Lessons learned in training ‘safe users’ of confidential data

Felix Ritchie*, Elizabeth Green*, John Newman** and Talei Parker**
* University of the West of England, Bristol
** Australian Bureau of Statistics

Abstract: Many statistical organisations require researchers using detailed sensitive data to undergo ‘safe researcher’ training. Such training has traditionally reflected the ‘policing’ model of data protection. This mirrors the defensive stance often adopted by data providers, which shifts the responsibility of failure onto the user, and which derives its behavioural assumptions from the neoclassical economic models of crime.

In recent years, there has been recognition that this approach is not well-suited in addressing the two most common risks to confidentiality: mistakes, and avoidance of inconvenient regulation. Moreover, it is hard to exploit the benefits of user engagement under the policing model, which encourages ‘them and us’ thinking. Finally, there is little evidence to suggest that students absorb “do/don’t” messages well.

There is a growing acceptance that a ‘community’ model of data protection brings a range of benefits, and that training is an investment in developing that community. This requires a different approach to training, focusing more on attitudinal shifts and less on right/wrong dichotomies.

This paper summarises recent learning about training users of confidential data: what they can learn, what they don’t learn, and how to extract the full benefit from training for both parties. We also explore how, in the community model, trainers and data owners also need to be trained as well researchers.

The paper focuses on face-to-face training, but also considers lessons for other training environments. We illustrate with an example of the conceptual design of a new training course being developed for the UK Office for National Statistics.

Acknowledgements: This paper was developed through extensive discussion with trainers and participants on training courses in the UK, Australia and Europe. All opinions expressed in the paper are those of the authors and do not represent the views of any organisation.

1 Introduction

The last two decades have seen great growth in the availability of microdata to researchers. For government data, the biggest growth has been in access to the most sensitive data (‘secure use files’ or SecUFs) which requires users to work in a restricted environment.

Where non-public data files are distributed to researchers (‘scientific use files’, or SUFs), users are typically given little or no training on safe use: some data providers send out information with the data files, others refer researchers to information on the web or in the application packs. This reliance on passive dissemination of good practice is sensible in the face of the costs and benefits. Requiring active training for a large number of users, geographically dispersed, may be prohibitively expensive, while there is very little evidence to suggest that users of SUFs pose a substantive risk to confidentiality.

In contrast, users getting access to the most sensitive data in SecUFs are likely to face at least some active training. SecUFs require a greater commitment on the part of the researcher to work in a restricted environment. This limits numbers, and also makes it easier for the data provider to require participation in training as part of the commitment to the research. Most organisations now require researchers using SecUFs to undergo some form of face-to-face 'safe researcher' training.

Training via passive media such as handouts or web pages has, with a few exceptions (eg Eurostat 2016; Sax Institute, 2016) traditionally reflected the 'policing' model of data protection: users are told "*be grateful, be careful, or you'll go to jail/be fined/lose your job"*. A defensive stance is often adopted by data providers, and shifts the responsibility for failure onto the user, and which derives its behavioural assumptions from the neoclassical economic models of crime. The training is primarily seen as a way of making sure that researchers obey rules.

In recent years, there has been recognition that this approach has two fundamental flaws. First, it is not well-suited to addressing the two most common risks to confidentiality: mistakes, and avoidance of inconvenient (to the researcher) regulation. Second, researchers resent the inference that they are fundamentally untrustworthy – thus engendering the behaviour that the training is designed to stop. There is also little evidence to suggest that students absorb "do/don't" messages well. The policing model encourages a 'them and us' mentality which discourages co-operation and can also act as a barrier for communication.

Instead, there is a growing acceptance that a 'community' model brings a range of benefits to all members of the research community. This requires training which focuses more on attitudinal shifts and less on right/wrong dichotomies. Training then becomes an investment in developing the community and allows the data provider to exploit the benefits of user engagement more effectively than under a policing model.

This paper summarises recent learning about training users of confidential data: what they do and don't learn, how to make learning effective, and how to extract the full benefit from training. We also explore how, in the community model, trainers and data owners need to be trained as well.

We focus on face-to-face training, with brief consideration of lessons learned from other training environments. For simplicity we will assume that the training relates to SecUF users accessing a research data centre (RDC), an environment which allows the researcher almost complete freedom to analyse the data, but limited opportunities to upload or download data. RDCs, particularly virtual RDCs dominate the environment for accessing SecUFs.

We illustrate with an example of a new training course being developed for the UK Office for National Statistics, intended for both external researchers and internal staff.

The next section describes briefly the empirical and theoretical models that have been used to frame questions of researcher training. Section 3 considers experiences from designing and teaching face-to-face courses (and, to a lesser extent, passive media), both our own and from observing others. Section 4 proposes 'good principles' for training, and Section 5 outlines how these are being used to develop a new face-to-face training course for the UK Office for National Statistics. Section 6 concludes.

2 Perspectives on the function of training

There are four relevant views on the purpose of training: public choice theory, which underlies the policing model; models based on community identity; the empirical, atheoretical approach adopted in many RDCs; and the WMI model, which synthesises the theoretical and empirical approaches.

2.1 The policing model

Historically, researcher training was based upon the policing model: show the researchers what to do, show them the consequences of not obeying the rules, and, if necessary, help them draw the obvious inference that following the rules is the way to avoid a whole heap of trouble.

An implicit outcome of this, is the transfer of responsibility from the trainer to the trainee. The trainee has been given the information necessary to act correctly, so if something goes wrong it’s now the trainee’s fault for ignoring the training.

The driving theoretical force behind this approach is ‘choice theory’ from neoclassical economics, the central paradigm in economics (Ritchie, 2015). Choice theory assumes a broadly rational individual who uses information about his or her environment to make informed choices; those choices include whether or not to follow instructions, keep to agreements or even commit crimes. Thus researchers will stay on the straight and narrow simply because the expected value of following orders exceeds the expected value of inappropriate action. In other words, make the risk of being caught and the subsequent penalties high enough to encourage compliance.

One result derived from this model is that researchers are fundamentally untrustworthy. The rational individual will only take actions that will not benefit him or her. If users need to be instructed in safe behaviour, this implies that the untrained individual would choose to take different actions which are more beneficial. Ergo, the purpose of training is to convince an untrustworthy individual to act appropriately.

This training model reflected the prevailing ‘default-closed’ perspective on data access described in Hafner et al (2015) and Ritchie (2016). When the main threat to data access is considered to be an intruder bent on malicious activity, the response is to demonstrate that the intruder’s assumption of successful enterprise is wrong.

The policing remains the dominant model of researcher training, particularly where training is given in the form of user guides and other passive learning tools.

2.2 Procedural justice and behavioural models

While policing models dominate confidentiality theory, the police themselves have explored more behavioural models of offender management. Beginning in the US in the 1990s, the concept of ‘organisational justice’ or ‘procedural justice’ has become an important part of police strategy [ref Jackson et al].

This builds on well-established psychological theory of Social Identity about how individuals respond to being part of a community. If individuals share common experiences, beliefs and values, these individuals are clustered with a shared identity group (Stainton Rogers, 2011). If the group feels they have low status with no power or control, this will ultimately create a dynamic situation where group members will actively seek to change and empower their group status (Tajfel and Turner, 1979). If a group feels able to mobilise their group status then they will use creative and often legal means of inspiring people to value their identity (for example through music, art, communication etc); however if the group does not feel they are able to mobilise their group status, the group will negatively compete against other groups (Turner and Reynolds, 2010).

The model would therefore argue that individual researchers are more likely to co-operate and engage with the data owners (or other administering force) if they believe (a) that they share a same community purpose, and (b) that the administrators will act fairly and honestly in pursuit of that common purpose. Correct behaviour becomes an outcome of building an affinity between the researchers and administrators such that the researchers become self-policing.

In respect of training courses, a procedural justice perspective would argue that courses which seek to police and utilise ‘scare tactics’ risk creating a dynamic where researchers feel undervalued and powerless; this would be expected to result in social competition which could lead to negative behaviours which actively contradict the trainers identity group. Put simply, researchers will not value (and may actively try to break the system) if they feel disrespected or unheard.

Note that the procedural justice model directly contradicts the predictions and policy implications of public choice theory. The latter suggests that sharing responsibility for data security through community engagement lessens the credibility of any punishment strategy, and so increases the chance of misbehaviour.

One obvious problem with procedural justice as a tool for managing offences is that it is not designed to tackle direct, deliberate and serious criminal activities. Similarly, in data confidentiality, a procedural justice model would not be expected to address the deliberate ‘intruders’ of the statistical literature.

2.3 An empirical approach: active researcher management

‘Active researcher management’ was an empirical, atheoretical approach adopted by the Virtual Microdata Laboratory (VML) team at the UK Office for National Statistics (ONS) from 2004 onwards. As noted in the introduction, the VML team initially adopted the policing approach, but this was abandoned fairly quickly. The VML team met, supervised, worked with and had coffee with researchers on a regular basis post-training; they were therefore able to get continuous feedback on the effectiveness of training, both verbally and by observing how they acted when using the VML.

During the first year the VML training evolved rapidly in response to these factors. For example, an early, serious, breach of procedure allowed the VML team to analyse both what failure of training had led to the breach, and how the presentation of that breach to new researchers affected their response to it. Also, the development of 'principles-based output statistical disclosure control' (PBOSDC; Ritchie and Elliott, 2015), allowed researchers and the VML team to have a common perspective on what sorts of output can be published.

Over time, VML practices coalesced into the strategy now known as 'active researcher management' (ARM: Desai and Ritchie, 2010; Wolters, 2015)[1]. ARM seeks to build a coalition of interests between the researcher, data owner, and support staff, so that data access solutions are recognised by all as fundamentally co-operative. ARM aims to make cooperation explicit by (a) paying attention to the language used between parties, (b) acknowledging the subjectivity of decision-making, (c) sharing understanding of the interests and constraints of all parties, and (d) requiring constant review. Wolters (2015) provides a detailed case study of ARM in practice.

ARM, and related developments such as PBOSDC, appeared to generate substantial efficiency gains. Perhaps as a result, this approach was being adopted or adapted by other organisations as early as 2005. In 2009 a Eurostat Expert Group recommended this model as best practice (Brandt et al, 2010). The rate of development slowed after the initial rush, and by around 2008-2009 the model was fairly stable with core principles established. VML development stopped in 2011 but the UK Data Archive's Secure Data Service continued to fine-tune the model until 2014, when the cross-organisation SURE training began to be developed [is there a reference?]

2.4 A recent synthesis: the WMI model

The popularity of the policing model was influenced by the domination of 'intruder' models of data security, as well as by defensive decision-making amongst data owners (Ritchie, 2016); both of which are strongly tied to the assumption of untrustworthy researchers. Research in the last decade into user behaviour has focused on the evidence base for such behaviour, and has largely debunked the assumptions underlying such attitudes (Ritchie and Welpton, 2014; Hafner et al, 2015; DSS, 2016). The developing paradigm amongst data access practitioners is WMI: the 'well-meaning idiots'[2]. This assumes

- Researchers' motivation is 'intrinsic' (that is, primarily driven by internal psychic needs rather than external motivations such as money or threats)
- Researchers genuinely want to do the right thing
- They will do the wrong thing if it's much easier than the right thing, or if it's the 'normal' thing
- Everyone (researchers and support team) will make a mistake at some point

[1] The name was coined in 2009, but others had independently developed some of these principles; Desai and Ritchie (2010) however seems to be the first article to systematise and describe this approach.

[2] In Eurostat (2015) and ABS (2016), this is referred to as the 'human' model, 'idiot' not being a government-approved description of users. The authors freely admit to being part of the WMI community.

- Humans find it easier to rationalise mistakes than to admit to making them (‘fundamental attribution error’)

These assumptions, unlike the intruder/policing model assumptions, have solid, if limited, empirical support to back them up (Ritchie and Welpton, 2016).

The WMI model provides the bridge between the empirical and theoretical approaches. ARM and procedural justice have many similar features, including the fundamental importance of a shared community purpose. As noted above, procedural justice models are not designed to deter ‘hardened criminals’ – or committed ‘intruders’ in the language of the statistical literature. However, in the world of WMI, procedural justice is exactly the right approach as it seeks to educate, inform, and learn from mistakes rather than punishing wrongdoing. These are also the empirical messages being pushed by ARM which has an implicit assumption that all misdeeds are the result of a bad attitude or lack of process knowledge.

This synthesis of ideas is reflected in the most recent training programmes (Eurostat, 2014, 2015; Wolters, 2015; ABS, 2016; Sax Institute, 2016; Cancer Research UK, 2016).

3 What and how do attendees learn?

In our experience, from our own courses and from observing others, there are a number of common factors that need to be addressed in any training programme:

Prior to the session beginning, researchers…

1. Are usually there because they have to be
    - In other words, researchers are likely to begin the course in a mental state ranging from resigned acceptance to outright hostility to participation
2. Are usually expecting a course that tells them ‘how to’ get access
    - Researchers tend to think of procedures (e.g. ethical approval, disclosure checks) as barriers rather than facilitators. They typically come wanting to learn specifics for how to get over those barriers. This poses problems for the community approach, where the learning objectives are likely to be much less specific than under a policing model. In the community model, for example, the presentation of stages such as output checking or ethical approval as positive steps on the research process can take some re-adjustment on the part of the researcher
3. Are unlikely to have thought about the perspective of the data provider or the support team
    - Again, this can pose challenges for the community-based course, as the researchers do not expect to have to think about the motivations of other organisations. An additional complication is credibility: researchers can be suspicious as to whether statistical organisations can really ‘walk the walk’ of the talk being talked by the trainer.

During the training, researchers…

4. Are interested in relevant procedures
   - That is, researchers are willing to get involved in minutae (for example, logon sequences) if they see a direct relevance to their work
5. Focus more on personal penalties rather than legal ones
   - In other words, researchers care about loss of access, loss of funding, and the opinions of colleagues. Jail and/or fines are not seen as highly credible. Researchers from international organisation quickly point out the data owner would have great difficulty prosecuting across borders; but they knew the non-legal consequences were devastating (loss of access, reputational impacts, etc.)
   - This also relates to the intrinsic motivation of researchers: stopping them carrying out their vocation forces them to re-evaluate their life priorities. Extrinsic legal threats do not have the same motivational force.
6. Are able to hold complex discussions on nuanced topics
   - For example, asking researchers whether they can access an RDC from work, home, McDonalds, the airport, libraries, and hotels will readily lead to a consensus of "yes", "no" or "maybe" for these and what conditions each place might have. One researcher may say it is fine to access the RDC from home, another will come up with the caveat that it shouldn't be done in the presence of flatmates/family/etc., to which the first researcher then invariably agrees, and the discussion progresses. It is possible that giving the researcher a list of do’s and don’ts could have instilled the same knowledge, but it does seem unlikely
7. Don’t like being seen as ‘untrustworthy’ or troublemakers
   - Most researchers react very badly when asked to consider themselves as ‘intruders’, or other negative terms, particularly if there is a suggestion that a whole class of people (i.e. researchers) is untrustworthy. If a person’s feelings of competence and self-determination are enhanced, his intrinsic motivation will increase; if diminished, intrinsic motivation will decrease (Deci, 1975)

After the course, researchers…

8. Aren’t interested in and don’t retain detail of laws
   - Courses based on the policing model emphasise the legal framework governing access. There is little evidence that this information is retained outside the course, and it reinforces the idea that the trainers view researchers as potential lawbreakers
9. Don’t absorb a lot of technical detail

- For example, in one course we observe statistical disclosure control is taught as an exhaustive and exhausting series of examples; but the typical comment from researchers when quizzed about it afterwards is that it was just ‘sudoku’

In summary researchers

- Want to know the minimum they need to learn
- Are prepared to be entertained if there’s a point
- Won’t retain detail
- Have almost certainly not considered wider issues
- Understand principles, but still want a solid foundation for those principles: a ‘safety net’ of things they can think “I know that”

In other words, much like everyone else on a practitioner training course.

4 Good principles in researcher training

The ARM-procedural synthesis, allied to knowledge of what works in practice, suggests that there are a number of common principles which should govern researcher training. Most of these apply to both face-to-face and passive learning models, although the former is more likely to achieve these goals.

There are three overarching principles:

1. Understanding of and engagement with processes is more important than any technical knowledge
2. Facilitated discussion and practical exercises are the most effective way to achieve the attitudinal shifts necessary for a community approach
3. Trainees should be able to relate their learning directly to their own (expected) experience

The first follows directly from the aims of the ARM-procedural model: it is better to have researchers who have less knowledge but are willing to work with the support team, then to have technical geniuses who see no value in the role of the support team. The second comes from practical experience, and well-established models in education and psychology. The third also follows directly from practical experience: researchers learn more from concrete examples than abstract ones.

On this basis, specific recommendations are:

Principle	**Rationale**
Minimise the amount of legal information, by moving from the detail of laws to a discussion of principles	Researchers aren’t interested in legal detail, and therefore retention rate is very low. It also emphasises the ‘them and us’ aspect. The law applies whether researchers understand it or not. Principles are simple and universal: all laws say who, what, why and for how long you can have access to data;

	for any change to that, ask the relevant professional. This is a message that researchers easily follow Researchers are intrinsically motivated, identifying their job as a vocation. Therefore using materialistic threats seldom works: training should focus on intrinsic threats, such as reputation or access
Make researchers understand that if they don't follow procedures, they may be breaking the law.	Emphasise the role of the data access procedures as protective, not prescriptive. For example, ethical committees are your friends who help you avoid making mistakes – and who will be on your side if things go wrong (and you have done what you said you would do)
Build a sense of community	Being part of a community encourages behaviour that accepts the norms of the community, even when those norms are notionally annoying to the researcher
Encourage positive challenge to the community	If researchers have a genuine role in the community, that means they should be able to influence how it develops
Encourage self-policing	This is done through two mechanisms: establishing the legitimacy of the access regime; and highlighting self-interest ("you don't want your dodgy colleague ruining your access!"). The former should take prominence as the positive message
Don't discuss policing, or self-policing	Both terms emphasise criminality
Use discussion instead of statement wherever possible	Discussion encourages ownership of ideas, and allows researchers to explore ideas in ways that are relevant to them, rather than the way the trainer may have intended an exercise to run
Examples should be illustrative, not exhaustive	Researchers' tolerance for extended technical examples is low, and excessive repetition undermines goodwill; researchers understand that further information can be made available post-training
Examples should demonstrate positive aspects of engagement	Negative examples should always be balanced by a positive reflection on the outcome
The trainer-researcher relationship needs to be clear	If trainers are not researchers, then (a) training materials need to reflect what the trainer can teach, and (b) the trainer needs to become much more of a facilitator rather than a leader of ideas. Even more importantly, an appropriate respect for the skills of the other parties needs to be established at an early stage. However good the rest of the course is, its impact will be diminished if the

	researcher has no respect for the trainer, or the trainer patronises the researchers.
Trainers need to be trained in depth	This approach requires a high level of commitment to the ideals of the community model; otherwise, the trainer may struggle to see the coherence behind the various messages being put forward.

5 Developing a new national training programme

We complete this analysis by reviewing a new training programme being developed by the UK Office for National Statistics for its RDC, and in which the authors of this paper are involved at various levels.

The new course, provisionally called Safe Researcher Training (SRT), tries to follow the above precepts. Accordingly, the course has the following features:

- A strong focus on community – both the relationship between different members of the community and the personal attitudes of the researcher
- Largely exercise-based teaching, with non-exercise slides being used to either introduce a concept or summarise it
- Discussion of legal frameworks focusing on principles
- Emphasis on the positive, protective role of procedures such as ethics, project approval or output checking
- Emphasis on the role researchers played, and are expected to continue to play, in the design of the RDC and its procedures
- Testing based upon attitudes rather than memory tests

In many respects, the focus on attitudes means that the training has more in common with a behavioural intervention than a standard training programme.

This course is currently in development and is intended to be in its final form in autumn 2017, when an updated version of this paper will be made available.

6 Conclusion

After many years of practical experience, there is now a coherent conceptual and empirical framework for training researchers in the safe use of confidential data. This new model contrasts sharply with the typical policing model in its focus on behavioural factors rather than abstract assumptions about the human condition.

This approach does require more commitment on the part of the trainer. It requires a thorough understanding of the underlying principles behind the community approach; a level of skill in facilitating discussion which may not come easily to some trainers; and mutual respect from all parties in respect of each other's skills.

The policing model provides a much simpler way of dealing with training, and there is no reason why the policing approach cannot incorporate exercise and class activities (as some do). However the fundamental problem remains: policing is driven by a fear of failure rather than a positive aim of engagement, and this ultimately is communicated to researchers.

This paper has focused on face-to-face training. Building a behavioural model around a passive medium such as handouts or web presentation is considerably harder, and rarer. Nevertheless, it can be done (see Eurostat, 21015; Sax Institute, 2016). More importantly, the lessons around what researchers actually learn, rather than what course designers would like them to learn, are at least partly transferable across media.

References

ABS (2016) *Datalab training programme*.

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

Deci, E. (1975) *Intrinsic Motivation.* New York: Plenum Press.

Desai T. and Ritchie F. (2010) "Effective researcher management", in *Work session on statistical data confidentiality 2009*; Eurostat; forthcoming

Eurostat (2014) *Treatment of Statistical Confidentiality*. European Commission training course 2014-2016.

Eurostat (2016) *Self-study material for the users of Eurostat microdata sets*

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

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.

Ritchie F. and Elliot M. (2015) "Principles- versus rules-based output statistical disclosure control in remote access environments", *IASSIST Quarterly* v39 pp5-13

Ritchie F. and Welpton R. (2014) "Addressing the human factor". Working papers in Economics no. 1413, University of the West of England, Bristol. December

Sax Institute (2016) *Secure unified Research Environment User Training*. Accessed May 2016.

Stainton Rogers, W. (2011) Social selves and social identities. In W. Stainton Rogers, *Social Psychology: Experimental and Critical Approaches* (2nd Edition) (pp. 292-297) Oxford: Oxford University Press

Turner, J. C. and Reynolds, K. J. (2010). The Story of Social Identity. In: T. Postmes & N. Branscombe (Eds.) *Rediscovering social identity: Key readings* (pp. 13-32). Psychological Press: New York.

Tajfel, H., and Turner, J. C. (1979). "An integrative theory of intergroup conflict". In W. G. Austin & S. Worchel. *The social psychology of intergroup relations*. Monterey, CA: Brooks/Cole. pp. 33–47.

Wolters A. (2015) *Researcher management and breaches policy*. Health Foundation presentation to the Five Safes Workshop on Secure Data Access. Available at https://ukdataservice.ac.uk/media/604140/14_5safes_safepeople_wolters.pdf