Principles- Versus Rules-Based Output Statistical Disclosure Control In Remote Access Environments

by Felix Ritchie[1] and Mark Elliot[2]

Abstract

In recent years, the level of detail in confidential data made available to social scientists has increased dramatically. One particularly important growth area has been ensuring that research outputs do not present any residual disclosure risk. Traditionally this has been managed by specifying rules for researchers (the 'rules-based' model), but it is increasingly recognized that a 'principles-based' approach is both more secure and more cost-effective.

The principles-based approach requires a higher level of expertise from those managing access to data, and places the subjective assessment of risk at the forefront of decision-making; these two factors make data managers uncomfortable. In addition, knowledge of this approach is concentrated amongst a relatively small community, whereas the rules-based approach has dominated for half a century; data managers may not be aware of an alternative perspective.

This paper reviews the arguments for the two different approaches. They are not mutually exclusive: both take simple rules as a starting point, but the rules-based approach also finishes there. This has advantages in some circumstances, but the value of the principles-based approach increases with the sensitivity of the data and the scope of researchers to innovate.

The paper considers how the two approaches can be implemented. Although the principles-based model requires greater initial investment by both researchers and those managing access to data, this can bring substantial auxiliary benefits to the latter. The paper therefore concludes that a principles-based approach is generally preferable, and it is essential for the remote research data centres which dominate access solutions for the most sensitive data.

Keywords

Data access, data security, statistical disclosure control, principles-based, output SDC, researcher management

Acknowledgements

We are grateful for comments by members of the Administrative Data Research Network; Don Webber and Richard Welpton; and the anonymous referees, who suggested the inclusion of the summary table in section 4

> The main advantages of a rules-based model is the certainty and lack of ambiguity

Introduction

Since the early 2000s, social scientists have seen an explosion in data availability. The most important has been the increasing research access to highly confidential but high utility data (see for example Trewin et al., (2007) or Elliot and Purdam, (2015) for a discussion of this trend). This has been made possible by the development of secure data access solutions.

These allow the managers of such facilities to control the data access process with a great deal of security, but researchers are no longer necessarily restricted to physically visiting the facility manager's site; instead, they increasingly do so through remote access systems.

Some organisations have invested in secure remote job submission, where the researcher submits code to a server holding the data and receives back statistical results; for some data this is the most appropriate model, but for many types of data the dominant model is the remote research data centre (RRDC), where researchers access data through 'thin clients' (that is, where all the processing is done by the server holding the data, and the researcher controls the research through a web browser, for example). These allow researchers almost complete freedom to work with the data (other than removing them from the facility) and so are popular with researchers; they are also popular with data access managers, as such systems still allow a high level of control to be applied without the need for continual surveillance.

A key element of that control is checking to ensure that statistical outputs do not present any residual disclosure risk: a researcher will use the data to produce statistical outputs, and it is possible that those outputs could inadvertently breach the confidentiality of the underlying data (for example, by revealing that only one person in a small area has a particular illness). A critical point here is that when a researcher publishes output they are effectively moving data (for output is still data) from a highly secure setting to a completely insecure one. The change in data environment, means that the status of the data can, in principle, change from non-personal to personal (see Mackey and Elliot, 2013, for a review). Hence, disclosure checking of output is a vital part of the governance of secure data access systems.

Traditionally, 'output statistical disclosure control' (OSDC) has been managed by specifying rules for researchers to follow; for example, a requirement for all table cells to have at least three observations contributing to that cell. If the table meets the rules it can be released; if not, not. This ethos is reinforced by a half-century of statistical research focused on making tabular outputs safe (Hundepool et al., 2012).

However, in the last ten years it has become clear that simple models for tables have limited value in modern research environments, and increasingly common to discuss 'output SDC' as a separate research field[3]. The complexity of outputs led some data managers (e.g. Ritchie, 2007) to argue that the rules-based approach was both unsafe and inefficient; instead, a 'principles-based' approach could be both more secure and more cost-effective.

The principles-based approach uses rules to first-approximate a decision to release or not; but all preliminary decisions are subject to review and change if the researcher or the person responsbile for approving the release of output can make the case. The key is that both parties agree on the aims of the SDC process – and one of those aims can be to use the resources of the facility efficiently. The principles-based approach does not accept that outputs can be definitively classified as safe or not, only that the balance of probability says so. Finally, the principles-based approach acknowledges the value of research output in any decision. These differences may seem subtle, but they have profound implications for the way the facility and the researchers are managed.

The principles-based approach requires a higher level of expertise from the managers of research facilities, who must have both technical knowledge and an understanding of the research environment and researcher. In addition, it places the subjective assessment of risk at the forefront of decision-making. These two factors often make facility managers uncomfortable, as such organisations are typically risk-averse (Ritchie, 2014a). In addition, knowledge of the principles-based approach is concentrated amongst a relatively small community, whereas the rules-based model has been the dominant approach for half a century. Hence, facility managers may not be aware that there is an alternative perspective; if they are, they may not appreciate the subtleties of the principles-based approach, preferring instead a simpler model of data security (Ritchie and Welpton, 2014).

This paper aims to help facility managers make decisions about SDC, using the current best understanding of the pros and cons of each of the two process methodologies. For explanatory purposes, it uses the example of deciding on an approach to SDC for a remote RDC. This is because this is the case in which the difference between the two is starkest, and this paper demonstrates that the value of the principles-based approach increases with the sensitivity of the data and with the degree of freedom that researchers have to innovate.

Our conclusion is that for remote RDCs the advantages of principles-based SDC are clear. Beyond this, there are also lessons for other environments. The two approaches are not mutually exclusive: both take simple rules as a starting point, but the rules-based approach also finishes there. This has advantages in some circumstances (for example, rules-based is more appropriate for European Statistical System outputs; Eurostat, 2014), but as the principles-based approach is the generalisation of the rules-based approach, facility managers would do well to consider both.

The paper notes that, although the principles-based model requires greater initial investment by both the facility managers and the researchers, the necessary training can bring substantial auxiliary benefits to the facility manager. Put simply, the necessity to train researchers gives the facility manager an opportunity to encourage other positive behaviours, leading to increased 'legitimacy' and improved researcher behaviour (Ritchie and Welpton, 2014). Again, this is not always feasible, which is why rules-based modelling is sometimes more appropriate. The aim of this paper is to show when these benefits can be realised.

The next section considers the definition, benefits and costs of the rules-based approach, whilst section three evaluates the principles-based approach. Both consider the evidence for claims made. This is particularly important for the principles-based model, which brings risk-assessment to the fore. Section four summarises the dicussion, and concludes that the principles-based approach is essential for RDCs, whether remote or not. Section Five reviews implementation issues; Section Six concludes[4].

Some definitions are necessary for the paper:

An 'output' in the context of this paper is any statistical product arising from use of the data, intended for distribution beyond the technical confines of the research facility. An output may be a table, graph, regression model, frequency count, survival function etc., or it may be a paper containing multiple statistical outputs.
A 'commentary' in the context of this paper refers to any discussion about the analysis produced by the researcher. This includes written and verbal discussion.

'Statistical disclosure control' (SDC) means techniques to ensure that a statistical product or data does not breach confidentiality guidelines. 'Input SDC' (often just referred to as SDC) is concerned with protection of data before researchers have access to it, and is the subject of a different paper. 'Output-based SDC' (OSDC) is only concerned with statistics for distribution, and is the focus of this paper.

'Disclosive' is used in this paper as a short-hand for 'output which should not be made generally available as it retains a non-negligible residual risk that an individual population unit could be identified within them'; there might be variations in outputs between what is strictly unlawful and what is unwise and undesirable. For the purposes of exposition, we assume that disclosure equals a breach of confidentiality, with legal consequences and/or ethical implications.

The discussion below sits within the 'five safes'[5] framework (Desai et al. , 2014; see Camden, 2014, or Sullivan, 2011, for examples of use), which is a way of identifying sources of risk in data access:

- Safe projects – whether the data use is lawful
- Safe people – whether the researchers can be trusted to hold and use the data appropriately
- Safe settings – whether the manner of accessing the data offers protection
- Safe data – whether there is any inherent protection in the data
- Safe outputs – whether the outputs from the research pose a disclosure risk

The final criterion recognises that, however well-intentioned and competent the researcher is, accidents can happen – either through ignorance or through complexity.

This paper only considers the 'safe outputs'; that is, it is based on the assumption that data access is lawful and that researchers are not deliberately trying to misuse breach data confidentiality. How the researchers access the data is not relevant to this discussion. Inherent protection in the data is mostly irrelevant to this discussion of general principles and procedures, (although it does have some bearing on training issues and rules-based approaches, to be discussed later). Therefore, this paper does not place any limits on the data used to generate these outputs.

Rules-based OSDC

How it works

Rules-based OSDC consist of the application of a set of rules to determine whether an output should be released or not. Example rules might be

- "A table may only be released if there are at least three observations for each cell"
- "A regression may be released if not based entirely on categorical data"
- "A Herfindahl index of over 0.3 should only be released as 'over 0.3'"
- "Variance-covariance matrices X'X may not be released"

These are hard rules; they are expected to be applied consistently. This generates certainty in what output is acceptable, and allows machine-based SDC to be applied

Rules-based OSDC

Rules-based OSDC is simple and transparent. It is popular with data owners as it reflects the guidelines used to create official statistics, which are largely tabular. It is also necessary for the increasing number of automated systems which allow researchers to produce tables and analysis on the fly.

The main advantages of a rules-based model is the certainty and lack of ambiguity. This allows untrained (or partially trained) staff to clear outputs, and requires no training on the part of the researcher. Researchers can be given all the information they need in the form of hand-outs. This, for example, is how researchers at the Eurostat Safe Centre have been advised.

Criticisms of Rules based OSDC

There are three main disadvantages with the rules-based approach. All three are a consequence of the inevitable trade-off between confidentiality and efficiency problems. Consider devising a rule for the number of observations that have to be in a cell for it to be released:

- The Confidentiality Problem: a low limit increases the probability of disclosive cells being published
- The Efficiency Problem: a high limit increases the probability of non-disclosive findings not being published

First, no rule can guarantee non-disclosure, and the application of strict rules can provide a false sense of security. For example, in some cases no amount of units in a cell prevents that cell breaching confidentiality in some way. Hence, the rules based approach may under-protect the data in some cases

Second, a rules-based approach tends to over-protect the data. Protection will tend to dominate user value: the rule is the only protection against disclosure, and hence has to be much stricter than if other factors (such as the specific data being tabulated) are taken into account. As well as being inefficient, this can also create credibility problems when expert users are asked to follow rules which do not make sense to them.

Third, rules cannot cover all conditions; new rules need to be devised and agreed as new possibilities occur. A proliferation of rules is possible and this can in turn lead to contradictions. Consider defining dominance rule for table cells of N units (that is, determining whether one or two units contribute so much to a cell total that the cell can be considered, for all practical purposes, to only include those units). Rules that have been put forward include:

- the top unit does not account for more than w% of the total of the bottom N-2 records
- the top unit does not account for more than x% of the cell total
- the top two units do not account for more than y% of the cell total
- the Herfindahl index does not exceed z%

Each rule identifies a potentially problematic distribution of data, but the rules will not necessarily agree. Requiring all four rules to be met is over restrictive. Finally, unless the person clearing the output knows how the output was created, it is not possible solely from the output to determine whether the rules have been met or not.

In summary, the simplicity of rules-based OSDC is also its main limitation, particularly when dealing with an expert user base when rules-based OSDC can suffer from credibility problems. The blunt instrument of rules-based OSDC can under-protect in some cases, but is more likely to over-protect. This can cause frustration in researchers, which in turn is one of the factors associated with confidentiality breaches (Desai and Ritchie, 2010).

Principles-based OSDC (PBOSDC)

How it works

Principles-based OSDC is characterised by:

- researchers and output checkers both trained in SDC
- rules-of-thumb rather than hard rules
- freedom to approve any output in principle
- no duty to release any output
- responsibility for producing good output resting with the researcher
- output checkers considering the value of the output
- output checkers considering resource constraints

PBOSDC starts from the same perspective as a rules-based model: a set of rules exist to guide output approval. The difference is that these now become rules-of-thumb, rather than hard rules. They are there to guide the output-checker, but do not necessarily need to be followed. The output checker has complete freedom to exercise discretion, both to release an output and to decide not to release it. Clearance for release thus becomes a negotiation between researcher and approver. But an unrestricted negotiation is inefficient and so the rules-of-thumb provide the starting point.

It is important that both parties recognise the costs and the benefits of checking an output for clearance. Both parties want outputs to be processed quickly. The researchers want their outputs to be cleared; approvers want to be satisfied that the outputs are non-disclosive. A rejected output imposes costs on both parties; both parties therefore want to avoid this.
The key to POBSDC is that the person best able to assess whether an output should be released is the person who created it. The researcher knows whether the data was sampled, whether there are any dominance issues, how many observations there are in a cell, and so on. Most importantly, the researcher knows the value of the output to his or her research.

If the researcher knows the broad criteria on which output is checked he/she can ensure that (1) the output meets those criteria (2) the output checker has the information necessary to come to the same conclusion quickly and easily. If the output does not meet the prima facie conditions but is of high importance to the researcher, the researcher can try to persuade the output checker that this case is a valid exception to the rules-of-thumb. As this is likely to involve effort on both parties (and the output checker is under no obligation to concede the arguments), the researcher will have to consider whether the output is worth it.

The final part of the system is that the outputs can be rejected not just on the basis of disclosiveness, but on the basis of whether they are a good use of the output checker's time. It is irrelevant how much time the output checker actually has; the point of this rule is to allow the output checker to reward compliant behaviour and punish the malcontents by setting up appropriate incentives (Welpton and Ritchie, 2011, Ritchie and Welpton, 2012).

Consider a (male) researcher wanting to produce a number of outputs which are (to his mind) non-disclosive, but nevertheless breach the rules-of-thumb. He has three alternatives

a. Send the outputs through and hope for the best
b. Not send the outputs through
c. Raise the issue with the output checker and try to identify a solution which works for both parties

Option (a) is often a good strategy as a one-off; output checkers should be tolerant of researchers periodically sending through output which requires more work to check. However, as a repeated strategy it risks aggravating the output checker who can delay or stop checking future outputs, particularly if the researcher shows a failure to understand the principles of clearance.

Option (b) is common in practice; researchers working in restricted environment typically produce a large amount of outputs, not all of which is wanted. As a general rule, PBOSDC aligns with good statistical practice; for example, in discouraging very low cell counts.

Option (c) is the most interesting one, and frequently used in restricted facilities using PBOSDC. Researchers learn that a 'no surprises' policy appeals to output checkers, who also want an efficient clearance process. This can lead to inventive solutions which work for both parties; for example, validating program code rather than outputs.

Hence, researchers are incentivised to produce good output, and are encouraged to talk to output checkers before problems arise. Output checkers are encouraged to promote good practice amongst researchers, and to listen to user perspectives on the value of certain outputs.

To be most efficient, a PBOSDC process should adopt the 'safe'/'unsafe' statistics dichotomy (Ritchie, 2008; Brandt et al., 2010; Ritchie, 2014b). A 'safe' statistic is something which has very little or no inherent disclosure risk, such as regression coefficients. An 'unsafe' statistic is one which presumed to present a disclosure risk unless proved otherwise; for example, a simple tabulation. 'Unsafe' statistics, being ex ante disclosive, are much more time consuming to check.

Under PBOSDC, researchers can expect that 'safe' statistics will be cleared unless the output checker can demonstrate that it is disclosive; by definition, there will be almost no instances where this is the case. In contrast, 'unsafe' statistics will not be cleared unless the researcher can demonstrate that there is no disclosure risk. 'Unsafe' statistics passed for clearance impose a cost on the researcher related to the output checker's cost of clearance. The researcher should therefore be incentivised to concentrate on producing outputs made up of 'safe' statistics; and because the researcher is aware of how the value of specific outputs, there is an incentive to focus on important results and not large quantities of 'nice to have' information. This dynamic needs to be emphasised in the service access training.

Finally, PBOSDC ideally attempts to raise awareness of speculating about the identity of records when commenting on results. For example, although a researcher may produce good statistical outputs, he or she might draw attention to the presence of a particular outlier which affects results. This cannot be dealt with by

a rules-based approach, but only by ensuring that researchers are aware of the risks posed by incautious discussions of data quality. In summary, the aim of PBOSDC is to create an atmosphere where clearance is seen as the joint responsibility of researchers and output checkers. The common interest encourages the production of safe and useful outputs cleared by an efficient process.

However, for this to happen effectively, both need to understand the incentives of the other party, as well as the principles of SDC. Hence, there is a need for training of researchers and of output checkers. This is a major difference compared to the rules-based approach.

Advantages of Principles based OSDC

The direct advantage of PBOSDC is that it should allow both more and safer outputs than the rules-based approach. Consider again the potential errors noted above arising from a threshold rule, the confidentiality problem (too low a limit —> disclosive outputs) and the efficiency problem (too high a limit—>safe outputs rejected). Under PBOSDC, there is no conflict: a high limit, likely to remove almost all disclosure risk, is set as the initial rule of thumb. If a researcher feels that this is too high in a particular case, he or she can make that argument, confident that the arbitrary rule of thumb will be replaced by a review of the circumstances in the specific case.

This works because most researchers requiring access to detailed data want it for analytical purposes, which are more likely to be safe statistics. When unsafe statistics are the main focus of output (for example, the OECD commissioned work on high-growth companies which required very many tabulations from the VML), output checkers and researchers have an incentive to work together to agree in advance the range of permissible outputs. The indirect advantage of PBOSDC is the ability to develop a culture of confidentiality awareness amongst researchers. This has positive feedback effects: a demonstrably educated and trustworthy researcher base can have systems designed to reflect that knowledge and trust, and a more appropriate working environment encourages positive behaviour from researchers (Desai and Ritchie, 2010; Ritchie and Welpton, 2014). This has been the pattern of development in UK Research Data Centres over the past decade.

Criticisms of Principles based OSDC

The three main criticisms of PBOSDC are uncertainty, inconsistency, and resource requirements.

PBOSDC, by design, introduces uncertainty into the process, as the whole ethos of PBOSDC is that decisions are taken in specific contexts. Training and ongoing engagement are therefore required to build trust.

If several output checkers are reviewing outputs, they may make different decisions; it is also possible that the same output checker will make apparently inconsistent decisions in different circumstances. This is because the checkers are making decisions responding to the particular circumstance of output, data and researcher and not just a rule about an output. To ameliorate this good training (of both researchers and output checkers) ensures that there will be agreement on the principles. An important part of the process is that is that there is no rules-based yes/no answer; therefore output checkers should be aware that any decision necessarily has an element of subjective judgement and may need to be justified.

It has been argued that the need for output checkers rather than automatic processes (or checking by staff with limited expertise) increase the costs of a facility, as does the explicit allowance for researchers to challenge decisions. While this has not been the case to date in facilities where PBOSDC is fully implemented, it is clear that some expenditure on training for staff and researchers is necessary; when one facility failed to train its new staff, there was an immediate impact on clearance rates and quality.

PBOSDC in practice

In practice, none of the criticisms suggested in the previous section have proved significant. The evidence for this comes from the eight years of PBOSDC at the Virtual Microdata Laboratory (VML) at the UK Office for National Statistics, where the process was developed, as well as more recent experience at SDS and HMRC Data Lab, also in the UK[6] .

Uncertainty does exist. However, there should be no uncertainty about the process or the criteria for deciding whether to release an output or not. The purpose of researcher training is to help researchers understand the uncertainty and manage it. Researchers have made mistakes, and while these are mostly oversights on the part of the researchers, a small number are due to researchers not understanding the principles of SDC. In general these were dealt with by discussion with the researchers, explaining the error. Recidivism rates were negligible, but a very small number of VML researchers were asked to re-attend training (less than five people in seven years, out of over eight hundred trained researchers). Although there are no figures, the number of requests for output refused at the VML was believed to be around 5%.

Inconsistency exists but there is little evidence to date of it being significant. Over seven years around twenty output checkers were employed at ONS; periodic checks were carried out, and although output checkers showed some slight variance, there was no practical difference in outcomes. It was discovered that one output checker was seen as 'softer' by some researchers, but even those outputs were well within the safety margin. In one case, a researcher's output was seen by five output checkers (including the 'soft' checker), all of whom independently gave the same opinion. So whilst differences do exist, in practice these have no notable impact.

If the research facility is not centrally run, or if several facilities are trying to co-ordinate SDC policies, there is another level at which variance may happen: between centres. There may be some cultural variability and the potential for local ethos to develop should be acknowledged and monitored. A common training framework where the principles of SDC are gone into in some depth, practice sharing between centres, "test" submissions and cross-centre case study reviews can help to mitigate this.

The reason inconsistency and uncertainty have not been major problems to date is because of the built-in safety margins. As noted above, ignoring the Efficiency Problem means that the rules of thumb to address the Confidentiality Problem can be made much stricter; a confidentiality breach therefore requires a considerable error by both researchers and output checkers.

This margin of error is also at the heart of the efficiency of the process. Output checkers can clear large amounts of output because they have confidence in the extra-safe rules-of-thumb for "unsafe statistics", and because "safe statistics" require little scrutiny. At its peak the VML was dealing with 2500 clearance requests a year, or roughly ten every working day; one person was allocated to be output checker for that day, and the target was that this should take no more than half an hour, a target generally achieved. As a 'clearance request' typically consisted of several regressions, a couple of descriptive tables (and in one case over fifty graphs), and/or a log file, the actual number of statistical outputs being checked was much more than ten per day.

This work rate was maintained by emphasising the researcher's role in clearance. For example, the VML output checkers had an informal policy of clearing the easiest outputs first; this was communicated to researchers at training sessions. This gave researchers an incentive to build up a reputation of producing 'good' (i.e. easy to check) outputs.

Finally, there is evidence that the training sessions (and the relationship built up between researchers and output checkers) do foster a culture of confidentiality awareness, with examples of researchers being self-policing. It is unlikely that academics using the VML, SDS and HMRC Data Lab would see themselves as being particularly SDC aware, as this is their only experience of it[7]. Nevertheless, a comparison with users of facilities in other countries shows that the UK academia is much better informed. This is why the UK model of researcher training was adopted largely unchanged by Eurostat as recommended best practice (Brandt et al., 2010), and why researcher training should be seen as an investment rather than an expense.

Rules-based versus principles-based OSDC

The various aspects of the two approaches can be summarised as follows.

	Rules-based	Principles-based
Complex	No	Generally not – rules-of-thumb usually sufficient
Flexible	No	Yes
Transparent	Yes	Yes, if output checkers record factors leading to exceptional judgment
Consistent	Yes	Generally, but scope for minor variations
Secure	Limited by efficiency	Yes - able to handle lower and higher risk cases
Efficiency	Limited by security	Yes - tailored to circumstances
Risk management	Yes/no model	Explicit 'balance-of-risks' model
Sensitive to context	No	Yes
Sensitive to user needs	No	Yes
Suitable for automation	Yes	Only as initial gatekeeper – humans are final arbiters
Requires researcher training	No	Yes
Requires researcher engagement	No	Yes
Requires co-ordination of output checkers	No	Yes
Cost	Low	High initial cost, low or negative ongoing costs

Although described above as alternatives, there is a relationship between rules-based and principles-based output SDC. The rules act as the starting point for PBOSDC output checking (see Brandt et al., 2010). However, 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.
- these rules of thumb can then be more restrictive as prima facie efficiency has a low priority in the setting of the default values.

For 'safe statistics' the 'hard rules' and 'rules-of-thumb' are the same; by definition, 'safe statistics' are those which are amenable to the identification of simple yes/no cases (Ritchie, 2008). 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 the RDCs. In situations where facility managers have less opportunity to manage researchers' activities and outputs, PBOSDC may be harder to implement, as the active engagement of researchers is essential.

Implementing PBOSDC

Conceptual development

The most detailed expression of PBOSDC is the Eurostat-approved guidelines of Brandt et al. (2010). These were derived almost entirely from the VML rules, with the exception of the dominance rules. Since the publication of the Eurostat guidelines there have been a number of minor developments. For example, on regression, several authors (e.g. Reznek and Riggs, 2005; Ronning, 2011; Bleninger et al., 2011) have investigated options for deliberately creating misleading regression results, and US and Australian works have studied regression models in remote-execution systems. Ritchie (2012, 2014b) incorporates most of these results, but new queries appear (for example, on the tabulation of binary variables, and how single-observation categories are treated in regressions) In addition, not all UK work (for example, on variance-covariance matrices) was adopted as it was felt to be too obscure for a general document.

None of these are major challenges but they highlight the need for updating the current state of knowledge and communicating this to facility managers and researchers. When the VML guide to SDC was the only source and was owned by the VML team, updating was straightforward. One of the negative consequences of the wider use of the PBOSDC approach is that there is no clear mechanism for maintaining and disseminating knowledge.

Facilities wishing to implement PBOSDC may therefore need to collaborate with other facilities to ensure that a consistent approach can be developed and applied[8].
It has been suggested that the theoretical basis for PBOSDC and the safe/unsafe statistics model does not provide sufficient reassurance to potential data suppliers. In addition, theoretical models may miss some plausible outcomes (that is, those likely to occur in practice) arising from, for example, naïve researchers.

One solution is to set up an 'ethical hacking' mechanism to probe both genuine and fake outputs to test what information could be acquired. A second would be to have periodic audits of outputs from the facility by a competent body, perhaps another facility. This has the added advantage of encouraging consistency across facilities.

Need for training
Researcher training is essential for PBOSDC. Untrained researchers cannot be expected to know the basis of SDC, and the training itself should be used to develop an affinity between the researchers and the facility. There are several PBOSDC training manuals/presentations which can be drawn upon and updated. This training could be part of any user accreditation process.

There is already significant training experience in the UK and other countries, and therefore any facility would not need to start from scratch. However, as one of the key elements is to encourage researcher engagement, training needs to be sensitive to the needs and indeed interests of the researcher community; what may appeal to Mexican researchers may be different from the expectation of Norwegians. Training also needs to be sensitive to the data. Much of the extant training focuses on social and business survey data, which are of more limited risk, but administrative data brings additional problems (data may be a census; health data is typically more prone to outliers and retains its sensitivity over time). This is particularly important if tables are likely to comprise a lot of the outputs.

Finally, there is the need to train and update the knowledge of output checkers. Periodic peer review has proved useful in the past. Discussion forums allow knowledge to be shared, discussed and updated across facilities. If made available to researchers, these would also have value for researchers, although not necessarily in the same detail and with a full range of views expressed. One option would be to set up a discussion forum for facility staff, with a summary/FAQ page for researchers.

Trusted researchers
One decision to be made is whether some researchers could have different rules applied. Either particular types of people (e.g. full professors) or those who have built up reputations with the output checkers could have fewer checks imposed on them. The argument is that the burden for output checking is unnecessary, and creates ill-will amongst senior researchers who have proven their expertise. Some facilities do use such differentiated models. In practice, these arguments do not hold. Seniority is no guarantee of good practice. Indeed the opposite can be true; experience shows that junior staff are more enthusiastic adopters of safe practices. Similarly, familiarity may make mistakes less likely but does not eliminate them. As mistakes are far and away the most likely reason for failed clearances, long-term usage does not seem sufficient cause on its own to remove checks.

These arguments are also based on the assumption that all outputs are equal. As should be clear from the discussion above, over time researchers should develop a sense of what is and is not allowed, and tailor their outputs accordingly. In other words, the PBOSDC encourages the development of expertise in output assessment by all parties, and thus efficient exchanges. There is no need to create an artificial group of 'good' researchers; besides, creating 'classes' of users could discourage knowledge and experience sharing.

Malevolent researchers
Any output disclosure control system will have additional difficulties with a user who deliberately sets out to breach the system. PBOSDC assumes that researchers are well-intentioned and interested in generating good statistical outputs. It may be possible to spot unauthorised outputs, but in practice a user set on breaching procedures would be able to disguise inappropriate outputs (for example by burying discovered data in complex model output). It should be noted that all the data used in current training programmes for controlled environments is wholly invented, yet plausible.

The success of PBOSDC therefore depends upon the 'safe people'/'safe project' dimensions of the Five Safes model. That said, at present, there are no known examples of academic researchers maliciously breaching confidentiality rules. There are numerous examples of well-intentioned researchers making mistakes, and a smaller number of cases of researchers deliberately ignoring procedures to make life easier for themselves (but again, without intending to disclose confidential data).

Setting up a system which would stand a chance of picking up malicious attacks therefore has no realistic prospect of success, and would increase greatly costs and clearance time. If a facility believed that malicious attack was a significant risk, then examining user logs and codes would be a more useful place to look for malpractice – and would also be necessary for forensic evidence in the event of a disciplinary event.

In summary, PBOSDC cannot stop ill-intentioned researchers; it is designed to deal with cases of accidental rule breaking and errors. If malicious attack is felt to be a problem, then this should be tackled at the 'people' level[9].

Multi-stage clearance
The UK VML operated a two stage-clearance procedure: 'intermediate' outputs could be released to researchers who could work on the further analysis at their home institution. 'final clearance' was given when papers were ready for general distribution. The UK Secure Data Service only allowed 'final clearance': papers are fully prepared within the SDS virtual facility. These two choices reflect physical differences. The VML involved travel to a specific location; it was not thought to be a good use of restricted facilities to have researchers editing papers, and the opportunity to discuss results with co-researchers was limited. In the SDS these two issues are less important.

Note however that VML applied full PBOSDC at the intermediate stage; the assumption was that once an output had left the VML's control it could end up in the wild however well-intentioned the researcher. This assumption turned out to be correct, even if unsanctioned releases were rare. The VML's 'final clearance' stage imposed a level of super-checking on outputs – asking researchers to limit outputs to the minimum necessary (compared to the

intermediate stage were multiple variants might be tried). As this reflected the research stages (exploratory work, produce many outputs, refine, and then publish) the model worked.
The two-stage model did introduce an extra layer of administration, as now both intermediate and final clearances had to be checked and recorded. However, it did speed up the intermediate level clearance, as VML staff knew that if they made a mistake they had a 'second chance' to address it. This increased the already-wide margin for error in the VML PBOSDC procedures.

Given that researchers were only bound by VML procedures (and not required by law) not to publish intermediate outputs, the 100% co-operation (allowing for mistakes) can be seen as a positive reflection of how researchers respond to appropriate training. However, the VML did make researchers aware that failure to follow procedures would affect the way that future access to data was viewed. This may have had more of an effect, and may be of relevance to a facility choosing to adopt two-stage clearance.
Allowing intermediate output out implies increasing risk unless adequate mitigation is in place. Relying 100% on trust of researchers without any bounds implies allowing unmeasurable variations in risk and therefore is not sufficient to provide that mitigation. Appropriate risk mitigation implies processes such as:

- Secondary licensing agreements.
- Minimum required security arrangements for intermediate output.
- Specified lists of persons who will have access to the intermediate outputs.
- Occasional random audits.

Summary

This paper recommends PBOSDC be adopted for use in RDCs, whether physical or remote. The reasons for this are twofold (i) principles based SDC can produce output that is of higher quality at the same or lower level of risk; and (ii) the opportunity of building a relationship with researchers can generate multiple benefits. In short, PBOSDC is both safer and more efficient that rules-based approaches and it encourages the development of a culture of expertise in confidentiality.

There are already guides and training programmes boasting several years' experience of PBOSDC. A facility wanting to adopt PBOSDC can build upon these, perhaps tailoring them more to its particular researcher group and data. PBOSDC needs to be integrated into a training programme; it assumes that researchers are well-intentioned (if liable to make occasional mistakes). There also needs to be a mechanism to ensure consistency across checkers (and possibility sites). The facility may also want to invest some resources in ethical hacking to provide extra reassurance to data owners. Finally, the facility needs to determine whether it wants a one- or two-stage clearance process. While PBOSDC at the point the output leaves the restricted facility can be the same, the perceived and actual security differs.

For non-RDC environments, the case for PBOSDC is less clear. For example, if researchers' only sensible interaction with the facility manager is receiving a partially anonymised file on CD plus guidelines for publishing safe statistics, then a rules-based approach may be simpler. However we would recommend that even a discussion of rules should be placed in the context of the principles of SDC: in general, the support officer is more welcomed than the policeman.

References

Bleninger P., Drechsler J., and Ronning G. (2011) Remote data access and the risk of disclosure from linear regression", Stat. and Op. Res. Trans. Special Issue: Privacy in statistical databases, pp 7-24 http://www.idescat.cat/sort/sortspecial2011/DataPrivacy.1.bleninger-etal.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

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

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

Duncan, G., Elliot, M. J. and Salazar, J. J. (2011) Statistical Confidentiality. Springer, New York

Elliot M.J. and Purdam K. (2015) 'The Changing Social Data Landscape' in Halfpenny, P. and Procter, R. (eds.) Innovation in Digital Research Methods. Sage

Eurostat (2014) Treatment of Statistical Confidentiality, manual and exercises, Luxembourg: Eurostat

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., and De Wolf, P. P. (2012) Statistical Disclosure Control. Chichester, UK: John Wiley & Sons.

Mackey E. and Elliot, M.J. (2013) 'Understanding the Data Environment' XRDS 20(1); 37-39. http://xrds.acm.org/article.cfm?aid=2508973

Ritchie F. (2007) Statistical disclosure control in a research environment, mimeo, Office for National Statistics; available as WISERD Data Resources Paper No. 6 http://www.wiserd.ac.uk/wp-content/uploads/2011/12/WISERD_WDR_006.pdf

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. (2012) Output-based disclosure control for regressions". Working papers in economics no. 1209. University of the West of England, Bristol. http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/economics2012/1209.pdf

Ritchie F. (2014a) "Resistance to change in government: risk, inertia, and incentives", Working papers in economics no. 1412. University of the West of England, Bristol. http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202014/1412.pdf

Ritchie F. (2014b) "Operationalising 'safe statistics': the case of linear regression", Working paper no. 1410, Department of Economics, University of the West of England, Bristol. http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202014/1410.pdf

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. and Welpton R. (2014) "Understanding the human factor in data access", Working papers in economics no. 1413. University of

the West of England, Bristol. http://www2.uwe.ac.uk/faculties/BBS/BUS/Research/Economics%20Papers%202014/1413.pdf

Ronning G. (2011) Disclosure Risk from Interactions and Saturated Models in Remote Access, IAW Discussion Papers No. 72, June http://www.iaw.edu/RePEc/iaw/pdf/iaw_dp_72.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

Trewin D., Andersen A., Beridze T., Biggeri L., Fellegi I., Toczynski T., (2007) Managing statistical confidentiality and microdata access: Principles and guidelines of good practice; Geneva, UNECE /CES.

Welpton R. and Ritchie F. (2011) "Incentive compatibility in data security", presentation to IASSIST 2012, Vancouver http://www.iassistdata.org/downloads/2012/2012_a2_welpton_etal.pdf

Notes

1. Corresponding author: Felix Ritchie, Bristol Business School, University of the West of England Bristol, Coldharbour Lane, Bristol BS16 1QY. Email: felix.ritchie@uwe.ac.uk.
2. Mark Elliot, School of Social Sciences and Data Research Institute, University of Manchester, Oxford Road, Manchester M13 9PL. Email: mark.elliot@manchester.ac.uk
3. There is a large literature on statistical disclosure risk assessment and control methods. We do not discuss this in detail as here we are talking about two top-level process methodologies rather than the specifics of individual technical methods. We would direct the reader who is interested in the detail to recent comprehensive field reviews (Duncan et al. 2011 and Hundepool et al. 2012)
4. A more extended discussion of this argument is available in Ritchie (2007).
5.. It is important to stress that "safe" is used here not in its absolute postpositive sense (free from danger or risk) but in its relative sense (the degree to which a solution affords security or protection from risk); see Ritchie (2014b, section 5).
6. PBOSDC is also used at other restricted facilities in Mexico, Germany and the Netherlands, as well as informally in other countries.
7. An exception is medical sciences where data protection has had a much higher profile, and researchers in all countries tend to have a greater awareness of confidentiality issues.
8. In January 2015, all the UK RRDCs initiated a working group on SDC; one of the group's functions is to determine how guidelines can be effectively maintained, distributed and updated.
9. Note that rules-based OSDC is even more susceptible to malicious attacks, as the yes/no approval process means that anything that looks acceptable will be approved without further checking.