Network (IPDLN) Conference, Banff, September 2018.

Data standardisation and ongoing rigour is essential and resource-intensive. Data changes are expensive. The Coordinating Centre is responsible for the Common Data Model, which is applied consistently and rigorously across all of the distributed information systems. There are very strict change controls and a conservative approach taken to minimise any changes to the Data Model as even a small change to the Common Data Model has significant ramifications on each of local distributed data systems.

KEY OBSERVATIONS FROM THE FDA SENTINEL PROJECT

1. FDA Sentinel is seen as a *national resource* for both the federal government, the private for profit sector, regulatory and compliance sector, and the university research and scientific community and not-for-profit sectors.
2. FDA Sentinel has had extensive public and private consultation, with success achieved and made possible by the collaboration and full participation of the healthcare community.
3. The Common Data Model has been a major undertaking and required sensitive and intense negotiation and consultation to establish. Ongoing maintenance involving all the parties has become easier over time due to the trust established through the Sentinel Coordinating Centre.
4. While a Common Data Dictionary exists, rigorous data checking for conformance is required and ongoing.
5. Any change to the Data Model is a very significant undertaking and consequently, changes are kept to a minimum. All changes go through thorough change management review and checking processes, i.e. ETL, analytical review, coding, testing and implementation.
6. Privacy and information security are expensive, with resultant questions of who pays, who enforces the policies and who validates security.
7. Given the use of public money, there is the question of dissemination of results and whether there is full transparency or not. What happens to the outputs or results? Are they published or not? Are they published with the names of participating organisations or not?
8. Extensive testing is done with partners.
9. FDA Sentinel has its own dataset to test data quality against.
10. FDA Sentinel code sets and data models (specifications) are all published.

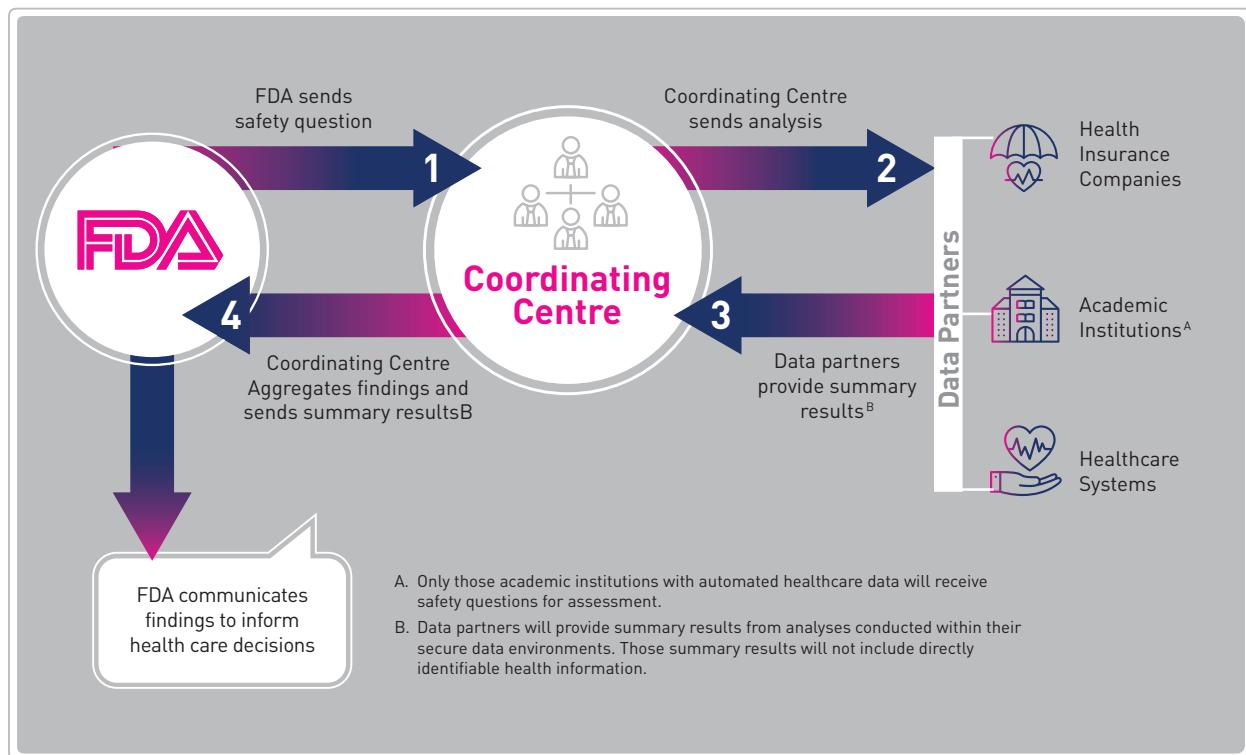


Figure 31. Overview of the Mini-Sentinel Safety Question Assessment Process

EXAMPLES OF NATIONAL DATA SHARING PROGRAMS

There are notable national data sharing examples from the health and medical sector overseas that have been used for the development of personalised services. In Scotland, for example, there is operational use of data from a large percentage of the population run through the Scottish Government's National Health System. In Canada, there is a non-government program operating out of Ontario. In Australia, the nationwide Population Health Research Network (PHRN) data sharing and access partnership operates between the Australian Government and the university sector.

SCOTLAND'S SHARE PROGRAM

The Share Program for the Scottish National Health System⁵⁸ provides an example of what can be done in partnership and with the trust of the public. The Scottish Government has taken a proactive yet personalised approach by seeking individual pre-consent for the sharing and use of people's deidentified health records in a secure manner.

Currently, 220,000 people are registered to share data that is securely stored and used in deidentified form to investigate better treatment and cures, and to improve the health system. The Share website provides details on longitudinal studies including genetic and chronic disease studies seeking to address a range of health issues.

Through the establishment of a similar research registry in North West London,⁵⁹ people in the Scotland and England registers are now also able to separately participate in health trials that would need direct contact and follow-up from the approved researchers.

58 The Scottish National Health System 'Share' initiative <https://www.registerforshare.org/index.php>

59 North West London's health research registry <https://www.registerfordiscover.org.uk/>

THE ONTARIO HEALTH STUDY

The Ontario Health Study⁶⁰ is an example of a pioneering non-government initiative lead by scientists and clinicians at universities, hospitals and research institutes.

Based on Ontario, Canada, this program operates as an ongoing research platform, with Ontario-based participants recruited as a large-scale research cohort from 2009 up until 2017. This is a consent study, with participant's health information harvested off a number of operational systems, deidentified and securely stored and able to be accessed by approved researchers.

This shared data resource supports a range of research studies including how lifestyle, environment, genetics and family history affect health over time, and evaluation and monitoring for better targeted interventions and strategies for the prevention, early detection and improved treatment and outcome of diseases.

The Ontario Health Study requires a participant's agreement to share their data in a safe and secure manner, with strict governance and separation controls.

The Ontario Health Study's data is seen as a fundamental asset to the Ontario community. It operates with a social licence to support a range of research studies, including the monitoring and evaluation of heart disease and stroke, cancer, obesity and diabetes, respiratory health, aging, hearing, nutrition, mental health, vision, neurology (nervous system related diseases), dentistry and the effect of the environment on health.

The Ontario Health Study is also now part of a recently established Canada-wide national health research platform, the *Canadian Partnership for Tomorrow Project*.⁶¹

This collaborative data sharing initiative across eight Canadian provinces supports ongoing research and cohort studies integrating data safely and securely.

60 The Ontario Health Study <https://www.ontariohealthstudy.ca/en/home>

61 The Canadian Partnership for Tomorrow Project <http://www.partnershipfortomorrow.ca/>



Appendix B – Increasing the size of the minimum identifiable cohort

Throughout this paper, the protection of data has inherently relied on limiting the minimum size of an identifiable cohort. Data is made more safe by either forcing an increase in the size of the minimum identifiable cohort or by breaking the context and personal features in the dataset. A question arises as to how this may be done while maintaining meaning for the feature. The number of discrete continuous variables may be continuously increased to create classes or ‘bins’ to reduce the number of values for a feature while retaining meaning for the feature, but a greater challenge exists for categorical variables (such as haircolour_is_brown, haircolour_is_grey, or suburb name).

INCREASING THE MICS BY GROWING CLASS RANGE FOR CONTINUOUS VARIABLES

There are many ways to generalise values for data with continuous values. Each approach leads to distortion of the underlying data values. The challenge is to achieve the MICS with the least possible distortion.

The example in Figure 32 shows two sample datasets of integers (age in days). The upper dataset is uniformly distributed over the range of 1 to 10. This creates 10 cohorts of size 1. The bottom dataset has values over the same range (1 to 10) but is not uniformly distributed, creating five cohorts where the MICS is 1.

The example below simply takes the approach of broadening the centre values. As the range of values broadens, centred around integers (for example, 8 is mapped to 7.5 ± 0.5 , then 8 ± 1 , then 8 ± 2), the number of cohorts change and the MICS changes. If the goal is to achieve a MICS of at least 3, the tables show how this simple, uniform growth of numerical range ultimately achieves a MICS of 5 after three steps in the upper set and a MICS of 10 after four steps in the lower set.

Age (days)											Cohorts	MICS	# at MICS
	1	2	3	4	5	6	7	8	9	10	10	1	10
Set 1	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	5.5+/-0.5	5.5+/-0.5	7.5+/-0.5	7.5+/-0.5	9.5+/-0.5	9.5+/-0.5	5	2	5
Set 2	2+/-1	2+/-1	2+/-1	5+/-1	5+/-1	5+/-1	8+/-1	8+/-1	8+/-1	11+/-1	4	1	1
Set 3	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	8+/-2	8+/-2	8+/-2	8+/-2	8+/-2	2	5	2

Age (days)											Cohorts	MICS	# at MICS
	1	1	1	1	3	3	4	4	9	10	5	1	2
Set 1	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	9.5+/-0.5	9.5+/-0.5	3	2	1
Set 2	2+/-1	2+/-1	2+/-1	2+/-1	2+/-1	2+/-1	5+/-1	5+/-1	8+/-1	11+/-1	4	1	2
Set 3	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	8+/-2	8+/-2	2	2	1
Set 4	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	1	10	1

Figure 32. Example of growing class range

With a small change to this approach, fewer steps (and so less overall generalisation) need be applied to achieve the target MICS. In the example above, the lower set goes through a stage where a MICS of 1 is created (after step 2). The approach eventually delivers the desired MICS of at least 3, but goes well beyond (MICS of 10), which significantly reduces the granularity of the data.

If a single cohort is found below the target MICS, that cohort's value can be simply mapped to the next nearest value range, growing that cohort. Figure 33 shows an example of this cohort-merging approach. The upper dataset develops a cohort of size 1 after the second step. This is merged with the cohort of 8+/-1 and the process ends. The lower dataset creates a cohort of size 2 after the first step. This cohort is merged with the cohort of 3.5+/-0.5 and the process ends. If the target MICS has not been achieved, the process can continue with further cohort merging.

Age (days)	1		2		3		4		5		6		7		8		9		10		Cohorts	MICS	# at MICS
	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	5.5+/-0.5	5.5+/-0.5	7.5+/-0.5	7.5+/-0.5	9.5+/-0.5	9.5+/-0.5	10.5+/-0.5	10.5+/-0.5	4	2	2	4	1	1	10	1	10		
Set 1	2+/-1	2+/-1	2+/-1	2+/-1	5+/-1	5+/-1	5+/-1	5+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	4	1	1	
Set 2	2+/-1	2+/-1	2+/-1	2+/-1	5+/-1	5+/-1	5+/-1	5+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	4	1	1	
Set 3	2+/-1	2+/-1	2+/-1	2+/-1	5+/-1	5+/-1	5+/-1	5+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	3	3	2	

Age (days)	1				3		4		4		9		10		Cohorts	MICS	# at MICS
	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	9.5+/-0.5	9.5+/-0.5	9.5+/-0.5	9.5+/-0.5	5	1	2
Set 1	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	9.5+/-0.5	9.5+/-0.5	9.5+/-0.5	9.5+/-0.5	3	2	1
Set 2	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	2	4	1

Figure 33. Cohort merging example of growing class range to force larger MICS

INCREASING THE MICS BY GROWING CLASS RANGE FOR CATEGORICAL VARIABLES

This data to be generalised relates to categorical variables, the broadened values may no longer have real-world meaning (such as hair_colour_is_red_black). Cohort-merging approaches may still be applied but in this case, rather than merging into the next nearest value, the smallest cohort may be randomly split and merged with multiple other cohorts, or randomly joined as a cohort to another cohort.

SPLITTING CONTEXT TO REDUCE THE PERSONAL INFORMATION FACTOR

Chapter 3 identified a mechanism for separately increasing the size of the MICS for features based on location (space), time and relationships separately from those of personal information features.

The example in Figure 34 shows two sample data sets of 2 features: Age in Days and Type. Once again, the upper dataset is uniformly distributed over the range of 1 to 10 with Type values from A to J. This creates 10 cohorts of size 1. The bottom dataset has values over the same range (1 to 10) but is not uniformly distributed, creating 5 cohorts where the MICS is 1.

In the upper set, the Age in Days feature is generalised in the same way as Figure 32, without changing the Type feature. As the Age in Days generalises, the MICS for this one feature increases; however the MICS formed by the two features does not change (MICS remains 1). Overall, the PIF reduces. The lower data set shows the same process for the non-uniform set described in Figure 32.

Age [days]	1	2	3	4	5	6	7	8	9	10	Cohorts	MICS	# at MICS
Type	A	B	C	D	E	F	G	H	I	J			
Set 1	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	5.5+/-0.5	5.5+/-0.5	7.5+/-0.5	7.5+/-0.5	9.5+/-0.5	9.5+/-0.5	10	1	10
Type	A	B	C	D	E	F	G	H	I	J			
Set 2	2+/-1	2+/-1	2+/-1	5+/-1	5+/-1	5+/-1	8+/-1	8+/-1	8+/-1	11+/-1	10	1	10
Type	A	B	C	D	E	F	G	H	I	J			
Set 3	3+/-1	3+/-1	3+/-1	3+/-1	3+/-1	8+/-1	8+/-1	8+/-1	8+/-1	8+/-1	10	1	10
Type	A	B	C	D	E	F	G	H	I	J			

Age [days]	1	1	1	1	3	3	4	4	9	10	Cohorts	MICS	# at MICS
Type	A	A	A	A	C	C	D	D	I	J			
Set 1	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	1.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	3.5+/-0.5	9.5+/-0.5	9.5+/-0.5	5	1	2
Type	A	A	A	A	C	C	D	D	I	J			
Set 2	2+/-1	2+/-1	2+/-1	2+/-1	2+/-1	2+/-1	5+/-1	5+/-1	8+/-1	11+/-1	5	1	2
Type	A	A	A	A	C	C	D	D	I	J			
Set 3	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	3+/-2	8+/-2	8+/-2	5	1	2
Type	A	A	A	A	C	C	D	D	I	J			
Set 3	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5+/-5	5	1	2
Type	A	A	A	A	C	C	D	D	I	J			

Figure 34. Example of growing class range in a two feature set to reduce PIF

In the example data sets in Figure 34, it is worth remembering that the DSF is dependent on the Feature Depth and the Coverage Probability. For a single feature data set (Age in Days), the addition of the second feature to the dataset increases the Feature Depth from 1 to 2 decreasing the DSF for the overall set.

REDUCING PERSONAL INFORMATION FACTOR BY GENERALISING SPATIAL FEATURES

Spatial features are often cited as a challenge when seeking to prevent re-identification in particular when a small number of individuals are located in regional areas.

The example in Figure 35 shows a sample data set of 2 features: location along two spatial axes. The dataset is uniformly distributed over the range of 1 to 10 for both spatial axes. This creates 12 cohorts of size 1. The bottom dataset has values over the same range (1 to 10) but each of the spatial dimensions has been generalised, creating 6 cohorts where the MICS is 2. In this case, the generalisation is not uniformly growing each of the centre values. Rather the generalisations are selected to ensure the cohort size is achieved without the ability to infer the location of the original data item as the centre of a generalised region.

	1	2	3	4	5	6	7	8	9	10	12	MICS	# at MICS
1													
2													
3		{2,3}			{5,3}			{8,3}					
4													
5			{3,5}			{6,5}			{9,5}				
6													
7													
8		{2,8}			{5,8}			{8,8}					
9													
10			{3,10}		{6,10}			{9,10}					

	1	2	3	4	5	6	7	8	9	10	6	MICS	#at MICS
1													
2													
3		{2+/-1,3+/-1}			{5+/-1,3+/-1}			{8+/-1,3+/-1}					
4													
5		{3+0/-2,5+/-1}			{6+0/-2,5+/-1}			{9+0/-2,5+/-1}					
6													
7													
8		{2+/-1,8+/-1}			{5+/-1,8+/-1}			{8+/-1,8+/-1}					
9													
10		{3+0/-2,10+/-1}			{6+0/-2,10+/-1}			{9+0/-2,10+/-1}					

Figure 35. Example of generalising spatial features to reduce PIF

Thanks

This paper was the culmination of more than two years' work by a taskforce which included ACS, the NSW Data Analytics Centre, Standards Australia, the office of the NSW Privacy Commissioner, the NSW Information Commissioner, the federal government's Digital Transformation Agency, CSIRO, Data61, the Department of Prime Minister and Cabinet, the Australian Institute of Health and Welfare, SA NT DataLink, the Government of South Australia, the Victorian Government, the Western Australian Government, the Queensland Government, the Communications Alliance, the Internet of Things Alliance Australia, Data Synergies, Creator Tech, Objective, EY, Microsoft, Clayton Utz and several other companies.

The data sharing taskforce has been run as a series of workshops and occasional intermediate conversations. Workshop participants provide their time freely to help address the challenges within the scope of the taskforce. Participants are free to join or not join each workshop. Special thanks go to some of the more diligent, enthusiastic contributors to these workshops:

Stephen Hardy, Peter Leonard, Chris Radbone, Geof Heydon, Sonya Sherman, Mathew Baldwin, Geoff Neideck, Frank Zeichner, Lyria Bennett Moses, Malcolm Crompton, Geoff Clarke, Kate Cummings, Ghislaine Entwistle, Ghazi Ahamat, Ben Hogan, Scott Nelson, Adrian Watson, Rachael Fraher, Alex Harrington, Andy West, Angelica Paul, Ashton Mills, Ben Hogan, Brian Thorne, Bridget Browne, Cassandra Gligora, Chris Mendes, Daniel Marlay, Dominic Guinane, Ghazi Ahamat, 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, Mukesh Mohania and Tiffany Roos.

Special thanks to Jessica Kashro and Marc Portlock for their organising and coordinating expertise.

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

ABOUT THE ACS

ACS is the professional association for Australia's Information and Communication Technology (ICT) sector. More than 40,000 ACS members work in business, education, government and the community.

ACS has a vision for Australia to be a world leader in technology talent, fostering innovation and creating new forms of value. We are firmly vested in the innovative creation and adoption of best of breed technology in Australia, and we strive to create the environment and provide the opportunities for members and partners to succeed.

ACS works to ensure ICT professionals are recognised as drivers of innovation in our society, relevant across all sectors, and to promote the formulation of effective policies on ICT and related matters.

Visit www.acs.org.au for more information.



**ACS**

Level 27, Tower One
100 Barangaroo Ave
Sydney NSW 200

P: 02 9299 3666
F: 02 9299 3997
E: info@acs.org.au
W: www.acs.org.au