Faculty of Business and Law

Five Safes: designing data access for research

Tanvi Desai[1], Felix Ritchie[2] and Richard Welpton[2]

[1] University of Essex
[2] University of the West of England, Bristol

Economics Working Paper Series
1601

Five Safes: designing data access for research

Tanvi Desai[1], Felix Ritchie[2] and Richard Welpton[2]

[1] University of Essex

[2] University of the West of England, Bristol

Abstract

What is the best way of managing access to sensitive data? This is not a straightforward question, as it involves the interaction of legal, technical, statistical and, above all, human components to produce a solution. This paper introduces a modelling tool designed to simplify and structure such decision-making.

The Five Safes model is a popular framework for designing, describing and evaluating access systems for data, used by data providers, data users, and regulators. The model integrates analysis of opportunities, constraints, costs and benefits of different approaches, taking account of the level of data anonymisation, the likely users, the scope for training, the environment through which data are accessed, and the statistical outputs derived from data use.

Up to now this model has largely been described indirectly in other papers which have used it as a framing device. This paper focuses specifically on the framework, discusses usage, and demonstrates where it sits with other data and risk management tools. The aim is to provide a practical guide to the effective planning and management of access to research data.

JEL Code: C81

Key words: data access, data management, confidentiality, security engineering, statistical disclosure control

Corresponding author: Felix Ritchie, Bristol Economic Analysis, University of the West of England, Bristol, Coldharbour Lane, BS16 1QY. Email: felix.ritchie@uwe.ac.uk

1. Introduction

The statistical research value of data to society is substantial. In the public sector, epidemiological data is essential for evaluating the long-term effects of drugs or health policies; relative performance of schools is used to target resources; rehabilitation charities can use re-offending rates to determine the effectiveness of their programmes; tax systems are interrogated to analyse the likely effect of changes in social benefit provision; and so on. In a world of 'evidence-based policy-making', good source data used effectively is the key.

The private sector also recognise the value of data: the information from a supermarket loyalty card will be used to drive purchasing and promotion decisions; a bank's data on security trades will help to define the next trading period's activity and to develop new trading models, for example. Personal data collected by private companies is covered by statute, but collection and use of the data is a hard-nosed commercial decision. For such organisations, the data are a valuable internal asset, and confidentiality is essential to preserving the value of that asset. The data owner will seek to extract as much value as possible. Data are likely to be distributed (even internally) only under restrictive arrangements.

In contrast, the research value of data collected by public organisations is maximised by making the data as widely available as possible (Trewin et al, 2007; Ritchie and Welpton, 2011; Desai, 2012); in some countries this is effectively a legal requirement. This poses a dilemma for the data owners, as unrestricted distribution of the data is not compatible with expectations of confidentiality. Much of the data acquired for research through administrative or survey sources is collected with either explicit or implicit guarantees of confidentiality. For example, surveys carried out by National Statistical Institutes (NSIs) normally have an explicit guarantee; information acquired as part of a tax submission or medical procedure is implicitly expected to be confidential.

Historically, public sector data owners have taken one of two approaches prior to distribution: anonymising the data, or having users sign confidentiality agreements to access confidential sources. An alternative has been not to distribute the data at all but to require researchers to visit secure facilities. More recently, advances in technology have expanded the range of options to include secure remote working solutions, real-time anonymisation, and synthetic data.

This variety of options for public sector data owners creates problems. First, the options reflect fundamentally different perspectives on risk management; for example, the underlying assumption of anonymisation is that users cannot be trusted to participate in protection, while the underlying assumption of confidentiality agreements is that they can. Second, different options reflect different finance models: anonymisation requires large initial investment, monitoring confidentiality agreements requires ongoing expenditure, and research data centres incur some of both. Third, the legal frameworks surrounding data access do not reflect operational solutions; they are defined in terms of broad concepts such as 'identifiable data' which are open to interpretation. Hence, institutional preference and custom can be misinterpreted as legal proscription (Ritchie, 2014a). Finally, data access decisions are likely to be framed in the language of the investing organisation; this makes discussions with outside bodies (for example, for data sharing) more complex, and may limit the opportunities for bodies to learn from each other (Ritchie, 2013).

These arguments may be intertwined: a data owner may feel unable to provide access to confidential sources due to its interpretation of legal frameworks, little ability to invest in data infrastructure, and little understanding about research potential and how researchers operate.

Indeed, some data owners may only consider data access for 'internal' use, and as an interesting by-product of their data management systems.

The combination of interrelated choice variables (risk, cost, technology, lawfulness, institutional factors) can make decision-making difficult. As Ritchie (2014a) notes, public sector bodies are more likely to focus on the 'what can we do?' operational model, rather than the 'what should we do?' strategic perspective. This provides a particular problem for data sharing and outsourced solutions, as organisations focused on implementation may have no common language with third parties. This is particularly true when, in the experience of the authors, data owners may focus on one, and only one, particular issue (such as the legal framework surrounding access to their data, or IT solutions). What is required is a more holistic approach to providing access to data.

One framework that has become increasingly popular in recent years is the so-called Five Safes model, which seeks to provide a common language and framework for data access irrespective of the particular circumstances. It breaks access discussions down into separate issues, allowing conversations to be more focused. Most importantly, it focuses on outcomes and objectives rather than specific solutions, although it has been used to design implementations. The overall impact is to provide a common frame of reference which can be applied to all data access questions.

This model has been employed in a variety of situations in different organisations across the world. Many of those uses relate to controlled access facilities, as this is where recent developments in data access have been concentrated; but the model is designed to be generally applicable: two statistical agencies have used it to frame, define and discuss entire data access policies (ONS, 2011a) or strategies (Webster, 2015), including anonymised data on the web, secure on-site labs, and outsourced access delivery. More recent examples of application to distributed data solutions are RCUK (2008), Corti et al (2014), and Hafner et al (2015).

This model has been described indirectly in a number of published works. The purpose of this paper is to focus on the Five Safes model and explain its purpose, use and limitations. The next section outlines the model in its simplest form and its use. Section three discusses each of the dimensions in detail. Section four considers its implementation in description, design and evaluation of data access strategies. Section five reviews a popular simplification of the framework, the 'Data Access Spectrum'; this simplification is particularly valuable as it shows how the 'traditional' model, which emphasises statistical controls, sits within the wider framework. Section six briefly discusses the implications for decision-making in this framework, which brings the subjectivity and uncertainty in decision-making to the fore. Section seven discusses the use of the framework in the context of qualitative data. Section eight concludes.

For the sake of clarity in exposition, this paper assumes that the data owner considering its access strategy is a National Statistics Institute (NSI), such as the US Census Bureau, the UK Office for National Statistics (ONS), or DeStatis in Germany; and we will mostly draw examples from quantitative data. However, the arguments here are relevant to all data owners in the public and private sector, and to qualitative data, as this is a generic framework to facilitate all data management discussions.

2. The Five Safes framework

The Five Safes framework, also sometimes called the VML Security Model[1], was originally developed in 2003 as a way to describe the Virtual Microdata Laboratory, a 'safe centre' for confidential research data at ONS, the UK's NSI. The purpose of the model was to simplify the complex discussion around data access into a set of related but independent questions, so that each topic could be dealt with succinctly and unambiguously.

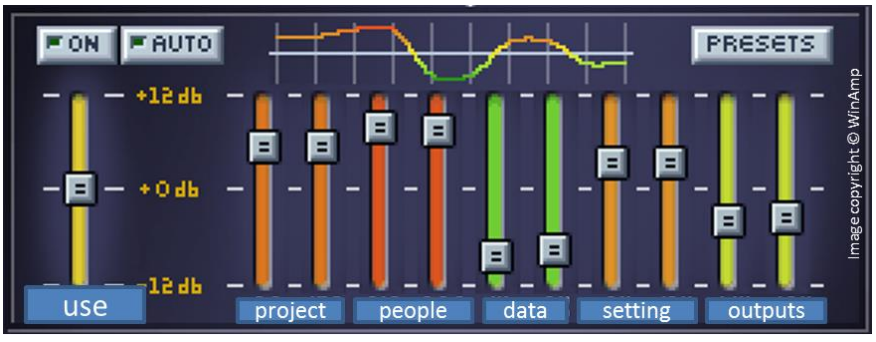
The basic premise of the model is that data access can be seen as a set of five 'risk (or access) dimensions'[2]: safe projects, safe people, safe data, safe settings, safe outputs[3]. Each of these dimensions provokes a question about data access:

Safe projects	Is this use of the data appropriate?
Safe people	Can the researchers be trusted to use it in an appropriate manner?
Safe data	Is there a disclosure risk in the data itself?
Safe settings	Does the access facility limit unauthorised use?
Safe outputs	Are the statistical results non-disclosive?

These dimensions embody a range of values: 'safety' is a measure, not a state. For example, 'safe data' is the dimension under which the safety of the data is being assessed; it does not mean that the data is non-disclosive. Nor does it necessarily specify how the dimensions should be calibrated. 'Safe data' could be classified using a statistical mode of re-identification risk, or a much more subjective scale, from 'very low' to 'very high'. The point is that the user has some idea of 'more safe data' and 'less safe data'. We return to the subjectivity of assessments in the penultimate section.

There is an analogy in multi-criteria decision analysis (MCDA; see Ishazaka and Nemery, 2013, for a description, and Nutt et al (2015) for an example). MCDA recognises that the many factors affecting a decision might not be specified in ways which lead to simple numerical models of 'best' outcomes. As such decision-making is explicitly subjective, expert-based, and negotiated across incompatible dimensions.

An alternative analogy was provided by McEachern (2015), who proposed that the model is akin to a graphic equaliser:



[1] The name Five Safes arose from a Google New Zealand search for 'felix ritchie Five Safes' in 2013. We are grateful to an unknown Kiwi for suggesting this much better name.
[2] The model was originally developed for access to identified data. 'Safe data' was added to the model in 2007 to allow it to be extended to partially or fully de-identified data.
[3] Some papers (such as Volkow, 2014) use 'secure' rather than 'safe'

The key to the model is that the five dimensions severally and jointly contribute to any consideration of whether a data access stratagem meets expectations:

- Dimensions are designed so that each can be evaluated independently of the others, as far possible
- All five dimensions need to be considered jointly to evaluate whether a data access system provides an 'acceptable' solution

These two precepts are considered below.

Precept 1: Dimensions are independent

The independence of dimensions is essential to the usability of the framework. The original model was designed so that decision-makers could focus on one specific issue at a time. Consider the question

> Question A: Should researchers be allowed to access Dataset X for epidemiological analysis?

The potential answers are yes, no and maybe; but each implies further questions:

Answer A	Question B
A1. Yes	B1. How do we make sure that the researcher uses the data appropriately?
A2. No	B2. Is it not lawful? B3. Is there concern that findings may be damaging to the data owner? B4. Are there ethical concerns?
A3. Maybe	B5. Does it depend upon the research question? B6. Does it depend on the researcher? B7. Does it depend on the facility? B8. Does it depend on the outputs produced? et cetera

The first two answers lead to a useful discussion of the acceptable range of purposes for which research is allowed. The 'maybe' does not, in general, because the follow-up questions add further uncertainty. However, it is not clear why the follow-up questions help Question A to be decided. If the NSI follows up Question A with Questions B6, B7 or B8, then the answer to Question A should have been 'yes'. To see this, suppose that the NSI thinks the answer is that 'it depends on the facility'. Then the discussion can be refined:

> Should researchers be allowed to access Dataset X for epidemiological analysis? *It depends upon the facility [Response: This answer is not helpful; why should the ethical correctness of the decision in principle depend upon technology used to analyse? Try again]*
>
> ⇨ Should researchers be allowed to access Dataset X for epidemiological analysis*?* ***Yes, if*** *the facilities are appropriate [No, still saying you can't make an ethical decision even in principle. Have another go]*
>
> ⇨ Should researchers be allowed to access Dataset X for epidemiological analysis? ***Yes, assuming that*** *appropriate facilities exist. [Better, but what do you mean by appropriate, and shouldn't the assumption be a criterion?]*
>
> ⇨ Should researchers be allowed to access Dataset X for epidemiological analysis? ***Yes, ceteris paribus*** *(all other necessary conditions being in place) [Much better; now focused on main issues – what makes a valid research project? - and not sidetracked by specific conditions]*
>
> ⇨ Should researchers be allowed to access Dataset X for epidemiological analysis? ***Yes***

The only 'maybe' follow-up answer that is relevant is 'It depends upon the research question'. This is within scope of the 'safe projects' risk dimension, and would require the NSI to refine the decision-making process – but without reference to any other risk dimension.

This focus on a specific issue has two benefits. First, it allows decision-makers to concentrate on the matter in hand. Second, it removes the co-dependence between independent decisions. Suppose that the validity of the project is tied to the researcher having been appropriately trained. If the training changes, then that implies the validity of the project changes too. Why? The legal framework, ethical concerns and reputational matters have not changed in terms of the project objectives. The only reason the project validity has changed is because it is (unnecessarily) dependent upon another risk dimension.

Precept 2: Dimensions need to be considered jointly

The Five Safes is a system model. That is, it is intended to review how all the elements fit together. Taking the example above, the answer to whether a researcher is allowed to access a dataset assumes that *all other necessary conditions are in place*. Supposing secure facilities do not exist; then this does not seem like a good use of the data. However, this does not mean the questions 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 settings' dimension.

Not all controls need be used; for example, data made available on the internet are open to anyone, including malicious use. This does not mean that 'safe projects', for example, should be not be considered; it is more correctly interpreted as 'checks on the suitability of data use are not feasible, and so project use is deemed to be uncontrolled'.

An analogy is with evaluating a meal at a restaurant. The key objective can be seen as 'creating a positive eating experience which meets a need'. Multiple criteria can be used: taste, calorific value, aesthetic value, price, convenience. Each of these can be assessed independently but the overall success of the dining experience depends upon the combination of factors – which may change in different circumstances. For a Michelin-starred restaurant, taste and aesthetic value are likely to be the only criteria under consideration; the kebab van is judged on calories, price and convenience; a mid-range restaurant may pass on an acceptable rating on all criteria, but without needing to be exceptional on any.

Overall, multiple solutions might meet the objective. This does not mean that one solution is better than another in any meaningful sense; different needs are reflected in different elements of the eating experience being seen as important. Similarly, creating a low-detail open dataset is not 'better' or 'worse' than creating a controlled-access detailed micro-dataset; it is just different.

Note that the standards are assessed as the minimum. A Michelin star may be a bonus but is not a necessary condition to be a successful kebab van (and is likely to be expensive). Similarly, if the architecture in a secure facility is designed to be proof against most hacking, an NSI can give its researchers minimal training but may decide to carry out additional training anyway. Just as for the kebab van, exceeding minimum standards is likely to cost money.

Welpton (2013) argues that a careful study of the costs associated with the underlying assumptions of each dimension provides a natural order for decision-making. Changing the 'safety' of the data is a one-off cost, but the other dimensions require both initial and ongoing costs. For example, anonymising a dataset for a Public Use File (PUF) requires a large initial investment to produce 'safe data', but then may incur no further costs. Producing Scientific Use Files (SUFs) may or may not cost

less, but it incurs additional costs for people and project assessment. Running a restricted facility may involve significant set-up costs, and will also require ongoing expenditure in all dimensions except 'safe data'; hence the marginal cost of each user is likely to be of more relevance to the data owner.

For PUFs, the marginal cost of additional users is effectively zero: after the initial investment, there need be no controls in other dimensions, implying no other costs. Hence, the size of the user base determines the cost-effectiveness of creating a PUF. For a restricted access facility, ongoing costs means the net benefit from each use of the data needs to be more carefully assessed. For example, manual checking of all outputs is a pure variable cost; however, this cost may be mitigated by considering protection in other elements of the framework, such as training.

Does the order matter?

The order in which the Five Safes are presented is a matter of some debate amongst practitioners. For example, Hafner et al (2015) see 'data' as the residual, and hence leave it until the end; others with the same perspective place 'data' at the beginning, to get it out of the way. In theory, independence of dimensions means that the order is irrelevant, and they are discussed below in a popular order. However, the order does have psychological and institutional impact; we return to this in the conclusion.

3. The dimensions in detail

3.1 Safe projects

'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.

Somewhat harder to identify is whether projects should be allowed when their use might damage the reputation of the data owner or the subjects in the data. For example, if statistical tables are likely to be distorted or misrepresented by pressure groups, should the outputs be produced? What if the data is of too low a quality to produce robust statistics to normal NSI standards? What if the project outputs are critical of the data owner? What if the project might produce a damaging impression of data subjects?

In terms of choosing a level of 'safe projects', the data owner is essentially deciding on the necessary checks. For example, the data owner could consider four levels of project safety:

Standard	*Advantage to data owner*
No restrictions on use	Lowest cost
Administrative information on researchers (name, residence etc)	Record of use – can measure demand, and check users in case of problems
Checklist for meeting legal, ethical requirements	Necessary minimum checks carried out – ensuring that, for example, valid statistical purposes are identified
Justification to be scrutinised by expert assessors	Potential to actively engage in dialogue with researchers – ensuring that the data owner is fully aware of use, and possibly leading to productive methodological discussions with users

Each option has different costs and benefits, as well as different risks. Note that, while the options listed here appear to decrease in riskiness, all other things being equal, this is not necessarily the case. Only the direct risk to the data owners is falling. Ritchie and Welpton (2012) argue that the risk of loss to society from not allowing valid projects to go ahead is rarely considered by public-sector data owners, and yet this risk may increase as project approval moves from an automatic to a manual process.

3.2 Safe people

'Safe people' reviews the knowledge, skills and incentives of the users to store and use the data appropriately. It considers the confidence of the data owner that those who will access to the data can be trusted to use it appropriately. In this context, 'appropriately' means 'in accordance with the required standards of behaviour', rather than whether statistical skills, for example, are up to the mark. In practice a basic technical ability is often necessary to understand training or restrictions and avoid inadvertent breaches of confidentiality; an inability to analyse data may lead to frustration, and increases incentives to 'share' access with unauthorised people.

3.2.1 Identifying the problem

Any release of data must assume that the users know what they are allowed to do with it. For example, the Scientific Use Files issued by the UK Data Service came with restrictions on where the data should be stored, and what it could be used for (UKDA, 2014).

The data owner needs to consider:

- Does the user know how the data should be stored?
- Does the user know how the data should be used?
- How do I communicate this information?
- What does it matter if this information isn't understood?

Even if the user does know how the data should be used, this does not mean the rules will be followed. Restrictions on data use, for example, might frustrate the user who then finds ways to get round those restrictions. A consideration of breaches in research access to data shows that the great majority of use errors in the academic community arise from mistakes by the user. The next largest group of errors are deliberate misuse by researchers who choose to ignore rules because they find them inconvenient, not because of any urge to re-identify individuals. The more restricted the use, the more likely it is that users will become frustrated; this is a well-known security risk (Desai and Ritchie, 2010). Hence there needs to be appropriate consideration of the incentives to ensure that use is appropriate.

Threats of legal action have little credibility, as no NSI has ever (to the authors' knowledge) pursued a criminal case against a researcher misusing data. OECD (2014) comes to similar conclusions about the unworkability of criminal sanctions, pointing out a number of practical difficulties. Despite this, much effort is still expended on the assumptions that data needs to be protected against hostile misuse, and that frightening researchers with tales of the legal penalties will prevent this[4]. This 'deterrence' model is popular with data owners because it allows them to distance themselves from potential risks: harm is created by unlawful action by trusted researchers, not by any failure in processes (Tyler et al, 2013).

[4] The authors have collected many examples of poor training practices in the government, academic and privates sectors. As the aim of this paper is to develop good practice rather than shame specific organisations, bad practices are not cited here.

In contrast, there is a substantial body of practical experience that procedural penalties (eg disciplinary hearings, loss of access to data or funding) are taken much more seriously by users; this is reflected in, for examples, the penalties associated with misuse of the UK Data Service. More importantly, it has been demonstrated that engaging researchers so that they understand the reasons for restrictions on use can create a virtuous circle: trustworthy researchers allow a more tolerant research environment to be used, reducing the likelihood of researchers finding ways round restrictions (Desai and Ritchie, 2010; Jackson et al, 2012; Ritchie and Welpton, 2015; Tyler et al, 2015).

The only case where hostile usage of NSI data is a serious consideration is unrestricted internet access. There are examples of NSI public outputs being attacked, with malicious intent, to uncover source data, perhaps for political or religious reasons (eg McHale and Jones, 2012); there are also examples (for examples, the Netflix case) of 'anonymised' data being de-identified, to demonstrate how badly the anonymisation was done. Amongst technically naïve users, there is also a higher probability of breaches of confidentiality being inaccurately claimed (for example, one NSI was accused of giving detailed Census microdata to a supermarket, when in fact the supermarket was merely making use of tabular data in the public domain). In such a case, the reputation of the data owner could suffer even if the claim is false.

3.2.2 The role of training

Software licences frequently require buyers to agree to an extensive list of conditions before software can be installed. Agreement is achieved by passive acceptance: having a large friendly button at the beginning of the installation process. Users can actively work through much legal phrasing to be informed about what they are agreeing to, but the assumption of the vendors is that no-one has the time and expertise to do so. This shows a good understanding of human psychology; unfortunately for software vendors, courts have the same understanding and increasingly the passive acceptance of licence conditions is being declared to be unenforceable.

For many data owners, however, the passive acceptance model still appeals. For example, Eurostat used to send users of its scientific-use files a detailed document explaining the conditions of access and how safe outputs can be achieved. For data access professionals this was an interesting document, but it was unlikely that researchers read it with the same avidity. In such a case, what assumptions can be made about the knowledge and incentives of these researchers?

As always, the answer depends upon context. Hafner et al (2014) argue that the Eurostat document did provide appropriate protection, *in the specific case of the dataset under discussion*: as the complex application process reiterates the penalties from misuse and the document itself is relatively short and non-legalistic, it is a fair assumption that the key elements of the document have been absorbed even if the detail is lost. Risk scenarios in Hafner et al (2014) therefore focused on accidental misuse. Interestingly, Eurostat has since reviewed its user guides and in 2015 produced a new user guide designed along modern psychological principles of positive engagement rather than deterrence (Eurostat, 2016).

As release environments become more complex, the need for active acceptance of access conditions through some form of direct engagement grows. In the case of controlled access facilities, a number of recent papers (Desai and Ritchie, 2010; Brandt et al, 2010; Hawkins, 2011;Welpton and Ritchie, 2012; Welpton and Kinder-Kurlanda, 2013; McEachearn, 2015; Webster, 2015; Ritchie and Elliott, 2016) have argued that training should go beyond mere acceptance of conditions. The primary focus should be to ensure that researchers share the same goals as the data owners, so that appropriate behaviour is a natural consequence of shared objectives. This has positive spillovers as the

replacement of 'them and us' planning with collective responsibility has been shown to produce a virtuous circle of behaviour and environment. This acculturation training is more expensive but can be shown to be cost-effective.

3.2.3 The myth of the 'good professor'

Data owners may have little knowledge of data users. For example, NSIs may have little knowledge of academic working practices. This encourages naïve assessment of researcher 'safety' by measurable criteria: job title, long lists of publications, membership of professional bodies, and so on. This leads to applications becoming a series of tick boxes for the assessor.

There is no hard evidence to support any of these criteria. However, there is a strong perception amongst data professionals that seniority is a very poor guide to trustworthiness; if anything, there is a negative relationship. This is not just limited to academic and government researchers; amongst data owners, resistance to change and a preference for precedent over evidence is observed to be correlated with seniority[5].

3.2.4 Reciprocity

So far, the discussion has focussed on researchers as Safe People. Yet, the relationship is two-way. One would hope that anybody accessing their data is willing to learn about safe use, and engage with the data owner to work together and achieve optimal data access solutions.

Sadly, cases exist where data owners expect users to behave in a certain way; but are not prepared to engage in return. Examples where data owners (or providers) are unwilling to listen to users who try to engage and work with them, lead to poor outcomes: at best, bad reputation; at worst, security breaches. We can include data owners / providers as Safe People too.

3.2.5 Summary

When considering the 'trustworthiness' of researchers, there needs to be consideration of the role of the data owner in ensuring that that level of trust is a fair assumption. In some circumstances no training or passive acceptance of licence conditions may be appropriate; in others there may be a need for active training to demonstrate that assumptions on researcher behaviour have some supporting evidence. Finally, it should be noted that more involved the training, the potential for acculturation of users to have positive spillovers for other dimensions.

3.3 Safe data

Safe data refers primarily to the potential for identification in the data. It could also refer to the sensitivity of the data itself, but for argument's sake we focus on the former case; without identification of an individual or group there is no breach.

There are various ways to assess 'safety' of a dataset. A popular approach when considering creating anonymised microdata is to map the probability of a successful match to an external database; see Hundepool et al (2012), or Spicer et al (2014), where the external database is an individual's private knowledge. Factors such as sampling probabilities, data accuracy and population uniques can be included in the assessment. An alternative is 'k-anonymity', which ensures that, for a set of identifying variables, there are always at least k records matching a set of characteristics; an intruder only has 1/k probability that an individual identified in the released data is the original respondent

[5] Source: personal discussions with colleagues in multiple countries and environments, usually in the context of how training should be targeted.

(Sweeney, 2002). Thus the safest data may have k set to ten and include all categorical variables as potential identifiers, whereas licensed datasets may have k set to 3 and only age, gender and detailed geography used as identifiers.

These and other measures provide numerical estimates which are particularly useful when comparing alternative protection mechanisms. Perhaps more importantly for data owners, they also appear to provide objective measures of re-identification risk. However, as Skinner (2012) points out, this 'objectivity' rests upon a set of assumptions which are entirely subjective: the amount of information available to the intruder, the comparability of external datasets, the time available to the intruder, and so on. Nevertheless, this can be a useful relative measure of risk.

There is an extensive literature on protection of microdata: almost all academic research in the past half-century addresses just microdata protection or tabular data protection, with a small literature on synthetic data arising in the past decade or so. None of this considers non-statistical protection factors beyond a brief description of a notional release environment, to enable breach scenarios to be specified. As Hafner et al (2015) note, this makes sense in the context of academic discussions; however it has had serious consequences in practice by emphasising the statistical element of data protection, to the detriment of other methods. Our response to this was the development of the 'data access spectrum', which makes clear that anonymisation of data should be treated as a residual control; see Section six, below.

When dealing with facilities rather than specific datasets, it may be more appropriate to identify general rules defining the types of data suitable for hosting in the facility. One option which has been used in the UK is to group datasets into classes of identification; for example, by whether they contain

- direct and recognisable identifiers (eg name and address)
- direct but not recognisable identifiers (eg social security numbers)
- indirect identifiers (eg postal code, age, gender and occupation for an individual)
- no identifiers in general but some exceptional values (eg 109-year old male living in Brussels, or a personal history describing how the user moved from Guinea-Bissau to Tiger Bay in the 1990s)
- no identifiers or extreme values in any tables of three or fewer dimensions

This scheme has the advantage that it links directly to the access environment. The first grouping is clearly identifying in all circumstances without further information; the second requires access to the internet and/or a list of external IDs; the third and fourth requires access to the internet and/or third-party information, and time to research; finally, the fifth assumes that even with the internet and unlimited time, the chances of an individual being identified are small.

The key is that the rules can be used in design without knowledge of a specific dataset; anonymisation takes place on a specific dataset once the context for data release is known (that is, its place on the data access spectrum). Hence anonymisation become a residual function.

3.4 Safe settings

'Safe settings' relates to the practical controls on the way the data is accessed. At one extreme researchers are restricted to using the data in a supervised physical location (a 'research data centre', or RDC); at the other, there are no restrictions on data downloaded from the internet. In between, there may be requirements that data is held on restricted access servers, or that CDs must be encrypted and only decrypted at the point of use, for example. Safe settings encompasses both

the physical environment (network access, use of portable storage et cetera) but also procedural arrangements such as the supervision and auditing regimes.

For safe settings, the likelihood of both deliberate and accidental disclosure needs to be explicitly considered. For example, if data are sent out on CD, then there are numerous examples of CDs going missing or not being stored appropriately; the *Guardian* (2015) reports one confidential leak every fortnight in the UK. In a more complex example from the authors' own experience, researchers had access to a data file on the condition that it was only accessed in a restricted server, and that all copies of the data were destroyed at the end of the project, including backups. As a result, the data had to be uploaded and removed each working day to evade the overnight backups.

In recent years, the most important growth area for research on government data has been the development of remote working solutions. Some countries have explored remote job servers (also known as remote execution, where researchers sent in code to be run by the data owners and receive results), but the major developments have been in remote RDCs. These use thin-client technology to allow researchers to have full access to data as if it held on their own machines By providing a familiar convenient environment, this increases the likelihood of 'safe' behaviour, since researchers avoid restrictions that create barriers to research.

The growth in remote RDCs raises interesting problems. These are designed to allow researchers to access very detailed data without the restrictions and supervision of the physical RDC; but that also implies that the researchers have more freedom to use the facility inappropriately. Physical supervision mechanisms are unlikely to be appropriate (and aren't used anyway, in the experience of the authors: CCTV cameras are not usually regularly monitored, for example); on the other hand, electronic supervision and audit logs can be much more effective.

3.5 Safe outputs

The final dimension covers the residual risk in publications from sensitive data. For example, a quantitative researcher could be tempted to suggest "Company X accounts for almost all of the variation…"; an observational study might note "respondent Y said she didn't like living over the wig shop…". 'Safe outputs' exists because, no matter how well-intentioned and competent a researcher may be, mistakes happen to all types of organisations.

Safe output is most easily managed in restricted access facilities, where facility managers can check all statistics being released and, if necessary, block their release. In other situations, the data owners must rely upon the training and knowledge of the researchers. For example, SUFs are often distributed with some guidelines on good practice when producing outputs; the data owner then hopes that the user will read the guidance and act appropriately, but cannot guarantee it. For example, SUFs distributed by Eurostat come with extensive guidance on what is expected in terms of minimum cell size and dominance in tables.

Historically, NSI and academic interest has focused on tabular outputs, as these are primary outputs for NSIs; tables are also inherently risky. Other outputs were largely ignored; where checking of outputs was deemed necessary (for example, from restricted research facilities), assessment was on a case-by-case basis. However, from the mid 2000s a number of authors have developed the more generalised field of 'output statistical disclosure control' (OSDC); see Ritchie (2007, 2014b).

Central to OSDC is the concept of 'safe statistics' (Ritchie, 2008, 2014b; Brandt et al, 2010). This is a system for classifying types of output (such as tables, regressions, or odds ratios), acknowledging that many research outputs pose no disclosure risk because of their functional form. This allows the

data owner to concentrate resources on the most risky outputs. This approach also provides a justification for the historical focus on tables to the exclusion of most other outputs.

3.5.1 Principles- versus rules-based OSDC

OSDC can be applied using either a 'rules-based' (RB) or 'principles-based' (PB) approach. RBOSDC means that hard-and-fast rules as to what is acceptable are applied; for example, every table must have at least ten units in every cell, and a dominance rule must be applied to every magnitude table. This is well-suited to automatic checking systems (such as remote job servers), or where a large number of similar outputs are being repeatedly produced (as in NSI outputs). Procedurally, it can simplify negotiations where output checkers must meet the needs of multiple data owners (for example, when Eurostat must reconcile different standards from member states). In addition, output checkers require little statistical skill or knowledge: RBOSDC is a 'box-ticking' exercise.

However, the rules-based approach suffers from the lack of context. This typically means that non-disclosive outputs are blocked unnecessarily (for example, by not taking account of transformations of the data; this is likely to be a significant concern in research environments). A more subtle concern is the mechanical nature of the rules-based approach, which means that inadequacies in the rules might go unnoticed (for example, in certain circumstances a table cell could be disclosive irrespective of the number of observations). This can only be countered in the rules-based approach by making the rules very strict, which increases the likelihood of unnecessary blocking of non-disclosive outputs. Worse, 'unsafe' outputs that meet a rule, regardless of context, may be released, as the output checker has less information to make an informed decision (lack of context).

In contrast, the principles-based alternative places context at the forefront of decision-making. All outputs are considered potential candidates for release, and 'rules-of-thumb' are used to provide a first approximation as to whether to approve the output or not. An output checker may decide that the output can be cleared even if it breaches the rules-of-thumb; or it may be blocked even if it meets the rules-of-thumb. A researcher can always make a case that a blocked output should be released; the output checker should consider such a case, but is under no obligation to accept it.

PBOSDC only works if both output checker and researcher understand the rules of output clearance, both technical and procedural, and therefore this requires training of researchers. To operate efficiently, thought also needs to be given to the incentives for researchers to produce good output, as this can dramatically affect the resources needed. However, if effective training can be implemented, then PBOSDC, when combined with the 'safe statistics' approach, is demonstrably more efficient and safer than other methods of outputs checking[6].

Although described above as alternatives, there is a relationship between rules-based and principles-based output SDC. PBOSDC is the generalised method; RBOSDC is the restricted model, appropriate in some circumstances. Both use rules, but the rules act as the starting point for PBOSDC output checking (see Brandt et al, 2010). There are two crucial differences:

- In PBOSDC, the rules are 'rules-of-thumb' – explicitly ad hoc, and amenable to adjustment, up or down, depending on circumstances.

[6] Evaluating costs and benefits is difficult because institutional features (costs) differ and inadvertent breaches (avoided costs) are almost non-existent; but Hafner et al (2015) note that the first PBOSDC system "…had operating costs 'an order of magnitude' lower than comparable systems across Europe, yet still supported more researchers with more data and faster response times." (p13)

- In PBOSDC, the rules of thumb can be more restrictive as the efficiency is not considered in the initial assessment.

Ultimately, PBOSDC takes rules-based SDC as a good first-order approximation, but gives expertise and experience the final decision. For this reason it is the recommended approach for research users (Ritchie, 2007; Ritchie and Elliot, 2016). In contrast, for NSI outputs something closer to RBOSDC is more appropriate (Eurostat, 2014). For a detailed discussion of the relative merits of RBOSDC and PDOBSDC see Ritchie and Elliot (2016).

3.5.2 Safe outputs and qualitative data

Safe output is a key problem when dealing with qualitative data. This is often very detailed, sensitive, and identifiable; qualitative research good practice therefore requires researchers to seek expert review. However, researchers typically work on their own machines and do not have output checking forced on them, so there is the possibility of publications not being reviewed for confidentiality. In addition, there are fewer guidelines on output checking because the types of data are so varied; this is expected to be one of the key development areas for SDC in the near future.

3.6 The subjectivity of 'safe' and risk

In data confidentiality 'safe' is a word loaded with implications of non-disclosiveness. However, in the above discussion the term 'safe' is used as a metric – how 'safe' are the data? The users? This use of 'safe' was a deliberate attempt to dismantle the binary meaning of safe/unsafe, and make it explicitly a relative concept.

This has the added advantage of bringing the subjectivity of data safety to the fore. If 'safety' is a relative measure, this implies the need to measure it, at least with reference to something else. It is immediately clear that none of the five dimensions can provide an *absolute* and *objective* measure of data safety. In certain circumstance it is feasible to provide *relative and objective* measures of safety, such as risk-utility maps comparing the effects of different anonymisation techniques. However, as Skinner (2012) points out, the disclosure scenarios within which the risk is estimated are parameterised as a result of subjective judgements: how is 'utility' measured? Skinner (2012) argues that risk assessment should always be seen as a subjective judgement, albeit with some objective support, and evaluated in that context.

In a similar vein, Ritchie (2014b) places the subjectivity of 'safe statistics' at the forefront of the classification of 'safe statistics'. There is no objective measure for determining as statistic as safe, but there is a significant amount of evidence on researcher behaviour which allows a 'reasonable' judgement that this is the correct classification. Brandt et al (2010) flag up the subjectivity by introducing an entirely arbitrary criterion (a minimum of ten degrees of freedom), even for 'safe' statistics.

Ritchie (2014a) notes that institutional perspectives are crucial in decision-making under uncertainty. First, if the default position is 'release unless not allowed', this should lead to more data release, all other things being equal. Second, becoming explicitly subjective about the nature of risk and uncertainty encourages use of evidence in decision-making, and allows accepted practice to be challenged. Welpton (2013) identifies 'classification creep' as a particular risk when data owners are being asked to make judgements on risk. In practice, this may not be the case, as greater emphasis is placed on empirical evidence rather than statistical theory; for example, Spicer et al (2014) demonstrates a significant gap between theoretical positions and empirical evaluation of re-identification risk.

Finally Hafner et al (2015) note that the term 'risk' is itself a misnomer; there exist no meaningful empirical risk assessments to support the disclosure scenarios used in analysis, because each scenario starts from a set of unevidenced assumptions. The relevant concept is Knightian 'uncertainty': not only do we not know the likelihood of an 'intruder' trying to break our protection measures, we have no quantifiable measures to start risk assessment. All we have is evidence from history and psychology; neither is a definitive guide to future behaviour. However, as the alternative is purely theoretical assumptions which are demonstrably wrong (eg Lenz, 2006; Hafner, 2008; Evans and Ritchie, 2009), acknowledging and allowing for subjective interpretations of evidence seems the more reasonable approach.

The subjectivity is reflected in the fact that the Five Safes framework is not designed to identify the 'best' solution. Whilst Ritchie (2013) suggests using arbitrary numbers to give a sense of disparities between options, the real value is to help identify different solutions (and possibly enveloping ones) by focusing on what matters in a particular context.

Returning to the restaurant analogy, conceptually the Michelin-starred restaurant may be 'best' from a gastronomic perspective; but to someone on a restricted budget the mid-range restaurant offers a better eating experience; and neither can compete with the kebab van next to the taxi rank at 2.30am when the clubs have closed.

4. Five Safes in practice

The Five Safes framework was originally specified in 2003 to explain a new RDC system to the UK Office for National Statistics (ONS) board. Since then it has expanded beyond its original home and is now widely used in description, design and evaluation across countries and facilities[7].

4.1 Five Safes as a descriptor

Various pre-existing facilities have adopted the Five Safes model as an organising description. For example, ICPSR, which has run the IPUMS international census data project for twenty years, now routinely describes its policies and strategies in terms of the Five Safes model. Statistics New Zealand's data lab has recently begun using the Five Safes to explain the criteria for access (Camden, 2014; SNZ, 2014). The Scottish Health Informatics Project (SHIP; Sullivan, 2011) adopted the framework once the project was underway; like the ICPSR, it was found to be a convenient way to describe the facility and discuss strategy, but it was also subsequently adopted for design criteria by associated organisation such as DaSH (Wilde, 2013).

4.2 Five Safes as a design tool

Increasingly, the Five Safes are used to provide the framework for designing data access strategies. For example, it was used by the Mexican Statistical Office, Statistics New Zealand (Camden, 2014) and Luxembourg Department of Social Security (Wagener, 2008) to design and develop their respective RDCs. In the UK, the tax department's DataLab used the framework as a design template (Hawkins, 2011), as did the Secure Data Service, set up at a similar time (Woollard, 2009); these two decisions reflected the design influence of the ONS data lab and the NORC Data Enclave,

[7] In the US, the Five Safes framework is often referred to as the 'portfolio approach', and is more associated with 'data enclaves' (remote RDCs) due to its early adoption in 2005 and effective promotion by the NORC Data Enclave (www.dataenclave.org) and influential strategists, such as Julia Lane of the National Science Foundation and the American Statistical Association's Confidentiality and Privacy Committee (eg ASA, 2009). The 'portfolio' usually omits 'safe data' as NORC adopted the pre-2007 definition.

respectively. The framework was used to identify the necessary business processes for the DARA European shared data project (Bujnowska and Museux, 2011); and the Five Safes principle of separating access into a set of interrelated but independent problems was used to operationalise workplans (albeit in a slightly modified form). Tubaro et al (2013) describe how a modified Five Safes was used to frame standards for the Data Without Boundaries international data sharing project. Most dramatically, in 2015 the Australian Bureau of Statistics redesigned its entire strategy using the Five Safes as the organising model (Webster, 2015), as has the UK's Health Foundation on a slightly smaller scale; see Wolters (2015) for a practical case study.

Ritchie (2013) proposed using the Five Safes as a way to resolve problems of international data sharing. The Five Safes framework naturally aligns to describing objectives. Ritchie (2013) argues that, while the technical and legal environments for data sharing may be country-specific, conceptual standards ("I do not want users being able to circulate this across the internet") can be agreed, allowing countries to find their own solutions which meet the standards. The analogy is with the internet: connection protocols allow computers to work together, even if specific operating systems or hardware varies substantially. Similarly, agreeing on the desired outcomes of any data access strategy allow different implementations to be developed.

The Five Safes has been less widely used for anonymising data, largely because of the dominance of the traditional data-centred models of SDC. Nevertheless, Hafner et al (2014) applied the Five Safes logic to the creation of SUfs for a multi-country business survey. They demonstrated that a more holistic approach could deliver substantial improvements in both confidentiality protection and data utility, with at little or no additional cost[8].

Finally, two cases where the Five Safes is a reference for a more detailed model.

Volkow (2014) use the model as the starting point to develop the OECD's practical recommendations on microdata access. The 'circle of trust' model used by OECD (2014) to operationalise specific proposals uses the Five Safes to describe how 'trust' is, or is not, established between partner organisations. For the 'inner circle' extensive control over people and settings allows detailed data to be shared; in the outer circle, the difficulty in establishing trust means that safe data becomes the primary dimension for confidentiality management.

A similar approach was taken in the Eurostat project described by Brandt (2010), where different configurations of 'star' or 'cluster' networks are considered. Irrespective of the network configuration, interoperability is given by ensuring that each of the nodes operated to the same standards for people, projects, settings and outputs.

4.3 Five Safes in evaluation

Because the Five Safes breaks down a complex problem into simpler components, it can simplify evaluations of data access plans (eg Bailey, 2012). Although the Five Safes is a discussion framework and therefore does not need empirical reference points, the underlying ethos is that evidence should be used to develop reasonable if subjective responses to known or likely problems. This makes it more suitable for evaluation than models which postulate unevidenced risk scenarios developed in the academic theoretical literature.

[8] Given (government) data owners' well-established preference for precedence when taking decisions, it is possible that the lack of application up to now has dissuaded others from experimenting with the Five Safes approach for microdata anonymisation. An alternative explanation is that almost no government data owners have practical experience in Five Safes thinking, and hence prefer to stick to familiar models.

For example, the formal evaluation and risk assessment of the Secure Data Service prior to its launch was structured around an evaluation of each dimension (ONS, 2011b). In 2008 Research Councils UK used the framework to critique an excessive focus on anonymisation in a government consultation on data access (RCUK, 2008).

In the US, Lane and Schur (2010) considered the appropriateness of the HIPAA regulations determining how behavioural and observational research should manage its data. One of the criticisms of HIPAA was that it focused too much on data anonymisation and technical solutions, but there was no mechanism to bring in other factors such as researcher training; the Five Safes approach would effectively generalise the existing regulations. In 2013 the workshop to revise best practice guidelines (NRC, 2013) came to a similar conclusion (amongst many others) and the National Research Council now explicitly recommends the use of the Five Safes framework as part of the replacement for (NRC, 2014, recommendation 5.1; Schomisch, 2014).

However, one of the difficulties in using the Five Safes for evaluation is its explicit subjectivity. In this, it has more in common with expert-led methods such as MCDA than with formal evaluation measures such as cost-benefit analysis, which seek to harmonise findings to a common quantitative measure.

4.4 Five Safes in communication and training

The Five Safes was originally designed to structure arguments towards, and convince, decision-makers. It continues to have this function in facilitating discussion, by providing a common frame of reference for organisations to resolve differences.

This is most noticeable when working internationally, where legal and cultural differences mean that discussions may even struggle to agree on what are the important things to be discussed. Ritchie (2013) argues that a framework like the Five Safes helps to structure the discussion because it does not ex ante place more importance on any organisation's view; instead it encourages putting concerns into homogenous groups. This was how it was used in the 'circles of trust' model of the OECD (2014) or the network models of Brandt (2010), for example.

This framework has also proved useful in training. A number of training packages (UKDA, NORC, Eurostat 2014, for example) structure their discussion around the Five Safes. This is particularly important for engaging users; the Five Safes can be used to show researchers how their contribution to safety fits in a wider context, and NSIs are not imposing arbitrary restrictions (for example). Anecdotal evidence suggests that most researchers do appreciate seeing the wider context, which should increase their willingness to act appropriately.

5. 'Safe data' as the residual: the Data Access Spectrum

A key insight from the Five Safes framework is that is emphasises the inherent undesirability of disclosure control techniques. The purpose of data access is to allow data to be analysed and results published. All dimensions apart from 'safe data' impose cash or resource costs on the researcher and/or data owner, but do not limit the research that can be carried out. In contrast, reducing the information content in the source data reduces the research potential. All other things being equal, fiddling with the data is to be avoided; safe data is the *residual* dimension.

In practice, all other things are not equal. On the supply side, as noted above, the costs of control in different dimensions vary in both size and whether they occur as one-off or repeated costs, reflecting one-off or ongoing risk management needs. However, these costs are often closely

interrelated. For example, it was noted that principles-based OSDC can be an extremely efficient and secure process, but that it requires substantial training; that is, the efficiency of both 'safe outputs' and 'safe people' are related. More generally, it can be observed that the non-data controls tend to move in step.

On the demand side, there is ample evidence that researchers are able to make intelligent decisions about whether the costs of access exceed the benefit of that access. As more detail in the data typically involves more restrictions on researchers, they balance desire for detail against resource costs of access. For example, Ritchie et al (2014) used restricted-access facilities and SUFs in the same research project. Again, from the researchers' perspective the non-data controls are typically seen as a single set of restrictions.

For example, data from ONS (the UK NSI) can be accessed via multiple routes:

- ONS staff have original data at their place of work
- All researchers can use ONS' RDC, accessible at ONS sites, which holds source data with direct identifiers removed
- Government researchers can access the ONS RDC from their offices
- Academic researchers can also use the Secure Data Service (SDS), a remote RDC available from university desktops which has almost the same data as the on-site RDC
- Researchers can get access to SUFs and PUFs through the UK Data Archive which come with use restrictions and less detail
- Custom tabulations can be requested from ONS
- The general public gets access to aggregate data, and some public data sources such as retail price data, and tabulations based on anonymised data via Nesstar
- For some datasets, third-parties run custom-built off-site secure facilities.

For most datasets, multiple forms of data release are used; for example, the Quarterly Labour Force Survey uses all of these access routes save the last.

A common expression of this perspective is the 'data access spectrum' (DAS) or 'continuum of access', the latter term being coined by Statistics Canada and the Canadian RDC Network; see Price (2014) for a recent expression. Put simply, on the assumption that the non-data controls move broadly in step, it reduces the dimensions from five to two by combining the non-data dimensions. Diagrammatically, this can be represented as follows (adapted from Ritchie, 2010, and Price, 2014), taking ONS as an example:

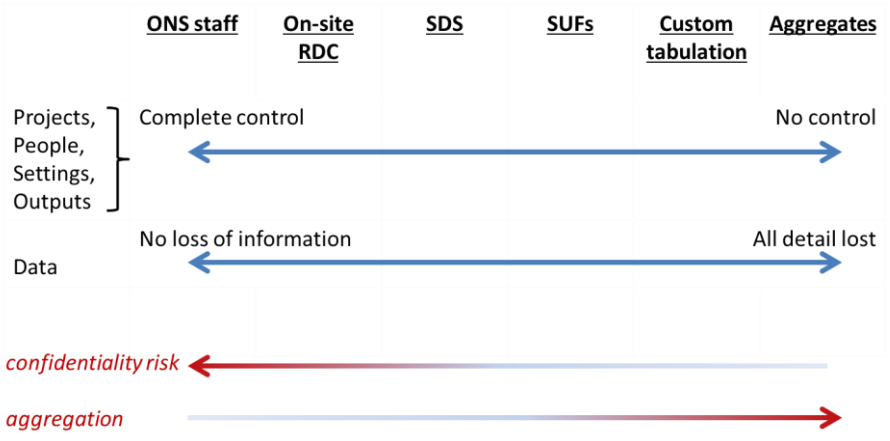


Figure 1 The Data Access Spectrum for ONS data

This two-dimensional representation is appealing as it shows how different access systems relate to each other. It also has value as a planning tool, by identifying gaps in current provision for certain types of data or facility; and it was used as such to provide a rationale for two new facilities in ONS' portfolio (ONS, 2011a).

The DAS should not be confused with two-dimensional diagrams commonly used in the literature to show 'risk-utility' trade-offs. The R-U maps typically mathematise hypothetical risk models and univariate perturbation, and hence have strictly limited practical application other than for the assessment of relative risk change. In contrast, the DAS is a symbolic representation of how a range of user needs is being met. It becomes less clear when considering options such as remote job servers and synthetic data, where the non-statistical elements do not move in step[9], but as an easily comprehensible (and non-specialist) overview of access routes it can be a useful tool for planning and communication.

6. Qualitative data

As noted in section 3.5.2, there are few guidelines on input or output SDC for qualitative data. This is because of the nature of qualitative data, where the narrative structure and the focus on personal experience makes data inherently more identifiable. One option being explored is the use of secure remote settings to view video recordings, for example.

At present qualitative data use is largely managed by non-statistical approaches, but without a strong conceptual framework to justify decisions made. The Five Safes provides a way to formalise this as a managerial rather than a data problem. If anonymisation of qualitative data is very difficult to achieve without destroying the value in the data, can the non-statistical dimensions help? Limiting access to high-benefit projects, training researchers, putting in place secure arrangements for holding and viewing the data, having double-checking of outputs for publication, and so on… All these can provide a way to reduce risk without unacceptably damaging the data of interest, and the Five Safes helps to conceptualise and structure the necessary decision-making. This is likely to be an area of significant development in the future as more unstructured and personal data becomes available for research.

7. Conclusion

A data owner wishing to allow access to data or statistical results whilst ensuring the confidentiality of the sources is faced with a vast amount of advice. Ethical works can provide information on which sort of project should be allowed; cybersecurity literature is filled with advice on how to set up secure systems; the academic and government literature on statistical disclosure control has been repeating much the same thing for fifty years.

None of these directly help because they focus on a specific practical aspect of the problem: ethics, law, IT, agreements, anonymisation, perturbation, and so on. Moreover, few works in these fields directly address the issue of researcher motivation, instead preferring to make simple assumptions that the researcher is inherently good or malicious.

The Five Safes is a way of addressing these matters in a structured way. It encompasses the other fields, and both differentiates them and shows how they are interlinked. That allows, for example, discussions with IT staff or lawyers for example, to be held in the knowledge that the necessary assumptions which frame that discussion are not immutable.

[9] Readers who can draw a useful 'data access plane' are encouraged to do so and let the authors know!

A key element of the Five Safes is that it focuses on the user objectives first: do they want summary stats? Do they want to do to create their own tabulations? Do they want to carry out marginal analysis on small subsamples? Do they really want this, or are they happy with slightly less? How much will they put up with to get that? How much are they willing to engage with us?

A consequence of Five Safes modelling is that statistical questions have no more weight in principle than any other potential controls. In fact, because (1) an extensive literature over forty years has demonstrated ways to anonymise data or outputs to any given level of likely re-identification, and (2) data value is the reason *ne plus ultra* for access, statistical controls become the least important part of the framework.

This "users first; statistics last" approach differs substantially from previous perspectives on data access. As Ritchie (2014a) notes, data access strategies often start with "what are we allowed to do?", rather than "what would we like to do?". The former is driven by the traditional focus on data disclosiveness as the touchstone of data owners. One data owner, responding to a paper entitled "Safe use, not safe data", rejected it with the (paraphrased) comment "I don't like the title. Data must always be safe." Similar comments have been made by heads of NSIs[10]. This is clearly nonsense, as attested by the huge growth in controlled environments in the last ten years giving access to reasonably identifiable confidential data. However, it usefully illustrates the change in perspective necessary to fully exploit the Five Safes framework, in particular the understanding that safe data should be a residual and not the goal.

Such changes in perspective are affected by the order in which the Five Safes are presented. Organisations are staffed by humans, who respond differently depending on whether they have an antipathy to or affinity with the subject: focusing on safe data first may address concerns or it may raise them. Practitioners do not agree on the appropriate order, but the fact that this is a matter of live debate reflects explicit recognition of the need to address psychological and institutional factors in data access.

As well as shifting perspectives into a more productive direction, the Five Safes framework provides a common frame of reference for organisations to resolve differences. This facilitates both strategic and operational planning, and enables discussion across cultural and legal divides, as well as organisational ones.

Finally, the Five Safes approach is explicitly relativistic, subjective and empirical, arising from the need to assess risk in non-comparable categories which are not easily open to quantification. As noted, this does create risks that the 'safest' option is always chosen, leading to less data access. However, the experience of the authors is that the accumulated evidence of several decades of data release can be effectively deployed to provide improved solutions for data owners and researchers alike.

The Five Safes framework does not solve all problems; indeed, it could be argued that solves no problem, as it is simply a structure and an ethos, helping to frame discussion. However, in the authors' experience this framing is crucial to effective strategic planning, operational assessment and evaluation; in cross-organisation discussions, the value of a common frame of reference is hard to overestimate. This helps to explain why the framework has been adopted across organisations, countries and release systems – simply, it has proved useful.

[10] Source: personal communication with the authors.

References

ASA (2009) *Statistical disclosure limitation: balancing data confidentiality and data access*. Accessed 13.8.2014. http://slideplayer.us/slide/709029/

Bailey M. (2012) "Workshop on security at the Cyberborder", *Security at the Cyberborder Workshop Report*, Summaries pp56-60 https://scholarworks.iu.edu/dspace/bitstream/handle/2022/14070/Cyberborder-Report-Final.pdf

Brandt M. (2010) *Decentralised Access to EU-Microdata Sets,* final report to Eurostat on grant agreement no. 61102.2008.001-2008.828, Eurostat, 2010. http://www.cros-portal.eu/sites/default/files//Final%20report%20DA.pdf

Brandt M., Franconi L., Guerke C., Hundepool A., Lucarelli M. , Mol J., Ritchie F., Seri G. and Welpton R. (2010) *Guidelines for the checking of output based on microdata research*, Final report of ESSnet sub-group on output SDC, Eurostat http://neon.vb.cbs.nl/casc/ESSnet/guidelines_on_outputchecking.pdf

Bujnowska A. and Museux J-M. (2012) "The Future of Access to European Confidential Data for Scientific Purposes*", Work session on Statistical Data Confidentiality 2011*, Eurostat. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/43_Eurostat.pdf

Camden M. (2014) "Confidentiality for integrated data" in *Work session on statistical data confidentiality 2013*; Eurostat. http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_3_NZ.pdf

Corti L., van den Eyden V., Bishop L., and Woollard M. (2014) *Managing and sharing research data: a guide to good practice*. London: Sage Publications Ltd

Desai T. (2012) "Maximising Returns to Government Investment in Data", presentation to IASSIST 2012, Vancouver. http://www.iassistdata.org/downloads/2012/2012_f2_desai.pdf

Desai T. and Ritchie F. (2010) "Effective researcher management", in *Work session on statistical data confidentiality 2009*; Eurostat. http://www.unece.org/stats/documents/ece/ces/ge.46/2009/wp.15.e.pdf

Eurostat (2014) *Treatment of Statistical Confidentiality: Introductory module.* Course notes. Eurostat.

Eurostat (2016) *Confidential data: self-study guide for researchers*, forthcoming.

Evans, P. and Ritchie, F. (2009) *UK Company Statistics Reconciliation Project: final report*, Report for the Department of Business Enterprise and Regulatory Reform; URN 09/599

Guardian (2015) "Public bodies are releasing confidential personal data by accident, activists say ". Guardian Newspapers Ltd 15th July. http://www.theguardian.com/technology/2015/jul/15/confidential-personal-data-release-accident-councils-nhs-police-government

Hafner, H.-P. (2008) *Die Qualität der Angriffsdatenbank für die Matchingexperimente mit den Daten des KSE-Panels 1999 – 2002*. Mimeo. IAB

Hafner H-P., Lenz R. and Ritchie F. (2014) *CIS Anonymisation: proposed methodological improvements*. CIS2010 Anonymisation Report, deliverable D2. Eurostat, April.

Hafner H-P., Lenz R. and Ritchie F. (2015) *User-focused threat identification for anonymised microdata*. Department of Economics working Paper 15/03, Bristol Business School. http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202015/1503.pdf

Hawkins M. (2011) *The HMRC data lab,* presentation to KAI International Conference on Taxation Analysis & Research, 2 Dec. http://www.esrc.ac.uk/hmrc/images/Session%2012%20The%20HMRC%20Datalab_tcm19-19599.ppt

Hundepool A., Domingo-Ferrer J., Franconi L., Giessing S., Schulte Nordholt E., Spicer K., de Wolf P-P. (2012) *Statistical Disclosure Control*. Wiley .

Ishazaka A. and Nemery P. (2013) *Multi-criteria decision analysis*. London: Wiley.

Jackson J., Bradford B., Hough M., Myhill A., Quinton P., and Tyler T. (2012) "Why do People Comply with the Law?: Legitimacy and the Influence of Legal Institutions". *British Journal of Criminology* v52 (6): 1051-1071 doi:10.1093/bjc/azs032 http://bjc.oxfordjournals.org/content/52/6/1051.full

Lane J.and Schur C. (2010) "Balancing access to health data and privacy: a review of the issues and approaches for the future". *The Free Library* October, 1; accessed 13.8.2014. http://www.thefreelibrary.com/Balancing%20access%20to%20health%20data%20and%20privacy:%20a%20review%20of%20the%20issues...-a0238476478

Lenz R. (2006) Measuring the disclosure protection of micro aggregated business microdata - an analysis taking as an example the German Structure of Costs Survey*, Journal of Official Statistics* 22 (4), 681-710

McEachern S. (2015) *Implementation of the Trusted Access Model* ASSA Policy Roundtable, November 2015. http://rsss.anu.edu.au/sites/default/files/SMcEachern_researcherperspective_ASSANov2015.pdf

McHale J. and Jones J. (2012) "Privacy, confidentiality and abortion statistics: a question of public interest?" *Journal of Medical Ethics* 2012;38:31-34 doi:10.1136/jme.2010.041186 http://jme.bmj.com/content/38/1/31.short

NRC (2013) *Proposed Revisions to the Common Rule for the Protection of Human Subjects in the Behavioral and Social Sciences: Workshop summary*. National Research Council, Washington, DC: The National Academies Press http://www.nap.edu/catalog.php?record_id=18383

NRC (2014). *Proposed Revisions to the Common Rule for the Protection of Human Subjects in the Behavioral and Social Sciences*. National Research Council, Washington, DC: The National Academies Press. http://www.nap.edu/catalog.php?record_id=18614

Nutt D., Phillips L., Balfour D., Curran H.V., Dockrell M., Foulds J., Fagerstrom K., Letlape K., Milton A., Polosa R., Ramsey J., and Sweanor D. (2014) "Estimating the harms of nicotine-containing products using the MCDA approach". *European Addiction Research*, 20 (5). pp. 218-225. ISSN 1022-6877

OECD (2014) *OECD Expert Group For International Collaboration On Microdata Access: Final Report*. Organization for Economic Co-operation and Development, Paris, July. http://www.oecd.org/std/microdata-access-final-report-OECD-2014.pdf

ONS (2011a) *Data Access Policy.* Mimeo, Office for National Statistics, Newport.

ONS (2011b) *Secure Data Service Risk Assessment.* Mimeo, Office for National Statistics, Newport.

Price D. (2014) *Statistics Canada: Microdata Access for Canadians*. Data Liberation Initiative Training Slides. http://cudo.carleton.ca/system/files/dli_training/3715/statistics-canada-microdata-access-canadiansapril112014.pdf

RCUK (2008) *Response to the ICO consultation on anonymisation.* Research Councils UK. http://www.rcuk.ac.uk/RCUK-prod/assets/documents/submissions/200812RCUKResponseICOconsultationanonymisation.pdf

Ritchie F. (2007) *Statistical disclosure control in a research environment*, mimeo, Office for National Statistics. Republished in WISERD Data Resources WDR/006 http://www.wiserd.ac.uk/index.php/download_file/view/59/248/151/

Ritchie F. (2008) "Disclosure detection in research environments in practice", in *Work session on statistical data confidentiality 2007*; Eurostat; pp399-406 http://epp.eurostat.ec.europa.eu/portal/page/portal/conferences/documents/unece_es_work_session_statistical_data_conf/TOPIC%203-WP.37%20SP%20RITCHIE.PDF

Ritchie F. (2013) "International access to restricted data: A principles-based standards approach". *Stat. J. of the IAOS* v29:4 pp289-300. DOI 10.3233/SJI-130780 http://iospress.metapress.com/content/x772718742503111/?p=d1e9b9e72436490da0efcaaf85bd800a&pi=6

Ritchie F. (2014a) "Access to sensitive data: satisfying objectives, not constraints", *J. Official Stat*., v30:3 pp533-545, September. DOI: 10.2478/jos-2014-0033. http://www.degruyter.com/view/j/jos.2014.30.issue-3/jos-2014-0033/jos-2014-0033.xml

Ritchie F. (2014b) "Operationalising 'safe statistics': the case of linear regression", Department of Economics working Papers 14/10, Bristol Business School. http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202014/1410.pdf

Ritchie F. and Elliot M. (2016) "Principles- versus rules-based output statistical disclosure control in remote access environments", *IASSIST Quarterly*, forthcoming

Ritchie F. and Welpton R. (2011) "Incentive compatibility in secure research facilities". Mimeo. http://www.felixritchie.co.uk/publications/Incentive%20compatibility%20v1.1.ppt

Ritchie F. and Welpton R. (2012) "Sharing risks, sharing benefits: Data as a public good", in *Work session on statistical data confidentiality 2011*; Eurostat http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2011/presentations/21_Ritchie-Welpton.pdf

Ritchie F. Whittard D. and Dawson C. (2014) *Understanding official data sources*. Final report for the Low Pay Commission, February.

Schomisch J. (2014) "National Research Council Proposes Common Rule Revisions", Guide to Good Clinical Practice Newsletter, Thompson, January 9th. http://prod-admin1.tmg.atex.cniweb.net:8080/preview/www/2.3427/2.3465/1.369607

Skinner C. (2012) Statistical Disclosure Risk: Separating Potential and Harm, *Int. Stat. Rev.* v80:3 pp349–368 http://onlinelibrary.wiley.com/doi/10.1111/j.1751-5823.2012.00194.x/abstract

SNZ (2014) *Integrated data infrastructure.* Statistics New Zealand web site. Accessed 13.8.2014. http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx

Spicer K., Tudor C. and Cornish G. (2014) "Intruder Testing: Demonstrating practical evidence of disclosure protection in 2011 UK Census" in *Work session on statistical data confidentiality 2013*; Eurostat http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_5_Spicer.pdf

Sullivan F. (2011) *The Scottish Health Informatics Programme*, presentation to Health Statistics User Group. http://www.rss.org.uk/uploadedfiles/userfiles/files/Frank-Sullivan-linkage.ppt

Sweeney L. (2002) "k-anonymity: a model for protecting privacy". *Int. J. on Uncertainty, Fuzziness and Knowledge-based Systems*, 10:5 pp557-570. http://dataprivacylab.org/dataprivacy/projects/kanonymity/kanonymity.pdf

Trewin D., Andersen A., Beridze T., Biggeri L., Fellegi I., and Toczynski T., *Managing statistical confidentiality and microdata access: Principles and guidelines of good practice*, final report for UNECE/CES, 2007. http://www.unece.org/stats/publications/Managing.statistical.confidentiality.and.microdata.access.pdf

Tubaro P., Cros M., Kleiner B., and Silberman R. (2013) "Access to official data and researcher accreditation: State of the art and future perspectives in Europe", presentation to IASSIST 2013, Cologne. http://iassistdata.org/downloads/2012/2012_d3_tubaro_etal.pdf

Tyler T., Jackson J. and Bradford B. (2013) "Psychology of procedural justice and co-operation" in BruinsmaG. And Weisburd D. (eds), *Encyclopaedia of Criminology and Criminal Justice.* Spring-Verlag

UKDA (2014) *Microdata handling and security: guide to good practice (rev. 05.00).* UK Data Archive, December. http://www.data-archive.ac.uk/media/132701/ukda171-ss-microdatahandling.pdf

Volkow N. (2014) "Standardised application process for microdata access" in OECD (2014), ch.6 pp75-81

Wagener R. (2008) *Microdata and Evaluation of Social Policies,* paper prepared for the colloquium "En route vers Lisbon", Public Policy Research Centre Henri Tudor, December. http://webserver.tudor.lu/cms/lu2020/publishing.nsf/0/FDECF548D12BC30BC12575140048AB73/$file/16h15_Raymond_WAGENER.pdf

Webster A. (2015), "Expanding access to public data: background paper", ASSA Policy Roundtable

Welpton R. (2013) "The UK Data Service: appropriate access", presentation to IASSIST 2013, Cologne. http://iassistdata.org/downloads/2013/2013_c5_welpton.pdf

Welpton R. and Kinder-Kurlanda K. (2013) "Achieving Real Data Security via Community Self-Enforcement", presentation to IASSIST 2013, Cologne.

Welpton R. and Ritchie F. (2012) "Incentive compatibility in data security", presentation to IASSIST 2012, Vancouver http://www.iassistdata.org/downloads/2012/2012_a2_welpton_etal.pdf

Wolters A. (2015) "Researcher management and breaches control", presentation to *The Five Safes of Secure Access to Confidential Data*, Manchester, September. https://ukdataservice.ac.uk/media/604140/14_5safes_safepeople_wolters.pdf

Woollard M. (2009) *Accessing and managing data in a secure environment: the Secure Data Service*. Presentation to Social Survey Research Workshop, Essex, January. http://www.esds.ac.uk/news/eventsdocs/how_to_set_up11nov2010mw.pdf

Wilde K. (2013) "Electronic Technology – Introducing Safe Havens – The Grampian Way!", *Quasar*, April pp28-29 http://www.drdenenelson.com/pubs/Quasar_123.pdf

Recent UWE Economics Papers
See http://www1.uwe.ac.uk/bl/research/bristoleconomicanalysis for a full list.

2016

1601 **Five Safes: designing data access for research**
Tanvi Desai, Felix Ritchie and Richard Welpton

2015

1509 **Debt cycles, instability and fiscal rules: a Godley-Minsky model**
Yannis Dafermos

1508 **Evaluating the FLQ and AFLQ formulae for estimating regional input coefficients: empirical evidence for the province of Córdoba, Argentina**
Anthony T. Flegg, Leonardo J. Mastronardi and Carlos A. Romero

1507 **Effects of preferential trade agreements in the presence of zero trade flows: the cases of China and India**
Rahul Sen, Sadhana Srivastava and Don J Webber

1506 **Using CHARM to adjust for cross-hauling: the case of the Province of Hubei, China**
Anthony T. Flegg, Yongming Huang and Timo Tohmo

1505 **University entrepreneurship education experiences: enhancing the entrepreneurial ecosystems in a UK city-region**
Fumi Kitagawa, Don J. Webber, Anthony Plumridge and Susan Robertson

1504 **Can indeterminacy and self-fulfilling expectations help explain international business cycles?**
Stephen McKnight and Laura Povoledo

1503 **User-focused threat identification for anonymised microdata**
Hans-Peter Hafner, Felix Ritchie and Rainer Lenz

1502 **Reflections on the one-minute paper**
Damian Whittard

1501 **Principles- versus rules-based output statistical disclosure control in remote access environments**
Felix Ritchie and Mark Elliot

2014

1413 **Addressing the human factor in data access: incentive compatibility, legitimacy and cost-effectiveness in public data resources**
Felix Ritchie and Richard Welpton

1412 **Resistance to change in government: risk, inertia and incentives**
Felix Ritchie

1411 **Emigration, remittances and corruption experience of those staying behind**
Artjoms Ivlevs and Roswitha M. King

1410 **Operationalising 'safe statistics': the case of linear regression**
Felix Ritchie

1409 **Is temporary employment a cause or consequence of poor mental health?**
Chris Dawson, Michail Veliziotis, Gail Pacheco and Don J Webber

1408 **Regional productivity in a multi-speed Europe**
Don J. Webber, Min Hua Jen and Eoin O'Leary

1407 **Assimilation of the migrant work ethic**
Chris Dawson, Michail Veliziotis, Benjamin Hopkins

1406 **Empirical evidence on the use of the FLQ formula for regionalizing national input-output tables: the case of the Province of Córdoba, Argentina**
Anthony T. Flegg, Leonardo J. Mastronardi and Carlos A. Romero

1405 **Can the one minute paper breathe life back into the economics lecture?**
Damian Whittard

1404 **The role of social norms in incentivising energy reduction in organisations**
Peter Bradley, Matthew Leach and Shane Fudge

1403 **How do knowledge brokers work? The case of WERS**
Hilary Drew, Felix Ritchie and Anna King

1402 **Happy moves? Assessing the impact of subjective well-being on the emigration decision**
Artjoms Ivlevs

1401 **Communist party membership and bribe paying in transitional economies**
Timothy Hinks and Artjoms Ivlevs

2013

1315 **Global economic crisis and corruption experience: Evidence from transition economies**
Artjoms Ivlevs and Timothy Hinks

1314 **A two-state Markov-switching distinctive conditional variance application for tanker freight returns**
Wessam Abouarghoub, Iris Biefang-Frisancho Mariscal and Peter Howells

1313 **Measuring the level of risk exposure in tanker shipping freight markets**
Wessam Abouarghoub and Iris Biefang-Frisancho Mariscal

1312 **Modelling the sectoral allocation of labour in open economy models**
Laura Povoledo

1311 **The US Fed and the Bank of England: ownership, structure and 'independence'**
Peter Howells

1310 **Cross-hauling and regional input-output tables: the case of the province of Hubei, China**
Anthony T. Flegg, Yongming Huang and Timo Tohmo

1309 **Temporary employment, job satisfaction and subjective well-being**
Chris Dawson and Michail Veliziotis

1308 **Risk taking and monetary policy before the crisis: the case of Germany**
Iris Biefang-Frisancho Mariscal

1307 **What determines students' choices of elective modules?**
Mary R Hedges, Gail A Pacheco and Don J Webber

1306 **How should economics curricula be evaluated?**
Andrew Mearman