Journal of Privacy and Confidentiality
Vol. 13 (2) 2023

NAHDAP-ICPSR

Submitted Nov 2022
Published Nov 2023

THE PRESENT AND FUTURE OF THE FIVE SAFES FRAMEWORK

ELIZABETH GREEN[†] AND FELIX RITCHIE[†]

[†] University of the West of England, Bristol

Abstract. The Five Safes has become the default framework for confidential data governance across multiple sectors and countries. Since its inception in 2003, the approach has influenced data management in many ways, particularly in the public sector. As it has become established and widely used, both its advantages and limitations have come to the fore, along with an understanding of modern data management principles.

This paper explores the history, application, strengths and limitations of the Five Safes. It discusses the different variations on the framework over time, as well as recent suggestions for deepening or extending the framework.

Finally, we discuss the framework's relationship to the emerging preference for principles-based regulation and design, showing how there is a concordance between the two that may lead to a new consensus on good data governance design models.

1. Introduction

At the beginning of 2003, the board of the UK Office for National Statistics (ONS) was presented with a proposal to set up a virtual research data centre (RDC) to allow academic researchers to analyse sensitive microdata under controlled conditions. Accompanying the proposal was a framework that divided confidential data management into four dimensions: *safe projects* (the ethics, purpose and plan for access), *safe people* (who are the users, and what training do they have/need?), *safe settings* (how are the data stored and transferred, and with what safeguards?) and *safe output* (could published results inadvertently breach the confidentiality of data subjects?). This was promoted by ONS and adopted by academic groups in the UK and US from 2005 onwards [Rahman et al., 2007, Lane et al., 2009]. By 2008 *safe data* had been added (what is the appropriate level of data detail? [Ritchie, 2008]) to allow the framework to describe the whole range of ONS' statistical outputs, including open data. The framework was initially known as the "ONS security model" or "VML security model" after ONS' secure research facility. The name became "Five Safes" in the early 2010s, following a suggestion from a New Zealand researcher.

In the last decade, the Five Safes framework has been widely adopted by organisations across the world, particularly by National Statistical Institutes (NSIs) and academic data

Key words and phrases: Five Safes, data governance, data management, confidentiality, privacy, principles-based.

Corresponding author: E. Green, elizabeth7.green@uwe.ac.uk.

www.journalprivacyconfidentiality.org
DOI:10.29012/jpc.831

Table 1: References to the Five Safes in the academic literature. Source: Google Scholar.

Year	2014	2015	2016	2017	2018	2019	2020	2021	2022
Citations	2	4	13	22	27	48	59	67	96

archives. It has reached countries as diverse as Australia, Mexico, Norway and New Zealand. The Five Safes framework has become a perceived international standard for data governance amongst both government and academic communities. Growth in the Five Safes has gone hand-in-hand with a growing literature; see Table 1. The particularly rapid increase since 2020 reflects the importance of data governance when dealing with COVID research.

As the literature grows and the applications of the Five Safes increase, its pros and cons become clearer. The advantage of the framework is its generality. It can be applied to any data management problem, such as research project planning, setting up secure onsite facilities, or releasing information on the Internet. For example, although the main use has been to develop research data governance frameworks, we are aware of cases where it has been used for human resources systems and staff surveys. It has even been used to frame economic evaluations [Whittard et al., 2022].

This advantage is also its main disadvantage: it is a framework for thinking, but does not explicitly state a solution [Ritchie, 2017a]. There is confusion over scope: is it a structure for data governance, or a checklist for practical data management? Can this checklist be quantified? Can each of the safes be objectively measured and tracked? Should the Five Safes should be equally weighted or are some elements more important [Ritchie and Tava, 2020, Groos and van Veen, 2020]? Are more than five dimensions needed [Oppermann (ed.), 2018, Groos and van Veen, 2020, UK Health Data Research Alliance, 2020, McEachern, 2021]? As the use of the Five Safes has grown and become embedded in public policy, these questions have become more important.

There is a wider context for these questions. The last twenty years have borne witness to many developments in confidential data management. Some are technical, such as differential privacy, synthetic data or table builders. Others are conceptual, such as the attitudinal model encapsulated in the "EDRU" approach (Evidence-based, Default-open, Risk-managed, User-centred), first formalised in Australian Government Department of Social Services [2016].

However, perhaps the most important change has been the increase in "principles-based" data governance, management and regulation. A principles-based approach focuses on outcomes rather than the specifics of implementation: regulated accreditation is the mechanism which allows acceptable use to advance alongside digital advancements. As a result, there is less need to revise statutes or regulations in the light of changing circumstances, as implementations can evolve without breaching core principles [Ritchie, 2014]. This suits the mutable context of the digital world, and is already reflected in legislation: for example, the European General Data Protection Regulation 2016 (GDPR) or the UK Digital Economy Act 2017 (DEA).

Whilst the Five Safes and principles-based regulation are different concepts, they are natural bedfellows: the Five Safes framework provides a basis to help derive principles and manage accreditation procedures. This has been taken furthest in Australia: the Data Access and Transparency Act 2022 requires federal government data sharing to take a principles-based approach following "Five Data Sharing Principles." The legislative framework is

accompanied by an extensive programme of government-developed support materials and accreditation schemes, as well as academic-led initiatives such as the CADRE programme (Coordinated Access to Data Researchers and Environments.[1]

Despite practical developments, there have been no formal attempts to integrate the two concepts. The aim of this paper is therefore twofold: to review the history, usage, current criticisms and likely developments of the Five Safes; and to explore how the Five Safes can be integrated with principles-based design to provide an effective design strategy for data governance and management

The structure of the paper is as follows; Section 2 provides a review of the concept, development and use of the Five Safes. Section 3 considers the future of the Five Safes:will new developments enhance or disrupt the framework? Section 4 considers the principles-based approach in detail, and its relation to the Five Safes. Section 5 concludes.

Whilst some authors (for example, Arbuckle and El Emam [2020]) have applied the Fives Safes to private sector contracting and data management, we focus here on the public sector. We also restrict attention to research uses such as producing statistical analyses, because operational use (e.g., designing HR systems or AI modelling) presents different challenges.

Before proceeding, a note on terminology. For clarity, the term "data access" is used to cover the whole range of ways data are used, and "the access system" refers to the implementation. Examples of access systems are:

- Allowing one's own analysts or third-party researchers to produce statistical results from confidential data;
- Generating and releasing an anonymised data file for unrestricted use;
- Creating files to be shared under licence;
- Government departments sharing administrative data for policy analysis;
- A private company sharing customer data with suppliers for operations or marketing.

2. The Five Safes

Data access is a complex, multidimensional process involving input from lawyers, ethicists, IT designers, statisticians, data holders, contract writers, human resources teams, and others. Griffiths et al. [2021] identify seven data governance frameworks just in the subset of those relevant to indigenous peoples' data. Authors have suggested multiple ways of categorising data to simplify access (for example, Marcotte et al. [2020]).

2.1. **The Concept.**

The Five Safes framework has probably gained most international traction. It was developed as a framework to outline confidentiality management through five dimensions: safe projects, safe people, safe settings, safe data and safe output. All of these contribute to the goal of "safe use." Often these are phrased as "key" questions: Table 2 gives examples.

The Five Safes framework addresses the multi-dimensionality problem by considering each "safe" as a separate but joint element of any solution [Ritchie, 2017a]. For each "safe," there is a wealth of knowledge about how to achieve the goals (how to run ethics committees, how to anonymise data, how to design secure IT systems, and so on). Therefore, safes can

[1] https://cadre5safes.org.au/).

Table 2: The Five Safes, represented as useful questions

Element	Typical question	Example of problems being addressed
Safe projects	Is this appropriate use and management of the data?	• What is the purpose of the access request? • Is this an ethical and lawful use of the data? • What is the benefit to society or to the organisations sharing data? • Is there a data management plan in place? • What happens to the data at the end of the project?
Safe people	How much can I trust the users to use the data appropriately?	• Do the users have the necessary technical skills? • Do the users need training in handling confidential data? • Are users likely to follow procedures?
Safe settings	How much protection does the physical environment afford to the data?	• How are data stored? • Are there physical restrictions on the users? • Does the IT prevent unauthorised use? • Are mistakes by authorised users likely to be detected?
Safe outputs	How much risk is there in the outputs of the access breaching confidentiality?	• If the aim of access is to produce statistics, is there any residual risk by, for example, showing outliers? • If the aim of the access is to produce data for onward transmission, how do we make sure that the released data are appropriate for the next use?
Safe data	Is the level of detail in the data appropriate?	• Is there sufficient detail to allow the project to go ahead? • Is there excessive data are not necessary for the project?

be considered separately, by assuming that the necessary controls in the other dimensions can be implemented as needed.

For example, a project lead considering the broad shape of the project (ethics, data collection methods, user groups, post-project use, and so on) can focus on these elements alone. The project lead can work on the assumption that user training, IT, data detail and output checks can be implemented to any relevant standard, once the shape of the project is clear. Similarly, the designer of the technical systems can make the assumption that users can be trained to work effectively with the system. Alternatively, the project lead might start from considering the types of users, what they can be reliably trained in, and what can't be trained out (the human fondness for short cuts, for example), and feeding this into the specification for system design knowing that it can be achieved.

The separateness makes for *efficiency* in decision-making: not everything has to be considered at once. The *security* comes from then considering all the "safe" dimensions together: for example, do our assumptions about the staff training hold, given the IT system that we have designed?

This joint-but-several approach makes explicit that the overall objective of "safe use" might have multiple solutions. For example, Bleninger et al. [2011] highlighted risks in the outputs from remote job systems, and suggest putting restrictions on outputs. However, Ritchie [2019] argues that the output is not the problem, but the researchers are; applying controls in the "safe person" dimension might have better outcomes. Meanwhile, Bailie [2020] demonstrates how differential privacy (DP) can "solve" the output problem; but he also notes that, in the context of the Five Safes, the most efficient solution is to choose the parameters for a DP solution in the context of the other safes.

These examples illustrate that the effectiveness of control efforts needed to be assessed against the risk being managed. In general, the Five Safes framework is focused on risk-reduction: what risk factors have been identified, and will such an intervention/control meaningfully reduce that risk? However, the framework can also help to identify the need for administrative processes (who controls access rights to the database?) and the coherence, or not, of those. This indirectly can reduce risk, as poorly designed procedures are themselves a risk factor.

These five dimensions of control are scales, not targets; different solutions will have different levels of control in different dimensions. Treating controls as scales demonstrate one of the most useful aspects of the framework—the ability to dial up or dial down controls, based on context and how the safes are considered jointly. Each control should be effective and contribute to reducing risk, therefore limiting unnecessary risk control efforts.

There may be non-existent controls in some dimensions, so long as the collective solution is appropriate. For example, Table 3 shows the controls that could be applied to different versions of the same dataset: one version released openly on the Internet, one downloadable as an end-user licence or "scientific use file," and one available through a secure research facility such as those run by statistical or health agencies.

For the open data, all the protection is contained in the data themselves, as there are no controls over use. In contrast, the secure facility needs only make minimal changes to the data (for example, removing direct identifiers such as names) because other controls provide an assurance of safe use. Release of the licensed download comes in between: the data holder has some confidence in the integrity and competence of users, but is aware that there is much scope for uncontrolled error, and so reduces the inherent risk in the data.

There is no inherent precedence among the Five Safes. Groos and van Veen [2020], Lane [2020], Ritchie and Tava [2020], and Boniface et al. [2022] argue for the primacy of safe projects: the other safes should only be considered once a clear understanding of the

Table 3: Differing control levels potentially applied to the same dataset.

Situation	Controls				
	Project	People	Settings	Outputs	Data
Open data	None	None	None	None	Very high (full anonymization)
Licensed download	Some (online application)	Some (licence)	Some (online guidance)	None	High (e.g., little geographical detail)
Controlled use	High (application with human review)	High (compulsory training)	High (isolated environment)	High (all outputs manually reviewed)	Minimal (de-identification only)

project rationale and outcome has been established. Ritchie [2017a] proposes that "safe data" should be treated as the residual: when you know who needs the data, for what purpose, and how they will access it, you can adjust the level of the detail in the data to the access arrangements and user need. However, Ritchie [2017a] acknowledges this is a suggestion, not a rule. There may be cases where, for example, the detail in the data is fixed in advance, and the other dimensions must adjust to it. This could happen in a research project where the data collection plans have not been integrated with planning the research phase.

The term "safe" itself has proved problematic. In the early 2010s, organisations began interpreting the framework to mean that a "safe" solution must be 'safe' in every dimension, in the ordinary sense of the word "safe:" a setting which is not "safe" must be 'unsafe" and therefore a risk. Desai et al. [2016] tried to correct these errors with the first detailed description of how to use the Five Safes. Although clearly a draft, this remains the most widely cited description of the Five Safes.

Desai et al. [2016] represents the Five Safes accurately, despite its draft nature, but this has not wholly eliminated misinterpretations. For example, Brennan et al. [2019] set up the Five Safes as deterministic and inflexible, yet also indeterminate; they use this to argue that the Five Safes framework is inappropriate, despite being unclear whether it is the determinism or indeterminism that they object to. Similarly, Culnane et al. [2020] interpret "safe" as an absolute standard, argue that is it impossible to define absolute standards, and therefore dispose of the straw man created. Tam [2021] carries out extensive statistical analysis on the assumption that the dimensions are independent of each other, which is one of the few things the Five Safes are very definitely not. Wirth et al. [2021] simply do not question the meaning of "safe."

Finally, we must recognise the use of the framework itself is no guarantee of good practice. A governance plan of "There's no need for ethics; we'll only give the data to people we trust, who say they'll look after them safely; we don't need to reduce the data detail, as we couldn't find anyone we know in it, so there's no risk in outputs..." is using the Five Safes, but not necessarily to good effect. This has led to some (such as Brennan et al. [2019]) to criticise the Five Safes framework as a "box ticking exercise," which it certainly can be. For example, one UK government organisation declares itself to be "Five Safes compliant;"

this is meaningless but presumably reassures a decision-maker somewhere. It should be clear though that the problem is not the framework, but the implementation.

2.2. **Use.** In its early days, the Five Safes framework was applied to existing systems as a way of describing them (for example, Ritchie [2008] and Corti et al. [2020] focused on secure research data centres—RDCs). Over time, application has moved away from RDCs to encompass all aspects of data governance. Nowadays, the growing familiarity with the model provides a common frame of reference even for organisations that do not formally use it in their internal models [Bujnowska, 2018]. This is particularly observable when considering interoperability of standards between organisations [Organization for Economic Cooperation and Development, 2014, Ritchie, 2013]. Conferences such as the biannual Work Session on Statistical Data Confidentiality increasingly see the Five Safes used as a framing device for presentations; the 2019 meeting summary explicitly suggested that more formal use should be made of the framework [United Nations Economic Commission for Europe, 2019].

The Five Safes framework has three main uses: description, including pedagogy; design and evaluation; and regulation, which we discuss next.

2.2.1. *Description, Framing and Pedagogy.* The Five Safes framework is currently used to describe the governance arrangements for all the general-purpose UK government and academic RDCs for health and social sciences (ONS, Her Majesty's Revenue and Customs, Northern Ireland Statistics and Research Agency, the Scottish Safe Havens, the SAIL [Secure Anonymised Information Linkage] Databank in Wales, the National Health Service Direct secure facility, and the UK Data Archive). Eurostat [Bujnowska, 2018], the German central bank [Bender et al., 2022], the Dutch research infrastructure ODISSEI, and the NSIs of Canada, Australia, New Zealand, Mexico, and Norway formally describe their RDCs in this way. The French secure data centre CASD retrofitted the Five Safes to existing governance arrangements to improve international comparability [Silberman, 2021]. As NSIs often have a strong influence over the data strategies of other parts of the public sector, the adoption of the Five Safes by NSIs has had significant spillover effects, particularly in Anglo-Saxon countries. However the UK Office for Statistics Regulation found that organisational commitment to the concept can vary when use of it is imposed by the public body on others [Office for Statistics Regulation, 2019].

Karrar et al. [2021] use the Five Safes to structure responses to a survey on data use registers. Atkin et al. [2021] appear to have used it in patient engagement groups to discuss multi-control approaches to safe data use; Coulter [2021] also notes its usefulness in this context. Griffith University Library used it as a "framework for all researchers working with potentially confidential data" [Weaver and Richardson, 2021].

It has also been used in confidential data management training since 2004 [Ritchie, 2008, Eurostat, 2016, Green et al., 2017, UK Office for National Statistics, 2020, Wiltshire, 2021]. The ready-made structure appeals to the trainees, and the framework provide a context for the training itself as part of the "safe people" element. In data governance training developed by the authors for the National Institute for Health and Care Research, homework exercises are structured around the Five Safes to help researchers build data governance plans.

There is a question of terminology. Describing these as five "elements" tends to work well for designers and implementers, as it implies a self-contained component which can be

tackled as part of a project plan. However, in discussions with data holders and the general public, who tend to be more concerned that risk is being managed, "dimensions of control" seems to be a more meaningful description; this carries the sense of "what you can't control, and what you can".

Other writers have described the elements as themes, security controls, components, or standards. The Mexican statistical office [Volkow, 2019] uses five "secure elements," as does the OECD [Organization for Economic Cooperation and Development, 2014]. Some organisations (for example, the NORC Data Enclave in the US, or the German federal statistical office) describe their operations in terms of a "portfolio approach." Lane [2020] suggests that "safe exports" is a better phrase than "safe outputs."

2.2.2. *Design and Evaluation.* In the last ten years awareness of the Five Safes has preceded planning, and so it has become more common to use the framework for designing data strategies. (Some public examples are Organization for Economic Cooperation and Development [2014], Australian Government Department of Social Services [2016], ICON [2016], Office for Statistics Regulation [2018a], and Cranswick et al. [2019].) Private sector organisations advising on data strategies have also begun to use the five safes (for example, Security Brief Australia [2019], Arbuckle and El Emam [2020], and Hafner et al. [2019] used it to minimise utility loss in scientific-use files.

The Five Safes framework is frequently used to build more complex, or more detailed, models. Sikorska et al. [2020] develop a risk assessment tool for company data. Raisaro et al. [2020] and Wirth et al. [2021] propose new medical infrastructures. Boniface et al. [2021] propose a "social data foundation model," Oppermann (ed.) [2018] a "human services outcomes" framework, and Bender et al. [2022] a model for maximising value from data; all take the Five Safes as the "base" model. Zheng et al. [2020] suggest that the Five Safes help to operationalise "privacy by design" in security automation.

Two groups have proposed the systematic adoption of the Five Safes for design. The National Research Council [National Research Council, 2014] incorporate it into recommendations for anonymization of US health data for sharing. The Australian Productivity Commission's comprehensive report [Australian Productivity Commission, 2017] led directly to the Data Access and Transparency Act.

The lack of detail in the Five Safes is a limitation, and several authors have tried to create practical guidelines [Arbuckle and El Emam, 2020, Boniface et al., 2022]). In 2018, the UK Anonymisation Network published the "Anonymisation Decision-Making Framework," a re-badged version of the Five Safes intended as a more practical guide [Elliot et al., 2020]. In 2021, the Australian research group CADRE (Coordinated Access to Data Researchers and Environments) was set up to provide detailed guidance on all elements of the Five Safes; the initial "CADRE Five Safes Framework" [McEachern, 2021] is now available as a web resource. CADRE notes that Five Safes framwork does not stipulate a quantitative evaluation of risk factors, and so it develops a checklist to provide consistency between data managers.

Finally, the predefined structure can simplify evaluation. This works particularly well in cases where the system being evaluated was designed using the Five Safes; for example, UK Office for National Statistics [2011] is a risk assessment of the UK Data Archive secure facility, which used the Five Safes as its design framework. But this is not necessary: ICON [2016] uses the structure to evaluate the Hellenic Statistical Agency's data strategy, and

Whittard et al. [2022] use the Five Safes to structure an economic evaluation of Bill and Melinda Gates Foundation investments into data governance in Ethiopia.

2.2.3. *Regulation.* The Five Safes framework is used in formal legislation, such as the UK Digital Economy Act 2017 and the state legislation in New South Wales, Victoria, and South Australia (DPMC, 2019). With the growth in principles-based regulation (see below) the Five Safes has become a useful way to frame legislation. Two examples illustrate this.

Microdata access at the UK Office for National Statistics: In 2002, when the Five Safes framework was conceived, access to the business data at ONS was governed by the Statistics of Trade Act 1947. This act had noconception of data being collected for any purpose other than producing aggregate tabulations, and so required extensive legal negotiations to ensure the lawfulness of access.

In 2007 the Statistics and Registration Services Act created a simple access gateway, the "approved researcher," and the basis for a flexible data acquisition and access path. These provisions reflected the growing awareness, embodied in the Five Safes, that data security is multi-dimensional, leading to a culture change about how ONS should approach data access issues.

By the time of the Digital Economy Act 2017 (DEA), the pieces had fallen into place, and the legislation served to embody this new model into law. In respect of research data access, the DEA set out the principles on which access would be granted, and created an accreditation process for each of projects, people, settings and outputs (safe data is effectively the residual, as Ritchie [2017a] recommends). The UK Statistics Authority oversees all the accreditation processes, and the Office for Statistics Regulation bases its guidance on the Five Safes [Office for Statistics Regulation, 2018b] Key academic funders (for example, the Administrative Data Research UK, Health Data Research UK and the Department for Education) have followed suit.

Data access in the Australian Federal Government: Until 2014, the Australian federal government took a data-centred default-closed perspective on data sharing and access. Following a critical review by the public auditors, the Australian Bureau of Statistics (ABS) took the decision to move towards a default-open environment for research access, adopting the Five Safes as the governance framework. The Department of Social Services followed suit [Australian Government Department of Social Services, 2016]. The Productivity Commission subsequently carried out a major review of all government data-sharing practices and recommended a wholesale adoption and development of the ABS approach [Australian Productivity Commission, 2017].

The Office of the National Data Commissioner was set up in 2018 to implement the recommendations, and began a two-year programme of public consultation and legislative drafting. A key element of drafting and consultation was developing guidelines to help practitioners implement the principles. Legislation came into force as the Australian Data Access and Transparency Act 2022, with the Five Safes transmuted into the "Five Data Sharing Principles."

The Australian experience therefore represents a very different development trajectory compared to the UK, with a step shift in attitudes and practice. However, none of the individual elements being introduced is unknown in the rest of the world, and the whole-of-government approach provides a unique opportunity to build in consistency and shared understanding of principles from the beginning.

3. What Next for the Five Safes?

We move to the future of the Five Safes.

3.1. **Should There be More Safes, or More Sub-safes?** As noted, the Five Safes framework is a generic. Some authors have tried to give more meaning to the dimensions. For example, Ritchie [2013], when considering how international data-sharing standards might be developed, suggested separating "safe people" into "knowledge" and "incentives," and "safe setting" into "access" and "networks." At least one NSI breaks the "safe people" component into organisational and personal criteria, so that this component is itself multi-dimensional. Health Data Research UK (HDRUK) suggests that "safe computing" be an explicit sub-field of safe settings (HRDUK, 2020).

The element that has grown most in scope since 2003 is "safe projects." Initially there to cover the process of getting approval to ONS' secure facility, it now is seen to cover all the project planning aspects: ethical approval, data management plans, benefit/cost assessment, user identification, post-project disposal. As Ritchie and Tava [2020] note, when designing data management from scratch this element takes centre stage. This has also been the focus of much of the Australian consultation preceding the new legislation, particularly the onward use of data from data access requests. Boniface et al. [2022] suggest that "safe collaboration" better reflects the growing need for inter-organisational co-operation.

"Safe outputs" has also changed. Originally concerned with the production of statistical outputs by researchers, different applications of the Five Safes means that this has described user queries to an HR system, datasets for onward management, machine learning models, or products to support compliance strategies.

There have been some proposals to increase the number of dimensions. The Australian Computer Society (ACS) [Oppermann (ed.), 2019] suggests adding:

- Safe organisations—ensuring the organisation has in place, and follows, agreed procedures;
- Safe lifecycle—time-relevant issues such as archiving and the sensitivity over time;
- Safe outcomes—the ultimate uses of the project output;
- Safe use—the impact of using the project output, apparently in a moral/ethical context
- Safe response—dealing with accidents

UK Health Data Research Alliance [2020] proposes "safe return" for clinical settings. This allows the return of analytical results generated in the safe setting to be "returned" to the original data holders for re-identification and thus clinical care. Meanwhile, Groos and van Veen [2020] suggest a "safe law" to ensure that law is always complied with.

While these are all important subjects, it is not at all clear that the extra "safes" help data planning. The safe projects/outputs/outcomes/use overlap underlies the difficulties: as more dimensions are added, the distinctions between the dimensions become weaker, and it is not clear what question is being addressed and how each dimension can be considered separately. Why is "safe law" not already covered in safe projects, for example? Groos and van Veen [2020] suggest that it needs to be separate and independent from the other safes as it must meet an absolute standard (compliance with law). However, it is surprising that the authors do not recognise that modern data protection laws are explicitly multi-dimensional and recognise the impossibility of prescribing absolute standards; this aligns directly with the Five Safes model of combining controls to achieve safe solutions.

Adding extra safes can also create unhelpful distinctions. For example, most organisations consider the safe organisations/safe people argument as part of the same question—what skills, training or assurances are needed to ensure that the users act appropriately? For bodies which explicitly consider organisations as part of the solution (such as Statistics Canada, the Australian Bureau of Statistics, or Eurostat), separating organisation from persons makes no sense. Excessive subdivision may also lead to micro-management and turf wars, outcomes the Five Safes was originally designed to avoid.

This is not to say that raising these issues is unhelpful in terms of thinking about data access. For example, "safe response" emphasises that a well-run organisations would have processes in place to identify, manage and communicate breaches of data security. It is not clear why it is helpful to think of this separately from the other dimensions. People and IT planning normally includes a breach policy, and part of a cost-benefit or privacy impact assessments should include the risk of failure and mitigation. However, it could be argued that this is necessary to make organisations explicitly confront the issue. Similarly, "safe returns" is simply allowing for a wider class of output than just non-disclosive statistical findings; but it raises the possibility of new uses for secure research environments. Addressing these issues is one reason why the Australian Government included advice on what to do before *and* after applying the Five Data Sharing Principles in its Guidance.

Perhaps the reason why these struggle to make the case for new "safes" is that the additional measures are about implementation rather than concepts or themes. Claiming that a designed system is "safe" because safe projects, people and setting have been considered assumes that project planning, people training and robust system design are done competently. There is of course nothing in the Five Safes or the extended scheme to guarantee this. In contrast, the extensions to "safe people/organisations" being trialled in some countries does not add new concepts, but refines them.

Nevertheless, the ACS's detailed critique of the Five Safes should be welcomed for challenging assumptions, precisely because they can be accommodated in the current framework. For example, when training users in data governance we encourage users to treat each processing stage as a different problem (data collection, processing, analysis, etc). Perhaps the "safe lifecycle" could provide a mechanism for integrating between stages. Similarly, "safe response" may not be helpful as a separate "safe," but it is clearly something that should be considered as part of data management.

One new problem that the ACS highlights is artificial intelligence (AI) systems. Should a new "safe" be created for them? Perhaps in the future there will be a new dimension for autonomous systems; but the current debate, whether AI fits into the "people," "settings" or "output" category, is stimulating much debate precisely because trying to fit AI into an existing category forces us to confront our understanding. On the other hand, the extensive review of Jefferson et al. [2022] concludes that AI models can be handled within the current framework. (New ethical problems are in safe projects; user and staff training in safe people; technical controls on disclosure in safe outputs.)

Finally, some authors have proposed ways to combine the safes more effectively. McEachern [2021] introduces "safe group," a combinations of different safe categories that would be automatically formulated as "safe" for classes of use. Similarly, Statistics Canada is undergoing a process of mapping user-data combinations that are "allowable" [Thomas, 2020]. Finally, Boniface et al. [2022] are trying to develop a classification system for all the safes; interestingly, they try to identify combinations which generate so-called "red lines:" for example, if releasing data without restriction, full anonymisation is the only acceptable

"safe data" standard. Boniface et al. [2022] also treat safe projects as a red line *per se*: if you can't demonstrate an ethical basis for the data access, all other considerations are irrelevant.

3.2. **Should the Five Safes Be More Objective?** The Five Safes framework is explicitly subjective. It does not attempt to quantify "safe" in any dimension, or objectively balance "safety" in one dimension against '1safety" in another dimension. There are five reasons for this.

First, there is no meaningful metric in any dimension, let alone a common metric. Consider defining a metric for "safe projects." There could be a case for a cost-benefit analysis of the value to the public: lives saved as a result of a data linking project versus implementation cost and expected cost arising from an unknown breach. Most of these costs and benefits are unknowable and not measurable [Alves et al., 2021]. This is why cost-efficiency analysis is more often used to assess the worth of a project, but by design, such analyses do not question whether the project itself brings an acceptable risk or not.

It could be argued that the common measure across all dimensions is "what is the risk of re-identification?" This is often the proposed measure (assuming it could be measured) but is not the only one. Re-identification risk can almost always be substantially reduced by expenditure in one or more dimensions, and so this risk alone is not a measure.

The second issue is that any discussion of "trade-offs," "risk-utility maps" or similar concepts implicitly assumes the independence of protection measures. Assume, for expository purposes, that people and setting are the only relevant dimensions. The trade-off argument assumes these are independent and linked by some function f to create safe use:

$$\text{safe use} = f(\text{people}, \text{setting})$$

However, in reality the risk of re-identification for each depends on the other:

$$\text{safe use} = f\left(\text{people}_{|\text{setting}}, \text{setting}_{|\text{people}}\right).$$

For example, the likelihood of an individual making a mistake might be dependent on the IT system; but the effectiveness of IT controls might depend on training for the user. This is no longer a linear trade-off, but a complex, non-linear model—again, with unknown values in every place.

Third, trade-off models do not allow for discontinuities. A secure RDC presents qualitatively different risks from an end-user licence model; both will be affected differently by ethical approval processes. This topic is characterised by discontinuities across all the different potential activities.

Fourth, any data that do come out of modelling are subjective. Consider the most objective measure in this whole field: assessing the disclosure risk in a given dataset. There is a large literature on this, going back to the 1970s. Techniques have been developed for many types of data, statistics and attack models. There are software tools, such as μArgus/τArgus or sdcTable/sdcMicro, to calculate risk probabilities to many decimal places, which are invaluable for exploring the *relative* risk in data and tabular outputs. Overall, there is large, stable and well-understood body of literature on the relative pros and cons of all the different statistical disclosure control (SDC) procedures, and metrics for comparing SDC risk in methodological settings.

None of these have external validity [Hafner et al., 2015]. The disclosure modeller faces choices at every stage of the process for which a decision is necessary:

- The attack scenario;

- The attacker's motivation;
- How much information the attacker already has;
- The time and resources available to the attacker;
- Alternative data sources available to the attacker;
- Alternative published statistics available to the attacker;
- What similar results will be published in future; and
- Whether repeated attacks are possible.

SDC analysts often discuss "worst case scenarios." The SDC literature has developed this concept to a common standard so that the results from analytical research articles can be usefully compared and lessons learned about the pros and cons of different measures. However, this does not mean that models used for methodological development should be applied in practical cases, as the "worst case" may no longer be relevant [Hafner et al., 2015]. Moreover, these models are typically the *mathematically tractable* worst-case scenarios. They cannot represent events that cannot be modelled (spontaneous recognition, for example [Ritchie, 2017b]) or unknowable events (the likelihood of an authorised user selling market-sensitive information). In short, claims to objectivity are spurious.

Fifth and finally, there is a "Hawthorne effect:" the design and operation of the system affects the psychological perception and hence interaction of the participants in the system. Most obviously, user training and IT systems are designed to affect behaviour, but as Green et al. [2017] show, the way that users are introduced into data governance regimes affects the way that they respond to them. In practice, most systems evolve over time in unexpected ways and develop institutional cultures that may resist change, however well-considered the change seems.

For many years, data access was dominated by the statistical approach, where quantitative risk assessments were given great prominence. This was most noticeable in the statistical agencies; as these had positions of authority in national systems, this was the prevailing viewpoint in the public sector. However, in the last decade this has increasingly been abandoned in favour of the multi-dimensional, explicitly subjective, EDRU approach (Evidence-based, Default-open, Risk-managed, User-centred; Australian Government Department of Social Services [2016]).

The choice of whether to quantify risk also affects the public debate. Numerous studies have shown that public perceptions of whether data sharing is a public benefit are highly sensitive to the questions asked (for example, Hallinan et al. [2012], Wellcome Trust [2013], Understanding Patient Data [2018]). Studies such as Jenkins et al. [2017] also show that numbers are more likely to be perceived as reliable than verbal description, irrespective of their actual credibility. For example, during the COVID-19 pandemic, there was substantial public debate about the value of R_0, the base infection rate [Bates Ramirez, 2020]. Public engagement centred on whether this value was greater or less than one. There was almost no debate outside scientific circles of confidence intervals, sample bias, distributions or other statistical factors that make the simple "What is R_0?" question meaningless.

In this light, the lack of quantitative measures for the Five Safes is not a gap in the literature, but an essential strength of the model. It focuses decision-makers' minds on the need to collect, evaluate and use subjective evidence, and to build consensus for decisions. This is certainly the perspective taken by McEachern [2021], who praises this flexibility. Quantitative models, such as those used by SDC professionals, can provide useful supporting evidence, particularly in terms of relative risk; but only supporting evidence.

4. The Five Safes, Principles-Based Planning and Accreditation

Here we link the Five Safes to principles-based regulation.

4.1. **Rules-Bases Versus Principles-Based Regulation.** Rules-based regulation aims to specify in a binary manner what is allowed or not allowed; the primary source of direction is the regulation itself. The strength of rules-based regulation lies in simplicity and a common understanding of the regulation. It works best when the terms can be unambiguously defined; for example, the UK Data Protection Act 2018 makes attempted re-identification of personal information in de-identified or anonymised data a criminal act, except in very specific, easily understood cases. In general, however, specification of rules in data governance is difficult. Regulating digital activities has severely tested regulations which rely upon clear statements; for example, no legislation provides an unambiguous definition of "anonymous" data because it relies on context except in very trivial cases.

In contrast to rules-based regulation, principles-based regulation focuses on what any system is trying to achieve, and then questions whether the system actually achieves those objectives. In a principles-based system, implementation decisions are primarily under the control of the implementer. Regulation is there to specify the goals, and to identify what evidence should be presented that the goals have been achieved.

A key element of making the principles-based approach work is accreditation. Rather than specifying actions, actors, systems and procedures undergo ex ante validation to ensure that they are fit for the purpose. In financial markets, for example, "good" conduct is hard to specify, and so many regulators rely on the accreditation of practitioners [Keenan, 2020]. The popularity of principles-based regulation is that it seems to address some of the flaws of older legislation, which struggled to provide adequate guidance.

The advantages of the principles-plus-accreditation approach are:

- Efficiency, as solutions can be adjusted to circumstance rather than a legislative context;
- Flexibility, as multiple accreditation pathways can be set up;
- Adaptability to circumstances, rather than requiring legislative/formal change;
- Adaptability to collective learning as processes develop;
- Cultural change through positive reinforcement and engagement with goals;
- Engagement with stakeholders.

The last is a consequence of the problems of the principles-plus-accreditation approach, compared to a rules-based regulation model:

- It may be less clear, at least at first sight;
- There may be concerns about unfair or uneven treatment, and so transparency in processes is essential;
- Practitioners might want clear guidelines in advance to implement solutions, rather than post-hoc approval.

An effective principles-based system, therefore, requires a much greater level of engagement.

4.2. **Principles-based Regulation, Data Management and the Five Safes.** In recent years, the Five Safes model has become increasingly associated with the principles-based approach to regulation. There is an affinity between the two concepts. The Five Safes framework provides for planning; the principles-based model provides a way of suggesting

Table 4: Applying rules-based and principles-based standards to the Five Safes.

Safe	Rules-based standard	Principles-based standard
Projects	Identify list of valid uses	Specify benefits that must be demonstrated, and risks to be considered
People	Require specific accreditation eg meet Civil Service appointments criteria	Require 'appropriate' training
Setting	Follow government IT standard	Follow ISO27001 practices (ie choose system and be able to demonstrate integrity of it)
Output	Apply threshold rule tabular statistics	Apply threshold "rules of thumb" but allow appeals if important and demonstrably non-disclosive
Data	Clear boundary between anonymised and other data	Data must be "appropriate" to the environment

how goals should be specified. Neither is specific on the actual implementation; but both provide a way to measure the effectiveness of any implementation.

In the context of the Five Safes, rules-based and principles-based regulation may be contrasted using the examples in Table 4.

The Five Safes therefore can accommodate either approach, or a mixture of the two, but moves towards principles-based models are more likely to be done in the context of the Five Safes. This is explicit in legislation in the UK and Australia but even the European GDPR follows this line. The GDPR is not explicitly principles-based and does not cite the Five Safes; nevertheless, its balancing of data detail against "procedural and technical measures," and the avoidance of specific technical or statistical standards in favour of solutions "having regard to" outcomes, places it in the same camp.

4.3. **So What?** Conformability between the Five Safes and principles-based regulation naturally occurs as both accept that specific solutions are hard to define and implement at strategic levels. It is better, instead, to identify structures and goals, and then test outcomes against the latter. Once again, it is the lack of specificity in the Five Safes that turns out to be the critical factor: there is no specification of what constitutes a "safe project," but there is a clear requirement of the system designer to ensure that any resulting solution is one such project.

Ritchie [2022] notes that these high-level concepts do not necessarilyhelp system designers, and proposes a nested perspective to support decision-making that is both efficient and robust to changes in design:

- First, consider how the design problem is to be approached (the attitudes towards evidence, risk and so on to be adopted);
- Second, specify the broad principles and aims of the data system;
- Third, use the Five Safes to provide the structure of the solution;
- Fourth, identify best/good practice in each area of the Five Safes;

- Fifth, implement a solution.

Ideally, this top-down framework should increase both public benefit and efficiency: public benefit because it starts with high-level goals and refines consistently with those goals; efficiency because the purpose of any implementation should be clear and directly relatable to the principles. The Five Safes framework is becoming an implementation tool. Starting from first principles is becoming common in law, but still relatively rare in practice; and very few organisations trouble to genuinely challenge their operating ethos. Nevertheless, this may be a useful way to conceptualise and implement data governance decisions.

5. Conclusion

The Five Safes model has become well-established, not through diktat but because it offers a useful structure for addressing data access, management and governance. It is increasingly the design structure for statistical organisations and secure research centres, and has prompted the development of useful practices such as output statistical disclosure control and active researcher management.

It remains an empty structure, by design. The usefulness and applicability of the concept is because it helps to order discussions, design, evaluation, and regulation, but makes no specific requirement on any of these. Using the Five Safes to design a data governance plan is no guarantee of competent delivery, any more than ignoring the Five Safes is a certain recipe for failure.

The Five Safes framework is explicitly subjective and qualitative, again by design. While some recent proposals to develop quantitative measures of risk are no doubt well-intentioned, the lesson of the last twenty years is that spurious metrics limit genuine risk assessment, lower credibility, and reduce the opportunity to extract public value from data sharing and use. Subjective decision-making, properly implemented, provides consistency and transparency but with the capability of responding to individual circumstances. There have also been proposals to extend the range of the Five Safes by creating sub-divisions, or by having more "safes." At present, it is not clear that the gain in coverage outweighs the increase in complexity.

What has become apparent in the last decade, as the number of users of the Five Safes increases, is that there is a need for more practical guidance on implementation. Many organisations have developed their own guidelines or interpretations; perhaps the next stage in the evolution of the Five Safes will be common understanding of the ten, or twenty, or however many "Sub-Safes." The scale of the Australian federal government project is likely to become a benchmark in this regard, and the CADRE project to provide explicit guidance to data managers.

Finally, we note that the Five Safes framework did not develop in a vacuum, and continues to interact with other concepts. The biggest change in the next decade is likely to be the growth in principles-based data regulation, where the Five Safes framework provides a natural structure for the accreditation processes essential for effective principles-based operation.

References

K. Alves, F. Tava, D. Whittard, E. Green, M. Beata Kreft, and F. Ritchie. Process and economic evaluation of the ODI R&D programme: Final report, 2021. Open Data Institute, London.

L. Arbuckle and K. El Emam. *Building an Anonymization Pipeline*. O'Reilly Publishers, Sebastopol, CA, 2020.

C. Atkin, B. Crosby, K. Dunn, G. Price, E. Marston, C. Crawford, M. O'Hara, C. Morgan, M. Levermore, S. Gallier, S. Modhwadia, J. Attwood, S. Perks, A. Denniston, G. Gkoutos, R. Dormer, A. Rosser, A. Ignatowicz, H. Fanning, and E. Sapey. Perceptions of anonymised data use and awareness of the NHS data opt-out amongst patients, carers and healthcare staff. *Research Involvement and Engagement*, 7:40, 2021. doi: https://doi.org/10.1186/s40900-021-00281-2.

Australian Government Department of Social Services. Data access project: Final report, 2016.

Australian Productivity Commission. Data availability and use—inquiry report, 2017.

J. Bailie. Big data, differential privacy and National Statistical Organisations. *Statistical Journal of the IAOS*, 36(3):1–8, 2020. doi: 10.3233/SJI-200685.

V. Bates Ramirez. What is "r-naught"? gauging contagious infections, 2020. URL https://www.healthline.com/health/r-naught-reproduction-number. Healthline.

S. Bender, J. Blaschke, and C. Hirsch. Data production in a digitised age: The need to establish successful workflows for microdata access, 2022. Deutsche Bundesbank, Research Data and Service Centre Technical Report 2022-02.

P. Bleninger, J. Drechsler, and G. Ronning. Remote data access and the risk of disclosure from linear regression. *Statistics and Operational Research Transactions*, 35:7–24, 2011. Special Issue: Privacy in Statistical Databases.

M. Boniface, L. Carmichael, W. Hall, B. Pickering, S. Stalla-Bourdillon, and S. Taylor. The social data foundation model: Facilitating health and social care transformation through datatrust services, 2021. Mimeo, University of Southampton Interdisciplinary Centre for Law, Internet and Culture.

M. Boniface, L. Carmichael, W. Hall, J. McMahon, B. Pickering, M. Surridge, S. Taylor, U.-I. Atmaca, G. Epiphaniou, C. Maple, S. Murakonda, and S. Weller. Privacy risk assessment requirements for safe collaborative research, 2022. DARE UK Sprint 'PRiAM' v1.1, July.

P. Brennan, M. Fitzpatrick, J. Larranaga, V. O'Donnell, M. Osman, C. Petterson, J. Powles, C. Twomey, and R. Wortham. Privacy and responsible information sharing for western australia, 2019. Community Submission to WA consultation on data sharing.

A. Bujnowska. Access to European microdata for statistical purposes, 2018. URL https://ec.europa.eu/eurostat/cros/system/files/04.access_to_microdata.pdf.

L. Corti, V. van den Eyden, L. Bishop, and M. Wollard. *Managing and Sharing Research Data: A Guide to Good Practice, 2nd edition*. Sage, Thousand Oaks, CA, 2020.

A. Coulter. Patient trust in plans to share primary care data. *British Medical Journal*, 373: 1413, 2021. doi: http://dx.doi.org/10.1136/bmj.n1413.

K. Cranswick, S. Tumpane, and S. Stobert. Virtual data labs—a more flexible approach to access Statistics Canada microdata, 2019. UNECE/Eurostat Work Session on Statistical Data Confidentiality, The Hague.

C. Culnane, B. Rubinstein, and D. Watts. Not fit for purpose: A critical analysis of the "five safes", 2020.

T. Desai, F. Ritchie, and R. Welpton. The Five Safes: designing data access for research, 2016. Working papers in Economics 1601, University of the West of England, Bristol.

M. Elliot, E. Mackey, and K. O'Hara. *The Anonymisation Decision-Making Framework: European Practitioners' Guide, 2nd Edition*. UK Anonymisation Network, Manchester, UK, 2020. URL https://msrbcel.files.wordpress.com/2020/11/adf-2nd-edition-1.pdf.

Eurostat. Self-study material for the users of Eurostat microdata sets, 2016.

E. Green, F. Ritchie, J. Newman, and T. Parker. Lessons learned in training "safe users" of confidential data, 2017. UNECE Work Session on Statistical Data Confidentiality 2017.

K. Griffiths, J. Blain, C. Vajdic, and L. Jorm. Indigenous and tribal peoples data governance in health research: A systematic review. *International Journal of Environmental Research and Public Health*, 18:10318, 2021. doi: doi.org/10.3390/ijerph181910318.

D. Groos and E. van Veen. Anonymised data and the rule of law, 2020. European Data Protection Law, 4:498-508.

H. Hafner, R. Lenz, and F. Ritchie. User-focused threat identification for anonymised microdata. *Statistical Journal of the IAOS*, 35(4):703–713, 2019. doi: https://doi.org/10.3233/SJI-190506.

H.-P. Hafner, R. Lenz, F. Ritchie, and R. Welpton. Evidence-based, context-sensitive, user-centred, risk-managed SDC planning: designing data access solutions for scientific use, 2015. UNECE/Eurostat Work Session on Statistical Data Confidentiality, Helsinki.

D. Hallinan, M. Friedewald, and P. McCarthy. Citizens' perceptions of data protection and privacy in Europe. *Computer Law and Security Review*, 28(3):263–272, 2012. doi: https://doi.org/10.1016/j.clsr.2012.03.005.

ICON. Hellenic Statistical Authority Mission on microdata access: final report, 2016. Eurostat.

E. Jefferson, J. Liley, M. Malone, S. Reel, A. Crespi-Boixader, X. Kerasidou, F. Tava, A. McCarthy, R. Preen, A. Blanco-Justicia, E. Mansouri-Benssassi, J. Domingo-Ferrer, J. Beggs, A. Chuter, C. Cole, F. Ritchie, A. Daly, S. Rogers, and J. Smith. Recommendations for disclosure control of trained machine learning (ML) models from Trusted Research Environments (TREs), 2022.

S. Jenkins, A. Harris, and R. Lark. Maintaining credibility when communicating uncertainty: The role of communication format. In G. Gunzelmann, A. Howes, T. Tenbrink, and E. Davelaar, editors, *CogSci 2017: Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, pages 582–587. Cognitive Science Society, London, 2017.

N. Karrar, K. S., S. Manohar, P. Quattroni, D. Seymour, and S. Varma. Analysis of data use registers published by health data custodians in the UK, 2021. medRxiv preprint.

P. Keenan. *Dictum meum pactum*: UK regulation: Rules or principles, 2020. Keenan Regulatory Consulting.

J. Lane. *Democratizing our Data: A Manifesto*. MIT Press, Cambridge, MA, 2020.

J. Lane, C. Bowie, F. Scheuren, and T. Mulcahy. NORC Data Enclave: Providing secure remote access to sensitive microdata, 2009. URL https://ec.europa.eu/eurostat/documents/1001617/4398365/S02P1-NORC-DATA-ENCLAVE-SCHEUREN.ppt.

J. Marcotte, S. Rush, and K. Ogden-Schuette. Tiered access to research data for secondary analysis, 2020. University of Michigan.

S. McEachern. CADRE Five Safes Framework—conceptualisation and operationalisation of the Five Safes Framework: Coordinated access for data, research and environments, 2021.

National Research Council. Proposed revisions to the common rule for the protection of human subjects in the behavioral and social sciences, 2014.

Office for Statistics Regulation. Joining up data for better statistics, 2018a. Systematic Review Programme Report.

Office for Statistics Regulation. Regulatory guidance—building confidence in the handling and use of data, 2018b. Systematic Review Programme Report.

Office for Statistics Regulation. Joining up data for better statistics, 2019. Systematic Review Programme Report.

I. Oppermann (ed.). Privacy in data sharing: a guide for business and government, 2018. Australian Computer Society.

I. Oppermann (ed.). Privacy-preserving data sharing frameworks people, projects, data and output, 2019. Australian Computer Society.

Organization for Economic Cooperation and Development. OECD expert group for international collaboration on microdata access: Final report, 2014.

M. Rahman, M. Jirotka, and W. Dutton. Lost in reality: The case for virtual safe settings, 2007. Mimeo, University of Oxford.

J. Raisaro, F. Marino, J. Troncoso-Pastoriza, R. Beau-Lejdstrom, R. Bellazzi, R. Murphy, E. Bernstam, H. Wang, M. Bucalo, Y. Chen, A. Gottlieb, A. Harmanci, M. Kim, Y. Kim, J. Klann, C. Klersy, B. Malin, M. Mean, F. Prasser, L. Scudeller, A. Torkamani, J. Vaucher, M. Puppala, S. Wong, M. Frenkel-Morgenstern, H. Xu, B. Maiyaki Musa, A. Habib, T. Cohen, A. Wilcox, H. Salihu, H. Sofia, X. Jiang, and J. Hubaux. SCOR: A secure international informatics infrastructure to investigate COVID-19. *Journal of the American Medical Informatics Association*, 27(11):1721–1726, 2020. doi: 10.1093/jamia/ocaa172.

F. Ritchie. Secure access to confidential microdata: four years of the Virtual Microdata Laboratory. *Economic Labour Market Review*, 2:29–34, 2008. doi: https://doi.org/10.1057/elmr.2008.73.

F. Ritchie. International access to restricted data: A principles-based standards approach. *Statistical Journal of the IAOS*, 29(4):289–300, 2013). doi: 10.3233/SJI-130780.

F. Ritchie. Access to sensitive data: Satisfying objectives rather than constraints. *Journal of Official Statistics*, 30(3):533–545, 2014. doi: https://doi.org/10.2478/jos-2014-0033.

F. Ritchie. The "Five Safes": a framework for planning, designing and evaluating data access solutions, 2017a. Data For Policy Conference 2017.

F. Ritchie. Spontaneous recognition: an unnecessary control on data access?, 2017b. European Central Bank Statistical Papers 24.

F. Ritchie. Analyzing the disclosure risk of regression coefficients. *Transactions on Data Privacy*, 12(2):145–173, 2019.

F. Ritchie. Frameworks, principles, and accreditation: Making data governance work, 2022. URL https://uwe-repository.worktribe.com/output/10630819. Presented at RSS Data Ethics and Governance—origins, progress and priorities, London.

F. Ritchie and F. Tava. Five Safes or One Plus Four Safes? musing on project purpose, 2020. Bristol Centre for Economics and Finance blog.

Security Brief Australia. IXUP embeds Five Safes framework in platform, 2019. URL https://securitybrief.com.au/story/ixup-embeds-five-safes-framework-in-platform.

J. Sikorska, S. Bradley, M. Hodkiewicz, and R. Fraser. DRAT: Data risk assessment tool for university-industry collaborations. *Data-Centric Engineering*, 1:e17, 2020. doi: 10.1017/dce.2020.13.

R. Silberman. Developing access to confidential data in France: results and new challenges. *Journal of Privacy and Confidentiality*, 11:2, 2021.

S. Tam. On a disclosure probability statement for the Five Safes framework. *Statistical Journal of the IAOS*, 37:693–698, 2021. doi: 10.3233/SJI-200677.

S. Thomas. Safe data—Statistics Canada's Data Confidentiality Classification Tool, 2020. URL https://copafs.org/wp-content/uploads/2020/03/COPAFS-Thomas-Safe-Data.pdf. Presentation to the Council of Professional Associations of Federal Statistics, Melbourne.

UK Health Data Research Alliance. Trusted research environments (TRE), 2020. Green Paper v1.0.

UK Office for National Statistics. Secure data service risk assessment, 2011.

UK Office for National Statistics. Safe researcher training 2017 onwards, 2020.

Understanding Patient Data. Public attitudes to patient data use: A summary of existing research, 2018.

United Nations Economic Commission for Europe. Unece/eurostat work session on statistical data confidentiality: Summary report, 2019. URL https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2019/mtg1/SDC_2019_Report.pdf.

N. Volkow. Harnessing the potentiality of microdata access risk management model, 2019. UNECE/Eurostat Work Session on Statistical Data Confidentiality, The Hague.

B. Weaver and J. Richardson. *Reinventing Library Research Support Services at Griffith University*, pages 267–269. IGI GLobal, Hershey, PA, 2021.

Wellcome Trust. Qualitative research into public attitudes to personal data and linking personal data, 2013.

D. Whittard, F. Ritchie, M. Rose, and R. Musker. Measuring the value of data governance in agricultural investments: A case study. *Experimental Agriculture*, 58:e8, 2022. doi: https://doi.org/10.1017/S0014479721000314.

D. Wiltshire. Using secure access data safely, 2021. URL https://youtu.be/svx-n4LEh6M. SSHOC Workshop.

T. Wirth, F.and Meurers, M. Johns, and P. F. Privacy-preserving data sharing infrastructures for medical research: systematization and comparison, 2021.

Y. Zheng, A. Pal, S. Abuadbba, S. Pokhrel, S. Nepal, and H. Janicke. Towards IoT security automation and orchestration, 2020. Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA).