Access to Sensitive Data: Satisfying Objectives Rather than Constraints

Felix Ritchie

The argument for access to sensitive unit-level data produced within government is usually framed in terms of risk and the legal responsibility to maintain confidentiality. This article argues that the framing of the question may restrict the set of possibilities; a more effective perspective starts from the data owner’s principles and user needs. Within this principles-based framework, the role of law changes: It becomes an ‘enabling technology’, helping to define the solution but playing no role in setting the objectives.

This shift in perspective has a number of consequences. The perception of ‘costs’ and ‘benefits’ is reversed. Law and established practice are distinguished and appropriately placed within a cost-benefit framework. The subjectivity and uncertainty in risk assessments is made explicit. Overall, all other things being equal, the expectation is that a move towards objective-based planning increases data access and improves risk assessment.

This alternative perspective also addresses the problem of the public-good nature of research outputs. It encourages the data owner to engage with users and build a case for data access taking account of the wider needs of society.

The UK data access regime is used as the primary example of the arguments in this article.

Key words: Confidential data; data access; data security; public goods; risk.

1. Introduction

It is nowadays widely accepted that access to confidential or sensitive microdata collected by government is essential for the research needed to produce an evidence base for policy; see Trewin et al. (2007) for a discussion. This data is usually collected either by statistical agencies to produce aggregates, or by government departments as part of their work. In both cases, use of the underlying microdata directly allows the collecting body to leverage their investment in data collection at minimal additional cost.

General agreement on the principle of research access is common; but principles can take a back seat when implementation is considered. In particular, the confidentiality of the data becomes paramount, and access to data focuses on how that confidentiality can be maintained. However, there is ample evidence to suggest that government is likely to be

1 Bristol Business School, University of the West of England, Frenchay Campus Bristol BS16 1QY, Bristol, UK. Email: felix.ritchie@uwe.ac.uk

Acknowledgments: This article is developed from a presentation for the Statistics New Zealand Official Statistics Forum in March 2010. I am grateful to SNZ and Motu for funding my visit and giving me the opportunity to draw out some of the themes here. The germ of this article arose from discussions with Richard Welpton of the Secure Data Service. I am also grateful to Tanvi Desai for detailed comments on an earlier draft, and to the referees and editors for incisive comments, particularly in relation to the appropriate role of risk.
collectively and individually risk-averse (see, for example, OAG 1998; House of Lords 2006; Pfeifer 2008; Buurman et al. 2012; Hall 2013) and so decisions taken may not be socially optimal.

This article argues that changing the perspective to concentrate on the principles governing data access can help to improve the quality of decisions taken, as well as clarifying exactly what risks are being run and what the benefits are. The basis of this argument, well-attested in psychology, is that the framing of the question affects the answers that are generated.

The next section proposes a perspective on access which emphasises the predominance of objectives over constraints. This leads to a model where law and technology are ‘enablers’: That is, they inform, and may constrain, decisions to be taken on how to implement an objective, but do not define the objective itself. The following two sections discuss these in more detail, and Section 5 examines how this facilitates an understanding of the role of risk.

Section 6 considers how this change of perspective can inform the debate on the ‘public goods’ problem of data access identified by Ritchie and Welpton (2012). The article concludes by noting that the arguments advanced here run counter to the natural decision-making structures in government, and so an efficient system of confidential data access may need an active and engaged sponsor.

For simplicity, the article throughout refers to the options of National Statistics Institutes (NSIs), who are generally the main or only holders of confidential government research data. However, it should be clear that the arguments apply to any owner of confidential data considering giving access to that data for research.

The author has been involved in data access in the UK for over a decade, has formally and informally advised the OECD and Eurostat, and has worked on data confidentiality with NSIs in many different countries. The examples used in this study are mostly drawn from the author’s experience in the UK, as it is difficult to ascertain whether an individual’s perspective truly reflects the experience of an organisation or country without having worked there. However, I am confident that the characterisation of NSI behaviour in this article, while simplified, is a fair reflection.

2. The Framework Principles

2.1. Constrained Decision Making: The ‘Constraint Model’

The usual decision-making process for giving access to confidential data can be framed as in Figure 1, which we will refer to as the ‘constraint model’.

![Fig. 1. The ‘Constraint Model’](image-url)
That is, organisations ask:

- What does the law say we can do?
- Given that, what do we need to ensure?
- What technologies are available to satisfy those constraints, and what are the needs of the user that can be satisfied within those constraints?
- How do we employ technology to meet the identified user needs in the best way?

The problem with the Constraint Model is the first step. Clearly, acting within the law is a requirement of any agency. The problem is that ‘the law’ is rarely a simple, unambiguous construct with only one possible outcome; a statement of practical law is an interpretation in relation to a specific set of circumstances. However, focusing on a particular interpretation constrains the set of solutions to a subset of outcomes, particularly if the circumstances surrounding the interpretation are not explicitly made known.

Consider the UK experience in 2003. The Office for National Statistics (ONS), the UK’s NSI, was reviewing the options for giving academic researchers access to confidential business data. The prevailing legal opinion was that this was not possible: The Act governing such access strictly limited access to employees of the UK government.

This seems crystal clear, until it is considered that the question being asked is the implicit one, “can academics, in their own right, have access to business microdata?” This is a very specific and, as it turned out, very limiting question. An alternative question was put to the government legal advisors: “Can academics become Civil Servants for the purposes and duration of their research?” There were several positive responses to this question; a form of secondment was taken as the most workable. As a result of changing the perspective, an outcome, previously considered impossible, was achieved with a solution in keeping with both the spirit and letter of the law. For details, see Ritchie (2009).

The specific legal arrangements continued to evolve as different circumstances came to light. ONS’s Legal Services unit periodically reviewed the secondment arrangements, and the team providing access was required from time to time to amend its procedures to address potential areas for challenge. For example, the team was asked to demonstrate that access was through ‘fair and open competition’, and to specify the criteria for determining whether access contributed to ONS’s ‘benefit’.

The important lesson from this is that the attitude of the NSI determined the outcome; that is, whether access could be granted or not. Both the research team and the legal team shared the same aim: to see wider research use being made of confidential data, lawfully. The specifics of implementation were just that: specifics of implementation, not a universal statement of law.

2.2. Principles-Led Decision Making: The ‘Objective Model’

This focus on objectives rather than implementation leads to a rather different framework for access, as displayed in Figure 2:
The questions now are, in order,

- What are our operating principles, and what do users want?
- What how do we turn this into a set of objectives?
- What legal and technological options are available?
- How do we employ these alternatives to meet the objectives in the best way?

This ‘Objective Model’ puts the aims of the NSI and the user at the start of the decision process. Law has the same status as technology: Just as all implementations are limited by the existing technology, so they are constrained by the existing law. But technology and law are both used to achieve the objectives; they do not count towards the definition of those objectives.

One of the implications of the Objective Model is that multiple legal and technological solutions may meet those objectives; there is no need to identify ‘the’ legal or technical solution. Several solutions might coexist; the aim is to find the combination of solutions that meets the objectives best.

A second implication is the primacy of the ‘user need’ (with the NSI itself as one class of user). User demands can be stereotyped as “give me all the data now, on my desktop, with no restrictions”, but this is an exaggeration. Researchers are generally aware that not all tasks need all data, particularly as more detail typically involves more restrictions. As an example, the UK Data Archive provides many datasets in both anonymised and detailed form, with the latter having more access restrictions. A bona fide UK researcher would have little trouble getting access to either, but the usage of the anonymised files massively outstrips that of the restricted-access detailed files. This model, of some sort of data archive holding files for distribution balanced by more restricted access to more detailed data, is relatively common across countries, indicating that users can make balanced judgements about costs and benefits.

The dichotomy characterised by the Constraint/Objective Models may be unfair to individual NSIs, but in the author’s experience, based on work with numerous NSIs and international organisations, this state generally prevails in the real world. There are units within NSIs that consider user needs and then consider how to meet them; but the majority still seek to identify the legal framework and then assess which user needs can be accommodated within that framework. An even smaller minority are prepared to consider NSI objectives and user needs jointly without reference to legal limits on implementation.

### 2.3. Semantics or Substance?

It could be argued that this is a largely semantic argument; that is, the real questions are always about implementation, and the same solution could be derived by individuals...
working from the different models. A rational organisation with all the necessary information would always come to the same conclusion, whatever its conceptual stance.

An analogy is with constrained optimisation. Take the typical undergraduate economics problem of maximising utility subject to a budget constraint, resulting in an optimal utility of, say, $U^*$. It can be demonstrated that minimising the expenditure needed to achieve that given level of utility $U^*$ recreates the budget constraint from the first problem, assuming the constraints are binding. Hence these are referred to as the ‘primal’ and the ‘dual’, with each generating ‘shadow prices’ for the cost of the constraint (see e.g., Varian 1992).

This focus on the equivalence of solution hides an important outcome. In the maximisation problem, the shadow prices are the benefit to be gained by loosening the budget constraint. In the minimisation problem, the shadow prices reflect the cost of any further increases in utility. These are clearly two different concepts, and the way the problem is posed reflects the analysts’ interest.

Similarly, the Constraint and Objective Models imply fundamentally different mindsets: the difference between “what can we do?” and “what would we like to do?” In coming to a solution, the perception of what has been given up to achieve that outcome differs, even if the outcome is the same.

However, this is not simply an alternative perspective. In the mathematical problem the choice of maximisation or minimisation does not affect the problem parameters, only the interpretation of results; but in the human world, the outcome can be substantially affected by the way that the question is framed.

Since the 1970s the psychological literature has repeatedly demonstrated the importance of framing effects; see, for example, Kahneman (2012, especially chap. 34) for an overview, Mellers et al. (1999) for experimental evidence, or De Martino et al. (2006) for a discussion of the psychological basis. This is also recognised in environmental and behavioural economics, politics, and marketing, for example – all subjects where the focus is on understanding what influences the decisions of people.

The relevance of this to the Constraint/Objective Model discussion is that the initial framing of the problem will lead to either ‘losses’ or ‘gains’ being identified. Losses tend to be felt more keenly than gains, and the certainty of outcomes affects decisions. As result, there is a tendency, all other things being equal, to stick to the starting point; see Kahneman et al. (1991) for examples. Samuelson and Zeckhauser (1988) identify this as ‘status quo bias’.

Consider Figure 3, below.

![Fig. 3. The importance of the starting point](image-url)
The Constraint Model could be considered as starting from ‘nothing is allowed’; as potential candidates for solutions are evaluated for conformity with the legal environment, constraints can be relaxed. The ‘Objective Model’ is starting from ‘everything is allowed’, and withdrawing from that position as solutions are shown to be unlawful or impractical. The theories of framing suggest that, while two NSIs starting from different perspectives could come to the same conclusion, the more likely outcome is that they will differ in their implementation. All other things being equal, the Objective Model is more likely to lead to more data access. Hence, this is not a semantic discussion: The conceptual stance of the organisation affects the outcome.

3. Law as an Enabler

Once law is seen as a tool to inform decisions about implementation, rather than a governing framework, some useful results appear.

First, attention focuses on the purposes of legal advice. Lawyers are professionally cautious: That is, one of their duties is to ensure that clients are warned about liabilities and consequences. Advice is likely to focus on avoiding negative outcomes. In the Constraint Model, legal advice sets the ground rules for all subsequent decisions. This places an inappropriate burden on lawyers, who are unlikely to be experts in data access. In the Objective Model, legal advice is taken in the context of specific solutions. Lawyers are not being asked to speculate on potential future interpretations, and any advice is reviewed in the context of the objectives. Both the giving and receiving of advice is more effective.

Second, changes to the law can be evaluated more easily. If the legal environment changes, the constraints on the NSI change; the set of feasible delivery options changes. Under the Constraint Model, the ‘value’ of the changed law is whether the outcomes being produced are now better for society. In the Objective Model, NSI objectives are invariant to law; therefore, a test of the likely effectiveness of any new law is simply whether it improves the way the NSI meets its objectives.

Again taking the case of the UK, in 2008 the Statistics and Registration Act came into force. This formally gave ONS a function of supporting research for the public benefit, and provided a simple universal legal gateway for access to ONS microdata. This greatly simplified the process through which researchers gained access to data, clarified the role of researchers’ use of ONS data, and brought all ONS data under the same legal framework for research. ONS objectives were largely unaltered, and so the impact of the law was a straight efficiency gain.

Third, the difference between law and established practice can be clarified and challenged. In the context of the diagrams above, ‘law’ includes the NSI’s procedures, which often go beyond the law into areas where the NSI feels it has an ethical or operational responsibility even if no legal responsibility exists. Fixed ways of working, particularly when in place for a long time, can also easily be confused with law. Even when procedures are explicitly recognised as NSI policy decisions, they can still be seen as immutable.

For example, at an OECD meeting on international data sharing, the author discussed country attitudes with a representative from a European NSI. The representative initially stated that such data sharing was not allowed in that country’s law; after some minutes of discussion, it transpired that the true position was that it was legally possible but the NSI would not allow it. This is a small difference but a very significant one.
Under the Constraint Model, challenging established practice is hard. A new or changed objective needs to demonstrate that it fits into the current understanding of the legal environment, which may be partly defined by established practice. But this begs the question: Why should objectives need to justify their value by reference to specific implementations? Surely the implementation has to address the objective, not the other way around?

Under the Objective Model, established practice has to justify its existence using fair and proper criteria for how well it addresses an objective: cost, benefit, effectiveness, legality, impact of disruption against the alternative solutions. These are also more easily quantifiable: The impact of a change on access rules on IT expenditure, for example, can be readily identified. Under the Objective Model cost-effective practice is what matters; the value of ‘established practice’ is only reduced costs of learning or change.

4. Technology as an Enabler

Technology (meaning all the practical matters surrounding access to data, including cost decisions) as an enabler is relatively self-explanatory. The technological options can be broadly grouped into six types:

- **Anonymisation** of the data: This is used for public files, such as those on the web.
- **Licensing** of researchers, sometimes combined with a degree of anonymisation, is still the most common way for researchers to get access to microdata.
- **Secure ‘research data centres’** (RDCs), laboratory facilities at the NSI or the researcher’s base; for many countries, this is still the only way to get access to detailed data.
- **Remote access**, where ‘virtual’ RDCs allow users to manipulate data unhindered by geography; although the technologies are common, implementation varies greatly from restricted-site access only to direct access from the internet.
- **Remote job submission**, where users send statistical programmes to be run and get back results, are relatively uncommon, but a number of NSIs have been exploiting web technologies to develop friendly interfaces.
- **Synthetic data**, which has the same characteristics as the real data but has been imputed from statistical models; the resulting dataset is then intended to be safe for distribution.

Most countries employ a number of these options, and often these solutions are combined. For example, some US Census Bureau data is made available at restricted on-site RDCs, but synthetic equivalents are accessible through a virtual RDC.

NSIs tend to be risk-averse and avoid new solutions, but in most of these areas a prospective data manager can draw on a wealth of international experience in implementation. As Ritchie (2013) notes, for strategic planning purposes an NSI can assume that an ‘off-the-shelf’ solution is available to meet its objectives. The everything-is-possible answer does not help planning, so Ritchie (2009) reduces the solution set by employing the concept of the ‘data access spectrum’. This suggests identifying a finite number of access options defined by class of user, and then developing appropriate legal or technological solutions based on NSI costs and the resulting risk profile. In the UK this model has been used both to classify existing operations and to justify the development of a third-party remote access system and an improved off-site RDC model.
5. Perceptions of Risk

The shift between perceptions is important for the evaluation of risk. Risk is often discussed as if it is measurable, and sometimes this is the case. For example, the large field of research on ensuring that datasets are anonymised to an ‘appropriate’ level quantifies risk as the probability of identification given intruder and protection scenarios; see Duncan et al. (2001) for a typical example.

However, the risk inherent in the data is only one element in a data access solution. The commonly-used ‘VML Security Model’ (also called the ‘Five Safes’ model) classifies risks into those arising from people, projects, the settings, the data and the outputs (see Ritchie 2013 for an expanded discussion). These risk elements interact, and most are not amenable to quantification; for example, how are the risks inherent in an NSI’s procedures for approving projects to be objectively assessed? The problem is made harder because NSIs generally have a very good record of managing confidentiality; examples of misuse of NSI research data are few and far between, and so there is no historical guide to the probability of confidentiality breaches. Risk is therefore a subjective measure, in general, which means it will be affected by the framing discussed earlier.

We earlier characterised the Constraint Model perspective as starting from ‘nothing is allowed’ and then evaluating individual ways to increase access in the context of the legal framework. Any solution therefore increases risk compared to doing nothing. Solutions may be compared to each other for their riskiness, but the default is always to do nothing.

Under the Objective Model, there is no risk baseline. If the NSI has an objective to make data available to a class of users, the default is to hand the data over. All solutions then involve placing some restrictions on that default by, for example, anonymising the data or restricting access. The aim of this is to reduce the risk of a breach of confidentiality, but from an uncertain level. As result, the only comparison that can be made is a subjective comparison to an alternative of equally subjective measurement.

Standard methods do of course exist, such as risk-utility models for evaluation of dataset vulnerability, as well as technical tools like tau-Argus, all adding an element of objectivity. However, like all models, these are parameterised subjectively; see Skinner (2012) for a perspective on perceived versus actual objectivity in NSI decision making. Moreover, while there are guidelines for good practice for nondata factors such as access environments and researchers (Brandt et al. 2010), these are entirely subjective. The Objective Model forces this subjectiveness to be acknowledged.

As noted earlier, losses tend to be weighed more heavily than gains. In the Constraint Model, the losses in security are balanced by gains in data access; in the Objective Model, gains in security are being balanced against losses in access. All other things being equal, the Constraint Model is likely to deliver lower access and higher security, the Objective Model more access and less security.

The psychological literature provides an additional insight. Certainty, all other things being equal, tends to have a higher weight than uncertain outcomes when comparing positive outcomes (see, for example, Viscusi et al. 1987). Consider now the options for an NSI. Benefits and changing risk are uncertain and subjective. The only fixed point is zero risk, which will have more weight in deliberations than the uncertain benefits and risks. Therefore, if the starting point is ‘nothing is allowed’, the outcome is likely to be more
restrictive than starting from an open solution and progressively adding restrictions. In the ‘everything is allowed’ case where there is no clear default measure, risks and benefits are more likely to be equally weighted, if still subjective.

Figure 4 summarises this discussion:

![Figure 4](image)

All other things being equal, the Constraint Model, which is closer to the ‘nothing is allowed’ option, is likely to place more weight on the loss of security and on the do-nothing option. By contrast, the Objective Model, faced with a set of competing but subjective costs and benefits, is more likely to weight the two fairly; and because the ‘losses’ are in access and ‘gains’ are in security, it is more likely to favour access.

6. Objectives, Constraints, and the Public Goods Problem

This article has considered how a change in perceptions to the Objective Model may improve NSI outcomes and operations. A natural question is why the Constraint Model predominates in NSI thinking. Part of the answer lies in the communal nature of research output.

Ritchie and Welpton (2012) argue that one reason why NSIs tend to focus on protecting data rather than maximising value is a ‘public goods’ problem arising from the unequal distribution of risks and benefits. The benefit from making confidential data available for research largely accrues to the wider public, but the risk of being blamed for something going wrong is typically borne by the NSI. For example, if a licensed user is sent a confidential dataset and loses it, the NSI may well get blamed for distributing the data no matter how well founded its distribution policy is. In contrast, if data is not released, or is only used by the NSI for its own purposes, then the NSI minimises risk; but the wider public loses the benefit of that data and runs an increased risk of bad decision making.

In these circumstances it is rational for NSIs to take a cautious approach to data release, and consider their own priorities over the wider public benefit. The NSI’s main function is to protect its interests: Risk avoidance becomes the goal, a conservative legal stance appeals, and the Constraint Model predominates. Even if the NSI takes the perspective of the Objective Model, it is still likely to underestimate benefits and overweight risks.

Ritchie and Welpton (2012) argue that one way to address the public-goods problem is to ‘negotiate’ the level of access with users; as part of that negotiation, issues of risk and responsibility are also addressed. Users are in a better position to identify the benefits from access; but then they need to acknowledge and accept joint responsibility for the risks being run by the NSI. For example, one of the key influences in establishing the ONS remote RDC was the explicit support from the UK Treasury, who made extensive use of the research outputs in their work.
This approach sits comfortably with the Objective Model, which puts agreement with users at the forefront of the decision-making process and views risk as something to be managed, not minimised. For the NSI to set objectives, it needs to consult with users – and this can be used to get the buy-in necessary to ensure a collective responsibility for the data release policy. If that buy-in is not forthcoming, the NSI is arguably justified in ignoring those user needs. Hence the Objective Model is consistent with the customer engagement necessary to avoid underprovision of data access from society’s perspective.

Ritchie and Welpton (2012) propose an alternative: the use of third parties to provide the data access services. In many countries, data distribution has been outsourced for years to third party providers in the form of data archives. However, the bulk of confidentiality protection in these cases is vested in the data. More interesting are the recent moves towards allowing third parties to provide distributed-access services such as remote RDCs; for example, the NORC Data Enclave in the US, the UK Secure Data Service, the IAB RDC-in-RDC in Germany, or the DARA project developing a remote access system for Eurostat. In these, the risk dimensions of people, settings and outputs become much more important than data protection.

The advantage of third parties is that the transparency of contractual agreements forces both parties to identify and acknowledge the risks and the acceptable level of risk management to be employed. This model presents difficulties for the Constraint Model perspective because of the need to refer to the zero-risk baseline. In essence, the Constraint Model specifies inputs to third-party processes (i.e., limits on working), whereas the Objective Model emphasises outputs (targets to be achieved, irrespective of how they are achieved). This gives third-party providers more flexibility in delivering the required outcome; and of course it ties in with the requirement to identify user needs as an initial step.

7. Conclusion

This article has dichotomised the decision-making process for data access into the ‘Constraint Model’ and ‘Objective Model’. Whilst this is clearly an oversimplification, it nevertheless usefully illustrates some different approaches to setting the objectives and solving problems associated with data access. In doing so, this idealised worldview also suggests that NSIs may be missing opportunities for both their benefit and the wider public.

This article has argued NSI decision-making processes tend to focus initially on what is allowed rather than what is desirable; the incentives for NSIs do not encourage exploration of the boundaries of their duties. This is not to argue that NSIs are deliberately acting against the public interest; as Buurman et al. (2012) demonstrate, risk aversion and ‘public spirit’ are two different concepts. Nevertheless, the tendency to risk aversion, however well intentioned, can mean that access to the data collected by NSIs and similar bodies is often unnecessarily restricted.

An alternative perspective focuses on the NSI objectives, and uses this to address questions of constraints in implementation, rather than the other way round. In this perspective, law, NSI procedures, and technology all become ‘enablers’: options for or constraints on implementation which affect the delivery of objectives, but not the objectives themselves.
This is not simply a semantic discussion; the subjectivity of decisions in this area means that the perspective of the NSI directly affects the outcomes achieved. In addition, basic human nature means that decisions about relative risk and uncertainty are affected by the starting point.

Focusing on objectives also provides a framework to bring users into the discussion on access principles, increasing the chance of community buy-in and reducing the NSI’s incentives to implement an overly risk-averse release policy. The objective-based worldview opens up the NSI to wider and deeper engagement with users. Knowledgeable users who recognise the risks but can also express the benefits can help to reduce the public-goods problem associated with research data access.

There are signs that attitudes might be changing. At the 2013 meeting of the major biennial UN conference for government statisticians (http://www.unece.org/?id=31938), a session was held on “Moving from risk avoidance to risk management”. The session papers described a number of positive developments in data access, using a variety of technologies. Of particular relevance here, while most papers focused on the idea of ‘widening access’ – that is, starting from a position of needing to justify any relaxation on data security – the Italian NSI (and, to a lesser extent, Eurostat and Mexico) took a strongly user-centred approach to work backwards from general objectives to specific implementations; see ISTAT (2013).

There is therefore a strong argument that NSIs could benefit from a change in perspective, and some shifts are happening. However, this user/objective-centric approach runs counter to the natural decision-making structures in government; these tend to be cautious, and reflect the Constraint Model. Ritchie (2013) notes that international sharing of confidential data has largely been driven by energetic individuals, rather than any corporate vision (the Eurostat DARA project http://www.safe-centre.info/ is an exception). An efficient system of confidential data access may therefore need active and engaged sponsors at a senior level to have any realistic prospect of success.

8. References


Received July 2012
Revised September 2013
Accepted January 2014