The 'Five Safes': a framework for planning, designing and evaluating data access solutions

Felix Ritchie

University of the West of England, Bristol and Administrative Data Service
felix.ritchie@uwe.ac.uk

Abstract

The 'Five Safes' is a popular way to structure thinking about data access solutions. Originally used mainly by statistical agencies and social science academics, in recent years it has been adopted more widely across government, health organisations and private sector bodies.

This paper explains the Five Safes, how the concept is used to organise and simplify decision-making, and how it helps to address concerns of different constituencies. We show how it aligns to recent regulation, anticipating the shift towards multi-dimensional data management strategies. We provide a number of practical examples as case studies for further information.

We also briefly consider what issues the Five Safes does not address, and how the framework sits within a wider body of work on data access which challenges traditional data access models.

Keywords – data access; confidentiality; Five Safes

1 Introduction

Decision-makers taking responsibility for data access face a complex problem where multiple factors need to be taken into consideration and balanced: ethical questions over access, technical and statistical protection measures, and use considerations. These factors interact: if less statistical protection is provided in the data, does this overly limit use-value? If licensing restrictions on how the data are stored allow more detail to be provided, how confident are we that the licences will be followed and that users will not make mistakes in data handling?

The problem is how to deal with these competing issues in a structured way that allows all factors to be discussed but without requiring that everything must be settled at the same time or in particular order. In the author's experience, it is common (and natural) for different constituencies to try to drive discussion towards their particular area of expertise: statisticians to focus on data protection, IT professionals to want to specify the environment first, data collectors worry who will use the data and for what purpose.

The Five Safes was developed to address this, creating a framework which allows

- Strategic discussion about the broad direction of a data access solution
- Specific discussion about factors without needing to address all factors simultaneously
- Transparency in the process of discussion

The Five Safes was originally developed at the UK Office for National Statistics in 2003. It was adopted by other statistical agencies, and social science academic organisation such as data archives in the UK, Germany and the US. Over time other government departments in the UK and abroad came to use it, as well as international bodies, and in recent years it has made its way into health research and legislation. In the absence of any other similar framework, it has become the de facto standard for addressing data access solutions, particularly in social science.

The rest of this paper is structured as follows. The next section describes how the Five Safes framework operates from a conceptual perspective. We then discuss (briefly) how it has been used in practice; this is done to shed light on the conceptual set-up, but is also intended highlight that there is a significant community of support and common interest, particularly in the UK. Finally we consider what problems the Five Safes does not solve, leading on to the other papers in this group. We conclude with pointers to further reading and resources.

For simplicity, we assume that we are considering access to data for research purposes. The Five Safes framework has also been used for operational data management, but the emphasis placed on control factors differs in this case.

2 The Five Safes framework defined

2.1 Key concept: joint but several control dimensions

The basic premise of the framework is that data access can be seen as a set of five 'control dimensions': safe projects, safe people, safe data, safe settings, safe outputs[1]. Each of these dimensions provokes a question about data access:

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

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:

- The 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

The several nature of this can be shown by considering the first dimension, 'safe purpose'. In answering the question "is this use of the data appropriate?", a natural response is "that depends on who uses it for what purpose". The Five Safes rejects this: the question should be addressed **making the working assumption** that appropriate facilities, users and training can be provided to support whatever decision is made about the purpose of the access. Discussion therefore can focus on the ethical/legal/logistical basis for access, and what controls are needed to facilitate that.

Clearly at some point the different dimensions need to be considered together; this is the joint nature of the framework. If you decide that data of type A is only available to researchers of type X whereas data of type B should be accessible to everyone, then the overall solution needs to make sure that the 'people'/'data' dimensions reflect these decisions being taken in the 'purpose'

dimension, which were **provisional** on assumptions made about other dimensions.

2.2 'Safety' as a scale, not a switch

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 data holder has some idea of 'more safe data' and 'less safe data' and can make relative judgments.

Overall, the aim is a 'safe' data access solution, but it is not necessary that all the control dimensions be set to 'maximum safety'. Consider the analogy of having a 'good meal' out, as measured on price, quality, ambience, and speed of delivery. A Michelin-starred restaurant, a good family restaurant, and a greasy spoon will differ greatly in price, quality and so on, but can all provide a 'good' experience. Similarly, statistics institutes will take one data source and create multiple data products available to a range of users with different restrictions on where, how and by whom the data can be used. These meet differing user needs and tolerances, and have vested their controls in different elements of the Five Safes, but all of these are 'safe' at the system level.

Note that this also provides a way of reviewing and comparing data access solutions where an organisation does make data available in many different ways. The UK, Australian and Greek statistics offices have formally described their range of outputs in this way.

2.3 Order and priority of control dimensions

The Five Safes does not give a higher preference to one or other dimension; however, the listing of dimensions above does reflect the natural order of typical decision-making process. Data access planning starts with identification of user needs, giving broad groupings such as "users of type X will be prepared to put up with a strict application procedure if they get detailed data, whereas users of type Y just want simple stats but expect them to be freely available". This leads to thinking of an overall solution in terms of people and settings, so that the broad shape of a system solution can be defined[2]. Implementation can then move on to specifics in each dimension.

In the EDRU model (see below), best practice now treats the 'statistical' controls as residuals; that is, these are

---

[1] 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, such as public use files or non-confidential source data. Some users (such as the OECD Microdata Expert Group) have used 'secure' rather than 'safe'.

[2] For a specific example of this working in practice, see the report for the Australian Department of Social Services (Green and Ritchie, 2016)

addressed once a satisfactory solution in terms of people, purpose and setting have been well defined. This arises from the EDRU perspective which treats user values as the outcome objective and confidentiality restrictions as constraints; in this view, the whole purpose of making data available is to support user needs, and restricting detail in the data is only done when better solutions are not available.

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'. The difference between "this control is not relevant" and "this control is conceptually relevant but not applicable here" is that the second formulation encourages transparency about where the data holder is accepting operational limitations.

2.3   Alignment with regulation

When the five safes was originally developed, data access regulation was framed around questions of whether data was 'identifiable' or not; this created a dominant role for statistical anonymisation in data access solutions in ensuring that confidentiality was not breached. This dominance was reflected in the academic and government literature, and in the practices of statistical agencies.

Recent years have seen the increasing availability of research data in a range of environments, better use of evidence about the risks of research access, and an awareness of lost public benefit from over-protected data. These have together led regulators to explicitly consider non-statistical controls on data as providing the necessary protection. Modern regulations and legislation increasing reflect the Five Safes model of statistical and operational controls.

For example, the UK *Digital Economy Act 2017*[3] explicitly states that assessing the lawfulness of disclosure of data for research purposes should take account of the purpose (subsection 64(8)), the setting (64(6)), and the people (various clauses, ss 64-66), and creates a legal role for accreditation of these. The Act also provides a gateway for output-checking processes (66(3)b). This conformity with the Five Safes is not surprising, as the research provisions were strongly influenced by ONS, which uses the Five Safes in its planning.

The European *General Data Protection Regulation 2016* (GDPR)[4] explicitly validates non-statistical measures of data protection. Articles 5, 25 and 38 repeatedly state that data must be stored and processed "using appropriate

technical or organisational measures". 'Pseudonymisation' (reducing identifiability) is one potential solution, but the GDPR is clear that identification is to be assessed in terms of the operational controls on data. In this, it differs substantially from the previous Regulation which only referenced de-identification.

In one case, the Five Safes is explicitly used as the basis for regulation. The *South Australia Public Sector (Data Sharing) Act 2016*[5] prescribes the Five Safes as 'Trusted Principles' (section 7) which all data access agreements should adhere to; the Australian Productivity Commission has recommended the wider adoption across federal government.

3   The Five Safes in practice: examples of use and users

3.1   Description and evaluation

The Five Safes was originally developed to describe ONS' secure research facility in 2003. Since then it has been widely used to describe individual data access systems and portfolios. Examples include governments in the UK, Mexico, Australia, New Zealand, France, and Germany. It has also been used to describe the activity of academic data archives and access facilities in the UK, US[6], South Africa, Japan, and several European countries. The Scottish Health Informatics Project adopted the framework in 2010 to bring clarity to descriptions of an extremely complex government-academic research centre.

The framework has also lends itself well to evaluation and review. For example, the formal risk assessment carried out on the UK Data Archive's Secure Data Service in 2011 by ONS built the assessment criteria from the Five Safes. More recently, Cancer Research UK has begun to use the framework for its data governance model.

3.2   Design

The Five Safes needed to be well-established as a standard terminology before its use in design was uncontroversial, but in recent years it has moved from passive descriptor to active planning tool. In one of the earliest uses, the Five Safes was embedded into the planning for the UK Administrative Data Research Network in 2012. The OECD Microdata Expert Group used the Five Safes to propose new inter-governmental data sharing processes. In the US, HIPAA revised guidelines on health data sharing reflect the Five Safes. In 2016 the most extensive design uses came with both the Australian Bureau of Social

---

[3] http://www.legislation.gov.uk/ukpga/2017/30/part/5/chapter/5/enacted

[4] http://www.eugdpr.org

[5]
https://www.legislation.sa.gov.au/lz/v/a/2016/public%20sector%20(data%20sharing)%20act%202016_61/2016.61.un.pdf

[6] In the US, it is often referred to as the 'portfolio model'; as early adopters in 2005, the US model omits 'safe data'.

Service and the Hellenic Statistical Authority using the Five Safes to structure and design complete organisation-wide data access strategies.

3.3 Training

Whether or not the Five Safes is used by the organisation for design, it has proved an important tool for training users of confidential data, as it makes for a clear course structure. It is currently used as part of the training materials for organisations as diverse as HMRC, Cancer Research UK, Eurostat, The Health Foundation, the UK Data Archive, NHS Scotland and Statistics New Zealand.

4 Five Safes in the wider data access context

The Five Safes is a way of organising, streamlining and clarifying decision-making processes, but it does not define how to make those decisions. For example, it does not state what sort of research proposals the organisation should accept, or what is a 'safe' level of data for an environment. Historically, government decisions over data access have been driven by 'defensive' strategies, designed to minimise organisational risk rather than maximise public benefit (see Ritchie, 2016, for a discussion). This has been supported by an academic literature which emphasises hypothetical risk and worst-case scenarios, with little reference to empirical evidence on risks or to cost/utility implications.

However, over the last decade or so there has been a movement toward re-thinking the traditional model. This has been driven by greater user demands, technical developments in access, analysis of evidence on breaches, and a re-assessment of the subjectivity of risk evaluation. This has coalesced in recent years into an approach sometimes referred to as the EDRU model of data access:

- **evidence-based**: focusing on the large body of empirical evidence rather than basing policy on hypothetical outcomes
- **default-open**: assuming that data is to be made available if at all possible; maintaining confidentiality is seen as a constraint, not an objective
- **risk-managed**: acknowledging that all risk assessments are subjective, and zero-risk options are not feasible; risk avoidance is downgraded in favour of risk management
- **user-centred**: defining solutions from the perspective of the user, rather than the data-holder

This approach has developed over the same period as the Five Safes, and by many of the same people and organisations. The EDRU approach, despite offering better security and user outcomes at lower costs, has been adopted more slowly, partly because it requires a shift in the organisation's attitude. In contrast, the Five Safes can be used by any organisation irrespective of its position on risk, users and evidence.

Nevertheless , there are substantial gains from integrating the two approaches. This is because the EDRU approach can pose difficult or uncomfortable questions for an institution; the Five Safes provides a framework for organising responses to those questions, particularly in relation to assessing evidence and risk. All organisations applying the EDRU model in full or part currently adopt the Five Safes as the descriptive/planning framework, as does the EDRU literature.

5 Summary

The Five Safes is a practical, proven, adaptable way of focusing thinking about data access. It has been used to describe, design and evaluate data access solutions, as a well as being a popular training tool for courses about data confidentiality. From its initial conception in statistical offices, it has spread throughout government and academia, and across to world. In the absence of any other similarly adaptable framework, it has become the standard model for discussing data access in the social sciences; increasingly it is making inroads into data management for health research.

The Five Safes helps to manage and structure discussion about data access, but it doesn't address all the issues. It does not specify an organisation's attitude to risk, or how it sees its relationship with data users, for example. As noted above, the EDRU approach can help an organisation decide what its attitude to data access should be, and is consistent with the Five Safes framework. The remaining presentations in this session will discuss these issues, use practical examples to illustrate the importance of evidence-based structured approaches to data access questions.

Resources

A more detailed discussion of the Five Safes theory and its practical use can be found in Desai et al (2016). A detailed practical example produced for a government body of the use of the Five Safes in designing an data access portfolio is Green and Ritchie (2016).

From January 2018, the website http://fivesafes.org/ will provide a collection of resources and documents for developers.

References

Desai T., Ritchie F., and Welpton R. (2016) The Five Safes: designing data access for research. Working papers in Economics no. 1601, University of the West of England, Bristol. January. http://eprints.uwe.ac.uk/28124/1/1601.pdf

Green E. and Ritchie F. (2016) *Data Access Project Final Report*. Australian Department of Social Services. http://eprints.uwe.ac.uk/31874/1/Data%20Access%20Project%20Final%20Report%20v2.00%20Final%20DSS.pdf

Ritchie F. (2016) "Can a change in attitudes improve effective access to administrative data for research?", Working papers in economics no. 1607. University of the West of England, Bristol. http://eprints.uwe.ac.uk/29646/1/1607.pdf